EdWorkingPaper No. 21-432

A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya

Beth E. Schueler
University of Virginia

Daniel Rodriguez-Segura
University of Virginia

Covid-19-induced school closures generated great interest in tutoring as a strategy to make up for lost learning time. Tutoring is backed by a rigorous body of research, but it is unclear whether it can be delivered effectively remotely. We study the effect of teacher-student phone call interventions in Kenya when schools were closed. Schools (n=105) were randomly assigned for their 3rd, 5th and 6th graders (n=8,319) to receive one of two versions of a 7-week weekly math-focused intervention—5-minute accountability checks or 15-minute mini-tutoring sessions—or to the control group. Although calls increased student perceptions that teachers cared, accountability checks had no effect on math performance up to four months after the intervention and tutoring decreased math achievement among students who returned to their schools after reopening. This was, in part, because the relatively low-achieving students most likely to benefit from calls were least likely to return and take in-person assessments. Tutoring substituted away from more productive uses of time, at least among returning students. Neither intervention affected enrollment. Tutoring remains a valuable tool but to avoid unintended consequences, careful attention should be paid to aligning tutoring interventions with best practices and targeting interventions to those who will benefit most.

VERSION: July 2021

Suggested citation: Schueler, Beth E., and Daniel Rodriguez-Segura. (2021). A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya. (EdWorkingPaper: 21-432). Retrieved from Annenberg Institute at Brown University: https://-doi.org/10.26300/43qs-cg37

A CAUTIONARY TALE OF TUTORING HARD-TO-REACH STUDENTS IN KENYA

Beth E. Schueler, University of Virginia
Daniel Rodriguez-Segura, University of Virginia

Abstract:
Covid-19-induced school closures generated great interest in tutoring as a strategy to make up for lost learning time. Tutoring is backed by a rigorous body of research, but it is unclear whether it can be delivered effectively remotely. We study the effect of teacher-student phone call interventions in Kenya when schools were closed. Schools (n=105) were randomly assigned for their 3rd, 5th and 6th graders (n=8,319) to receive one of two versions of a 7-week weekly math-focused intervention—5-minute accountability checks or 15-minute mini-tutoring sessions—or to the control group. Although calls increased student perceptions that teachers cared, accountability checks had no effect on math performance up to four months after the intervention and tutoring decreased math achievement among students who returned to their schools after reopening. This was, in part, because the relatively low-achieving students most likely to benefit from calls were least likely to return and take in-person assessments. Tutoring substituted away from more productive uses of time, at least among returning students. Neither intervention affected enrollment. Tutoring remains a valuable tool but to avoid unintended consequences, careful attention should be paid to aligning tutoring interventions with best practices and targeting interventions to those who will benefit most.

Keywords: Tutoring, distance learning, teacher-student phone calls, international education, educational disruptions, field experiment, Covid-19

Notes: We are grateful to the Center for Reinventing Public Education (CRPE) and Innovations for Poverty Action (IPA) for generously funding this work, and to Tim Sullivan and Sean Geraghty at NewGlobe Education for partnering with us on this project, as well as all of the administrators, teachers, parents, students, and assessors who made it possible. We received valuable feedback from Lee Crawfurd, Jim Soland, Robin Lake, Betheny Gross, Noam Angrist, and seminar participants at the University of Virginia, including Isaac Mbiti and Vivian Wong, among others. Shannon Kontalonis provided exemplary administrative support. We registered a pre-analysis plan for this project via the American Economic Association under ID: AEARCTR-0006954. Please address correspondence to Beth E. Schueler, Ridley Hall 268, University of Virginia, P.O. Box 400277, Charlottesville, Virginia 22904; beth_schueler@virginia.edu.

INTRODUCTION

Schools serving over 1.5 billion children globally mandated temporary closures due to the coronavirus pandemic (World Bank, 2020). Even prior to the pandemic, the international community was grappling with what the World Bank calls a "learning crisis" or the fact that in low- and middle-income countries, over half of all 10-year-olds are unable to read an age-appropriate story. The Covid-19 pauses on in-person schooling threaten to exacerbate these conditions. This is particularly true in developing contexts with low levels of overall educational attainment where the length of the school closures represents a non-trivial share of the average student's overall time in formal schooling. Closure time is estimated, for example, to be roughly seven percent of the overall average schooling time in Kenya (Evans, Hares, Mendez Acosta & Saintis, 2021). The 2014 Ebola Crisis in West Africa illustrates the dramatic effects a pandemic can have on educational outcomes (Bandiera et al., 2019; Santos & Novelli, 2017). Furthermore, given all that social scientists know about the unequal effects of natural disasters, crisis-induced economic downturns, absenteeism, trauma, and learning time (e.g., Andrabi, Daniels & Das, 2020; Sacerdote, 2012; Shores & Steinberg, 2017; Bacher-Hicks, Goodman & Mulhern, 2021), Covid-19 is likely to not only harm student learning overall but widen educational opportunity and outcome gaps between the world's most advantaged and disadvantaged students.

The likelihood of growing educational inequality has led to significant policy interest among researchers, policymakers, and pundits alike in tutoring as a promising strategy to address COVID learning loss, support struggling students, and curb growing educational inequalities (e.g., Dynarski, 2020; Blitzer, 2020; Brooks, 2020). This would require targeting tutoring to relatively disadvantaged students as access to private tutoring is currently inequitably distributed both in developed countries (e.g., Kim, Goodman & West, 2021) and developing nations (e.g., Dang & Rogers, 2008; Jayachanran, 2014) where it is sometimes referred to as “shadow education” (Bray, 2021). Indeed, the U.K. government has launched a national tutoring program to help disadvantaged students catch-up on lost learning time (Daily News, 2021) and the U.S. Department of Education points to tutoring as one of only a handful of highlighted interventions for addressing lost instructional time (U.S. DOE, 2021).

The attention to tutoring in this context is well-deserved. An unusually large and consistent body of rigorous causal studies, mostly from the U.S., illustrate that tutoring can deliver consistently large academic benefits across subjects and grade levels (Nickow, Oreopoulos & Quan, 2020; Fryer, 2016). Studies of related interventions suggest that efforts to individualize instruction to students’ learning levels can also generate meaningful academic improvements in both developed and developing contexts (Romero, Chen & Magari, 2021; Banerjee, Cole, Duflo & Linden, 2007; Dang et al., 2008; Duflo, Dupas & Kremer, 2011; Banerjee et al., 2017; Muralidharan, Singh & Ganimian, 2019; Schueler, 2018). However, an important theme in this literature is that program design matters for a tutoring program’s effectiveness in a variety of contexts (Cabezas, 2011; Jayachandran, 2014; Song et al., 2018; Kraft & Goldstein, 2020) and, despite the existence of gold standard interventions, scholars have also documented examples of ineffective tutoring programs (e.g., Heinrich, Meyere & Whitten, 2010; Heinrich et al., 2014). More specifically, the most effective programs tend to be delivered in high doses with three or more 30-60-minute sessions per week and conducted during the school day rather than after-school or summer (Robinson, Kraft, Loeb & Schueler, 2021).

To date there has been limited research on whether tutoring can be effectively delivered at a distance although technology could improve both the safety and scalability of individualized instruction (Rodriguez-Segura, 2021; Ganimian, Vegas & Hess, 2020). Recent evidence from

Italy suggests that free online tutoring for disadvantaged middle schoolers delivered by university student volunteers for three to six hours per week improved academic achievement, socio-emotional skills, aspirations, and a measure of psychological well-being (Carlana & LaFerrara, 2021). Relatedly, Roschelle et al. (2020) conducted a small-scale evaluation of an elementary math one-on-one online tutoring program and found that participants made larger math gains than non-participants. Previous work on remote in-school instruction via satellite in developing contexts (e.g., Johnston & Ksoll (2017) in Ghana, Naik et al. (2016) in India, and Bianchi et al. (2019) in China) provides relatively promising results, but these studies have mostly focused on group instruction, not personalized lessons.

Much ongoing scholarship on minimizing coronavirus learning loss from a distance focuses on online learning. However, many young people around the world have no access to internet at home, such as those in rural communities, refugee camps, conflict zones, and low-income communities in digital deserts. Worldwide, roughly 42 percent of all individuals in low- and middle-income countries report using the internet (World Bank, 2017). This figure is likely lower when considering home-based internet. Even in the U.S., at least 14 percent of households are without internet (NCES, 2018) and access to technology is inequitably distributed. In Kenya, the context of our study, for homes where the “head of household” has received postsecondary education, 60 percent have internet at home, whereas this is true for only 3 percent of households led by someone with less than a postsecondary education (Kenya National Bureau of Statistics, 2018). These unequal levels of access to technological solutions to mitigate learning losses both across and within countries have raised concerns that the pandemic will have long-lasting effects on enrollment and educational attainment for some of the world’s most disadvantaged children (Muhumuza, 2020; Musu, 2018; Mays & Newman, 2020; Parker et al., 2020).

Therefore, there is a need to understand the potential for using more basic forms of technology to support hard-to-reach learners. Some research has demonstrated that phone-based communication can improve educational engagement, though this work has mostly been limited to teacher-parent rather than teacher-student communication (e.g., Kraft & Dougherty, 2013; Doss et al., 2019; Bergman, 2015; Berlinski et al., 2016; Hurwitz et al., 2015; Kraft & Monti-Nussbaum, 2017; Mayer et al., 2015), or to adult learners (Aker & Ksoll, 2019; Ksoll et al., 2015), high school graduates (Bird et al., 2019; Castleman & Page, 2016; Oreopoulos & Petronijevic, 2018), school officers (Dustan et al. 2019), or teachers (Jukes et al., 2017; Vakis & Farfan, 2018), rather than K-12 students. Researchers have focused on phone-based technology for informational or accountability purposes (e.g., "nudging" college-intending students with registration information or texting parents data on children's attendance). Little is known about whether individualized instruction itself can be accomplished from a distance through low-cost and wide-reach technology like phone calls.

One notable exception is a recent study testing the impact of text messages providing a "problem of the week" and weekly 15-20-minute phone calls aimed to minimize educational fallout due to the pandemic in the context of Botswana. Twelve weeks of phone calls improved math achievement in the short run, though the text messages alone had no effect (Angrist, Bergman & Matsheng, 2021). However, this study was unable to disentangle the impact of the accountability aspect of the phone call from an educator, from the instructional component, leaving open questions about the viability of a potentially more scalable accountability-focused intervention using shorter calls, as well as about the extent to which tutoring itself is necessary or viable when delivered at a distance. Additionally, the outcomes for this study were relatively short-term based solely on a phone-based, rather than in-person assessment.

There are some reasons to believe that contact with a teacher during periods of educational disruptions, even without direct instruction, could improve outcomes. Student engagement with learning is essential for the educational process (e.g., Wang & Holcombe, 2010; Fredricks et al., 2004). Engagement can be challenging to cultivate in typical times let alone when educators are not seeing students in person. Teacher demonstrations of caring have strong associations with student academic effort (Connell, 1990; Finn & Rock, 1997; Battistich et al., 1997; Furrer & Skinner, 2003; Wentzel, 1997, 1998). Additionally, a substantial behavioral science literature suggests basic accountability can encourage engagement (Gill et al., 2017; Gill, 2020; Lerner & Tetlock, 1999). In the context of distance learning, the research on online charter schools in the U.S. suggests that the dearth of synchronous exchange helps explain why these programs haven't been more successful (e.g., Gill et al., 2015). Therefore, direct teacher-student live communication could powerfully encourage engagement through relational and accountability mechanisms. On the other hand, given the aforementioned literature on the impressive benefits of high-dosage tutoring, accountability alone may not be enough to generate learning gains or the impacts may be much larger for a tutoring than an accountability-only intervention, particularly for students who are struggling academically.

We study whether providing direct teacher-student phone-based communication and instruction can improve academic achievement and directly compare the relative effectiveness of accountability versus instructional mechanisms by randomly assigning 105 schools serving 3rd, 5th, and 6th graders to either (1) five-minute teacher-student phone calls focused on accountability, (2) 15-minute teacher-student phone calls focused on a mini-tutoring session in addition to basic accountability, or (3) a control group. The intervention occurred over seven weeks while students engaged in distance learning in Kenya where cell phone access is relatively ubiquitous but the large majority of households (93 percent) are without internet (Kenya Bureau of Statistics, 2018). We measure achievement outcomes in multiple waves up to four months after the intervention, using both phone-based and in-person assessments.

Although we find that both types of phone calls increased student perceptions that teachers cared about their students, unfortunately, accountability checks had no effect on math performance up to four months after the intervention and tutoring actually decreased math achievement among those students who returned to in-person schools after they reopened. This was, in part, because the relatively low-achieving students who appeared most likely to benefit from the phone-based interventions, at least in the short-run, were least likely to return to school after reopening and therefore least likely to take in-person assessments and contribute to our estimates of the long-term program effects. We provide suggestive evidence that the relatively low-touch phone-based tutoring substituted away from more productive uses of time, at least among returning students. Notably, the negative tutoring effects were on in-person exams with relatively weak alignment to the content of the intervention while effects were null on shorter-term phone-based assessments more closely aligned to the content of the intervention. Neither intervention affected enrollment.

In no way do we take these results as a referendum on tutoring programs as a whole, as in hindsight there were several key differences in the design of the intervention under study here from those high-impact tutoring interventions for which researchers have documented impressive benefits using rigorous methods. Notably, the phone-based tutoring intervention we study was light-touch relative to the state-of-the-art high-dosage tutoring models and was delivered outside of school and at a distance, in large part due to logistical and financial constraints. Nonetheless, our results have important implications for policymakers and educational leaders in the midst of determining how best to spend relief funding and to support students who lost learning time due to the pandemic. As policy interest in tutoring grows, our study provides important lessons about the need to pay careful attention to designing tutoring programs so that they are aligned with those high-impact models, to effectively target interventions and limited funds to those groups of students most likely to benefit, to weigh trade-offs between scale and intensity of an intervention, and to attempt to anticipate and minimize unintended consequences.

TEACHER-STUDENT PHONE CALLS

Setting

Our study was conducted in the East African nation of Kenya. Coronavirus-induced school closures in Kenya lasted for a total of 37 weeks, with some variation by grade level (UNESCO, 2021). Although Kenya is among the top ten countries in Africa in terms of total coronavirus cases, it not among the hardest hit nations in the world, experiencing roughly 3 percent of the number of both cases and deaths per 100,000 people as that of the U.S (CNN, 2021). The study occurred in partnership with NewGlobe which operates Bridge Kenya, a network of low-cost private schools. NewGlobe also operates or supports several other programs, including both public schools operated as part of public-private partnerships and low-cost private schools. NewGlobe has served over 750,000 children in over 2,000 schools operated in India, Kenya, Liberia, Nigeria, and Uganda (Bridge 2018).

Students participating in the intervention came from low-cost private Bridge schools across Kenya which, at the time, were closed for in-person learning due to the Covid-19 pandemic. These students and their families are likely somewhat different from typical families enrolled in Kenya's public schools. For example, nationally, 27 percent of families report the mother having no formal education while this is true for only one percent of our sample.

Conversely, roughly 29 percent of parents in our sample received at least a secondary education while this is true for less than one percent of parents nationally (Twaweza, 2014). Parental education in this context is strongly correlated with other factors that could moderate learning loss. For instance, only 4.5 percent of all Kenyan households where the head only has primary school has a computer, and 2.9 percent have internet at home, whereas these figures are 60.3 percent and 57.7 percent respectively for households where the head of the household has some higher education degree (Kenya Bureau of Statistics, 2018). In sum, our sample comes from more educated households, likely with better technological tools and stronger practices of at-home learning during school closures than the average Kenyan home.

Control Condition

During the period of school closures, all Bridge students in Kenya – including those in the treatment and control groups alike – had access to a relatively robust set of distance learning materials, especially for the Kenyan context. This included a webpage called Bridge@Home--a portal containing daily learning guides and activity bundles for each grade and subject. Bridge also was reaching out to parents via text messaging to provide daily recommendations for homework assignments and supplementary opportunities. They moderated WhatsApp parent groups by grade level where Bridge employees shared daily learning guides and activity bundles and answered questions about learning from home. Bridge had digitally catalogued mobile-friendly storybooks so families could access age-appropriate texts via phone. The organization also developed a series of mobile quizzes that students could take directly from a cellular phone designed to provide additional curriculum-aligned practice opportunities from a distance. Families with radios also had access to educational radio programming throughout this period provided by Kenya Institute for Curriculum Development and encouraged by Bridge.

Participating Schools and Students

We focused on primary schools and within those schools, grades three, five, and six as the focal grades that received the intervention. We selected grades to include a range of ages while remaining within budget and because grades four and eight returned to school earlier than other grades. Our sample included a total of 105 Bridge schools for which we had baseline data. As shown in Appendix Figure A1, the schools were located in three main clusters around the major cities of Kisumu in the west, the capital city of Nairobi in the center, and the eastern coastal city of Mombasa. However, the schools were not entirely urban, as 11 schools are located in rural areas and 50 in “peri-urban” settings. Table 1 describes the sample of students. On average, these students came from communities where 34 percent of residents were living in multi-dimensional poverty and the female literacy rate was 85 percent. On average, participants came from schools where student attendance rates were around 50 percent while teacher attendance was higher at 84 percent. (Note that in the tables we have used the most conservative student attendance measures available. Average attendance rates are more like 72 percent in these schools based on the metrics that are more typically publicly reported and the metrics are are highly correlated.)

Training and Staffing

The intervention was implemented through a cascade model of training in October 2020. A central Bridge staff member trained regional managers on the intervention. Regional managers then trained school supervisors and provided coaching and mentorship support on their roles and responsibilities. School supervisors then trained school principals (called “academy managers in this context) and principals trained their teachers on how to complete the teacher-student phone calls. School supervisors were also responsible for troubleshooting problems related to airtime provision or usage and responding to issues raised by teachers or revealed via weekly online reporting forms teachers were required to complete to monitor progress. Teachers received weekly transfers of airtime to be used for completing calls to students as well as weekly stipends that were, on average, equivalent to ~$37.40 per month (~$9.40 per week) via a mobile phone-based money transfer service. This was a non-trivial amount relative to typical monthly teacher salaries in this context. Due to in-person school closures, teachers' other responsibilities were scaled back during this period. The overall per pupil cost of the full intervention was roughly $3.90 for the accountability intervention and $6.80 for tutoring.

Interventions

The teacher-student phone call interventions were conducted over the course of seven weeks in October and November of 2020, while in-person learning was suspended, and prior to the typical December vacation break and school re-openings in January 2021. Treatment group students received one of two versions of a teacher-student phone call intervention which we describe below, randomly assigned at the school-level. Teachers called students from their own class who they had previously taught in person prior to Covid-19 shutdowns. Treatment group students in both arms of the intervention received weekly SMS text messages sharing practice problems that could then be referred back to during the calls themselves. The weekly SMS communication included a welcome message, a practice problem, guiding tips to master the topic, an additional challenge problem, and recommended mobile quizzes related to the content. All of the content for this intervention focused on mathematics. We provide an example SMS communication in the Appendix. For both intervention arms, teachers were trained that phone calls should occur once per week per pupil over the seven-week period.

The first of the two treatment arms, we call "accountability." For this intervention, teachers provided brief phone calls to families focused on accountability checks. Teachers were instructed to first gather information on whether the student attempted the SMS problems, answered the problems correctly, and completed the recommended interactive quizzes. Then, they discussed next steps including recommended additional mobile interactive quizzes and reminding them of the next time the teacher plans to call to check in. Teachers were trained that each individual call should last approximately five minutes. We include sample transcripts that were distributed to teachers, separately by treatment arm, in the Appendix.

The second treatment arm, we call "tutoring." This version of the intervention includes all aspects of the accountability arm but also includes the provision of direction instruction and feedback to students on academic content. More specifically, teachers were instructed to first gather information on whether the student attempted the SMS problems, answered the problems correctly, and completed the recommended interactive quizzes. Next, the teacher asked whether the student had questions when solving the SMS problem. After answering any questions, the teacher delivered a brief demonstration, teaching the student the skills necessary to complete the SMS problem. The teacher then instructed the student to complete additional practice problems after the call. The teacher concluded by discussing next steps, including recommended additional mobile interactive quizzes and reminding them of the next time the teacher planned to call back. Teachers were trained that each individual call should last approximately fifteen minutes.

In total, teachers in the accountability group were expected to make an average of 130 minutes worth of phone calls each week (a little more than two hours) while teachers in the tutoring group were expected to make 390 minutes (six and a half hours) worth of calls per week (based on an average class size of 26.4). All students in a given teacher's class were part of the same treatment arm. Airtime was transferred to teachers weekly on Tuesdays, teachers were asked to complete all calls and the online reporting form by Thursdays, and reporting form data were reviewed by administrators on Fridays. Stipends were provided at the end of the intervention. Airtime and stipend amounts varied by teacher, calculated based on the expected length of calls and the number of students enrolled in a given teacher's class.

Teachers in both treatment arms were advised to call pupils in alphabetical order and to start back at the top of the list later in the day or week for any students not reached the first time around or to call back at a pre-agreed time if the parent picked up the phone but was not able to speak. Trainings emphasized that the goal was to reach all pupils each week. Teachers were instructed to begin by greeting the parent, explain that they are calling to provide support in math, and then ask the parent to put the phone on speaker so that the pupil can join. If teachers reached all students and still had airtime remaining, they were instructed not to complete more than one call per pupil each week. If teachers ran out of airtime before they could reach all students, they were simply asked to keep their calls shorter the next week. Treatment was assigned at the school level in part to minimize spillover effects within schools or contamination that could have occurred had we randomized at the student or classroom level.

The weekly online reporting forms asked teachers to provide their name, grade, and school, as well as the number of calls completed for that week (the number of families that were successfully reached), the number of students they attempted to call (regardless of whether they were reached), the average length of calls, a list of any students they were not about to successfully reach (and an indication of whether this was because the mobile number was unavailable or because the family simply did not pick up the phone), any issues with the transfer or amount of airtime, and finally, any important issues or questions raised by parents during the calls. These were typically completed via Google form accessed by smartphone.

During the first two weeks of the intervention, a team of twelve Bridge staff members made “confirmatory calls” to a random sample of approximately 400 parents from the treatment groups per week to assess whether parents were receiving the calls as intended. If parents had not received calls, administrators followed up with the teacher concerned to provide additional support. After two weeks, it was determined that calls were being completed with enough consistency that there was not a huge value-add to continue the confirmatory calls.

EXPERIMENTAL DESIGN

Randomization

We include 105 schools in the experimental sample, representing 94 percent of all Bridge schools in Kenya, but excluding those for which we were missing exact latitudes and longitudes or digitized baseline achievement data (both of which we used for the randomization process). More specifically, to improve statistical power, we began by blocking schools based on covariates that likely explain variation in outcomes. First, we created three bins for the size of the population living within a five-kilometer radius surrounding each school as a proxy for urbanicity. The GIS population data comes from Bosco et al (2017), downloaded at a resolution of 1-km grids at the equator. These bins span from approximately 6,000 people to 55,000 for the rural category, 55,000 to 170,000 for the peri-urban category, and greater than 170,000 for the urban category (with a maximum of roughly 1,850,000). Second, we split each of the three bins into quintiles representing school-level baseline exam scores. We then randomly assigned schools within each of the 15 resulting randomization blocks to one of three groups: (1) control, (2) accountability, or (3) tutoring. Ultimately, there were 35 schools in each group.

We describe the baseline characteristics of the control, accountability, and tutoring groups in Table 1, demonstrating balance on observable baseline characteristics across groups. Only one baseline difference for one group (less than five percent of the differences) was statistically significant: the accountability group had lower baseline English test scores than the control group. We observe two significant baseline differences between students in the accountability and the tutoring treatment arms, as shown in the last column of Table 1. Specifically, the tutoring group is somewhat younger in age than the accountability group and also came from schools with a slightly lower baseline pupil attendance rate. As a result, we are careful to examine both unconstrained estimates of treatment effects as well as estimates controlling for this and other baseline characteristics.

We estimate that we have 0.80 power to detect a minimum treatment/control contrast of 0.12 standard deviations and a minimum contrast between treatment arms of 0.14 standard deviations, comparable to the short-term effects of phone-based tutoring documented by Angrist, Bergman & Matsheng (2021). This is after accounting for the inclusion of covariates, including pre-shutdown baseline assessment results, as well as blocking explaining 50 percent of the variation in our outcomes. For the purpose of this power calculation, we assume an intra-school correlation of 0.10 (observed at baseline) and that teachers would on average successfully reach 80 percent of the students on their call list (based on a pilot phone-call program our partner had previously conducted). We pre-registered our analysis plan via the American Economic Association RCT Registry after randomization but prior to analysis.

Compliance

Based on the school rosters from February 2020, prior to pandemic-induced school shutdowns, there were 8,319 students in grades 3, 5, and 6 enrolled in the 105 schools randomized to one of the three conditions. We rely on these 8,319 students as our primary analytic sample, although this is a conservative approach because we know that many students left school or transferred schools between February 2020 and October 2020 when our intervention was delivered. The intervention was communicated to schools starting in September 2020. We do observe which students appeared on an updated call list that was provided to teachers in treatment groups making phone calls in October-November of 2020. These call lists were updated to exclude students who administrators knew had left the school and to include any new students who joined the school. This was done to minimize the extent to which teachers spent time calling students who were no longer enrolled in their school. Unfortunately, such lists were not generated for control schools, as teachers in control schools did not need updated rosters. We use these updated call lists to generate our "treatment-on-the-treated" estimates, described in more detail below. However, we are unable to observe whether an individual student was successfully reached by phone. Therefore, even these treatment-on-the-treated estimates are likely an underestimate of the effect of receiving a phone-based intervention.

Again, we use the February 2020 rosters for our intent-to-treat estimates because we do not have access to an updated roster for the control group at the time of the intervention. In Table 1, we show that students who ended up on the call lists (from among those in the original experimental sample) were somewhat different on observable baseline characteristics than those who did not, particularly among those who had been assigned to the accountability treatment arm. More specifically, these "compliers" were slightly younger, had a somewhat lower placement score, came from slightly somewhat larger communities, and came from slightly smaller schools than those who had, in February 2020, been enrolled in schools that were assigned to the accountability intervention but did not remain in the school as of fall 2020.

However, none of the magnitudes of these differences is especially large. For the tutoring treatment group, students who remained on the call list had slightly lower baseline Kiswahili scores and came from slightly larger schools. Importantly, attrition did not occur systematically across treatment assignment groups and is not, therefore a threat to internal validity.

Although we do not observe data on receipt of phone calls at the student level, we do have some information that speaks to the aggregate rates at which students assigned to a treatment group actually received the intervention. First, among a subsample of students to whom we administered a post-intervention phone-based assessment and survey, described in more detail below, children in both the accountability and tutoring groups reporting receiving roughly one more phone call per week from their teacher than children in the control group. This aligns with the design of the intervention, meant to provide one weekly phone call to all treated students. Additionally, teachers were not provided with phone numbers for the families in the control group and these students were not in their classes or even enrolled in their school when it had been open for in-person learning, therefore, it seems unlikely that a significant number of control group students would have received the intervention.

We also have data from the weekly online form that teachers implementing the interventions completed. Ninety five percent of teachers completed these forms (between 75 and 95 percent in any given week). On average, teachers reported successfully reaching approximately 75 percent of the students who they attempted to call. In an average week, 31 percent of teachers reported that their average calls lasted between 15 and 20 minutes, 29 percent said between ten and 14 minutes, 36 percent said five to nine minutes, and four percent reported average calls between one and four minutes. Unfortunately, we are not able to break these numbers out by treatment arm, but these numbers at least give some level of confidence that although compliance was not perfect it appeared to have been reasonably high.

METHODS

Data

Administrative Data. We rely primarily on student-level administrative data provided by NewGlobe. We begin with the full roster of the 8,319 students enrolled in the 105 schools in our analytic sample as of February 2020 prior to school closures. These data include demographic characteristics and baseline achievement information at both the student and school level, as well as basic enrollment information such as grade and school. We merge these data with the geospatial data described above for the purpose of generating randomization blocks, and then information on treatment assignment as well as presence on the call lists as of October 2020. There are four different sources of data that we use to measure our outcomes, described below loosely in the chronological order in which they were collected.

Phone-Based Assessments. First, we administered a phone-based assessment (PBA) and accompanying student and parent survey between December 7 and 23, 2020. This occurred after the intervention had been implemented but before schools reopened. To do so, we randomly selected a sample of 6,295 students to be assessed from our analytic sample of 8,319. This random sample was representative of the full sample on all observable dimensions. However, assessors were only able to reach 2,552 students within the period leading up to the holiday break. As we show in Table 2, the sample of students for which we have PBA data was somewhat higher achieving at baseline on all subjects (math, Kiswahili, and English) than the full sample, came from somewhat larger communities, and came from schools with slightly higher baseline attendance rates and larger enrollments. In short, the sample for which we have

PBA outcome data is not fully representative of the broader sample, as it seems to be somewhat more advantaged. We discuss the implications of this in our findings section. Importantly, as we illustrate in Table 3, we observe treatment/control balance on baseline characteristics for those in the PBA sample. Therefore, we do not see a threat to the internal validity of our estimated effects on this outcome. Generally, we observe balance between the two treatment arms. The exception is that the tutoring group was somewhat lower achieving at baseline in Kiswahili than the accountability group, and also had slightly lower student attendance rates at baseline. We are therefore careful to control for these characteristics in our models.

The PBA was conducted by hired enumerators and consisted of 14 questions, covering two predetermined sections on (1) core numeracy, and (2) curriculum-aligned standards based on what students would have been learning had schools been open and what they were supposed to be learning as part of the phone-based interventions. At the end of the assessment, we included a short student and parent survey, with one question for students on the extent to which they feel their teacher cares about their learning, along with five questions for parents on at-home study habits, COVID-19-related shocks, and their educational attainment. The curriculum-aligned questions varied across grades while the core numeracy section and survey questions were the same across grades. This allows for benchmarking to the typical annual growth on core numeracy for students in the same grade levels. For the math scores (overall and by section), we generate IRT-based outcomes and use these as our main PBA outcomes but confirm that our results are not sensitive to this choice. Elsewhere we provide evidence that results from this measure are correlated with in-person assessment results and fairly accurately classify students based on their mathematics performance, albeit more noisily than in-person exams (Rodriguez-

Segura & Schueler, 2021). Similar measures have been used by scholars evaluating other educational interventions from a distance in developing contexts (Angrist et al., 2020).

In-Person Assessments. We also merge in data from an in-person assessment administered by Bridge in February 2021 (n=5,665) as well as an in-person endterm assessment administered by Bridge in March 2021 (n=5,527) for those students for which it was available among members of our analytic sample. These are internal standardized exams typically administered by schools within this network multiple times per year. Therefore, these data help us assess the extent to which the phone-based interventions had an effect on student achievement three and four months after the treatment occurred and once the students were back to in-person learning. However, it is important to note that these exams cover content that goes beyond that which was covered during the teacher-student phone calls. A mapping of content comparing those topics covered during phone-based tutoring and the percent of exam questions covering those topics suggests that across grades, there was 50 percent alignment on the February 2021 in-person math assessments and 57 percent alignment on the March 2021 in-person math exams. This is lower than the phone-based assessment for which we observe 76 percent alignment between the exam questions and the content of the intervention.

As we show in Table 2, the samples for both the February and March in-person assessments were slightly younger than the analytic sample as a whole. They were also somewhat higher achieving at baseline on all three subjects and came from somewhat larger communities as well as schools with slightly higher attendance rates. These patterns were consistent across the February and March in-person samples, suggesting that these differences likely reflect patterns in the types of students more likely to return to in-person schooling in the early months of reopening. In Table 3 we show that there was balance across treatment and control groups for both the February and March in-person assessments. The one consistent exception is that, for both the February and March in-person assessment samples, the treatment group members had lower baseline English test scores than the control group members. Additionally, on the midterm, the tutoring treatment group came from communities with somewhat higher rates of female literacy than the control group members. Therefore, we are careful to check the robustness of our findings to the inclusion of controls for these, and other, observable baseline differences.

Post-Covid Enrollment Data. We also rely on updated enrollment rosters from March 2021 to estimate the effect of the intervention on whether students returned to in-person learning at Bridge schools after in-person schooling resumed. This is a policy-relevant outcome given the uncertainty about whether students will return to school after this significant pandemic-induced learning disruption, and whether rates of return will vary depending on characteristics of students, their families, and their schools. More specifically, we generate a variable equal to one if a student was found on Bridge's March 2021 enrollment roster, regardless of whether we have other outcome data (e.g., non-missing test score values) for that student. In addition, we estimate effects of treatment on whether a student is missing a value on each of the outcomes, regardless of whether they are enrolled in a Bridge school and find no evidence that any of our results are an artifact of treatment effects on missingness, or of differential attrition by treatment arm.

Phone Survey. Finally, we administered an additional phone-based survey to a subset of families in April 2021 to learn more about the mechanisms behind the impacts we were observing in our preliminary analysis. For this survey, we asked students and parents to think back to the remote learning period and to tell us about the parental help they or their child was receiving with studying as well as their perceptions of the adequacy of learning supports from the school. We focused on all 2,132 students in a single grade level—grade three. This choice was in part because we were simultaneously collecting data for a second project for which assessing grade 3 made the most sense. Of all third grade students on the original 2020 roster, assessors were able to reach 1,036 students (51 percent). As we show in Table 2, this sample was higher performing than the full population of third graders on the 2020 roster. However, this sub-sample was more representative of third graders who had returned to in-person school in 2021 than the original December PBA sample was of the 2020 roster sample. Table 3 demonstrates that we observe treatment/control balance on baseline characteristics for this sample.

Analytical Approach

Intent-to-Treat Estimates. We begin by estimating the causal effect of being enrolled in a school randomized to receive one of the phone-based interventions on outcomes by generating intent-to-treat (ITT) estimates of the following form:

$$Y_{is} = \beta_0 + \beta_1 Accountability_{is} + \beta_2 Tutoring_{is} + \mathbf{P}X_i + \mathbf{M}S_s + \theta_r + \varepsilon_{is} \qquad (1)$$

Here, Y_{is} is an outcome measure for student *i* in school *s.* β_1 is the coefficient associated with assignment to the accountability arm of the intervention and β_2 the coefficient corresponding to assignment to the phone-based mini-tutoring arm. To increase precision, we include a vector of baseline achievement and demographic controls measured at the student level (X_i) as well as a vector of controls varying at the school level (S_s). Student-level covariates included gender, age, placement score at the time of joining Bridge Kenya, and a baseline test score. School-level covariates included population count in a 5-kilometer radius around the school, average multidimensional poverty rate in a 5 kilometer radius around the school, average adult female literacy rate at the school, and distance to nearest cell tower in meters. We also include fixed effects for randomization strata (θ_r). However, in this case, our results are not sensitive to this choice. Following the advice of Abadie, Athey, Imbens and Wooldridge (2017), we cluster standard errors at the school level.

Treatment-on-the-Treated Estimates. Because assignment to the treatment group did not guarantee that a student remained at a school randomized to treatment by the time the interventions were delivered, we also generated treatment-on-the-treated (TOT) estimates to assess the effect of being on a call list to receive a phone-based intervention. We note, however, that this is likely an underestimate of the effect of actually receiving a phone call, as we do not have the data to distinguish between children on the call list who were and were not successfully reached by their teacher. Nonetheless, we generated these pseudo-TOT estimates in two stages, using the following model for the first stage:

$$\text{Call\_List}_{is} = \alpha_0 + \alpha_1 \text{Treatment\_Group}_{is} + \mathbf{P}\text{X}_i + \mathbf{M}\text{S}_s + \theta_r + \varepsilon_{is} \tag{2}$$

where the outcome is an indicator equal to one if a student was on the call list provided to teachers conducting the phone-based interventions and α_1 represents the relationship between assignment to treatment (being enrolled in a school in February 2020 that was assigned to the treatment group) and ending up on the call list. We run this separately for those assigned to the accountability phone calls, dropping those observations assigned to the tutoring phone call intervention, and vice versa. The rest of the controls are the same as those included in model (1). Assignment to treatment does indeed predict whether a student was on the call list (α_1=0.88, $p<0.001$). We then used predicted values for presence on the call list generated by the first stage model to estimate the following second stage equation:

$$\text{Y}_{is} = \beta_0 + \beta_1 \widehat{Call\_List}_{is} + \mathbf{P}\text{X}_i + \mathbf{M}\text{S}_s + \theta_r + \varepsilon_{is} \tag{3}$$

where Y_{is} is an outcome measure and β_1 is the coefficient associated with being on a call list for a phone-based intervention. Again, we include the student- and school-level covariates described above and randomization strata fixed effects. We cluster standard errors at the school level.

Sample Weighting. Due to the fact that the subsamples of students for which we have outcome data available are not always perfectly representative of the original analytic sample, we run versions of the models above but weighting our sample to reflect the composition of the full sample of students from which each sub-sample was drawn. We begin by estimating each student's propensity to be assessed in a given data collection wave by regressing a dichotomous variable for whether that student was in the assessment sample on student, school, and community characteristics, as well as school and grade fixed effects, clustering standard errors at the school level. We then re-run the above models but including weights such that an observation with a high likelihood of not being assessed in a given wave contributes more to our estimates.

FINDINGS

What Impact Did Phone Calls Have on Achievement?

We display ITT results in Table 4. We find that phone-based accountability checks had a small positive effect on short-term math performance on the phone-based assessment (PBA) by 0.04 standard deviations, but this effect does not achieve statistical significance. The direction of the effect is negative for phone-based tutoring (-0.03 standard deviations) but again this effect is not statistically different from zero. Our conclusions remain unchanged after controlling for baseline covariates and weighting to reflect the full sample. The TOT results for the PBA, reported in Table 5, are very similar. We also confirm in results not reported here that our findings are unchanged when we include fixed effects for PBA enumerator and when we control for the order in which a student was reached and assessed by the enumerator, given other work suggesting that these factors can influence measurement of the outcomes (Rodriguez-Segura & Schueler, 2021). Although we cannot rule out very small effects, with a 95 percent confidence interval, we can rule out effect sizes of 0.12 standard deviations or larger for the accountability intervention and of 0.06 standard deviations or higher for the mini-tutoring treatment arm.

Turning to the February 2021 midterm results, gathered after schools had reopened, our ITT estimates in Table 4 suggest that the accountability intervention had a negative effect on math performance. When including covariates in the model, the coefficient for math achievement remains negative (-0.08 standard deviations), though not statistically significant. The patterns are very similar based on the TOT results. For tutoring, we observe negatively-signed effects on the order of -0.10 standard deviations that again do not achieve statistical significance. These patterns are identical based on the treatment-on-the-treated estimates. Again, findings are generally consistent when weighting to reflect the full sample. We find no differences in the effects of the accountability versus tutoring intervention on the midterm exams.

For the March 2021 in-person endterm assessment, we observe negatively-signed effects of the accountability intervention on the order of -0.11 standard deviations that are not statistically significant and that are attenuated to -0.06 standard deviations when we include covariates in our model. This pattern is similar across the ITT and TOT results though the magnitude when including covariates is even smaller with the TOT specification (a non-significant -0.03 standard deviations). In contrast, for the tutoring intervention, we observe statistically significant negative effects on the order of -0.16 standard deviation units based on the ITT. These effects persist when controlling for baseline covariates, weighting the sample, and estimating TOT effects. None of the differences between the effects of accountability compared to tutoring were statistically different from zero, regardless of the outcome, although in many cases we are underpowered to detect differences. In Table 6, we demonstrate that the interventions did not impact whether a student was enrolled in a Bridge school at any wave of data collection and findings are not due to differential missingness on outcomes across the treatment and control groups at any given wave.

Why Do Results Differ Across Assessments?

Importantly, the evidence does not suggest that students receiving the tutoring intervention, on average, experienced null short-term effects and then negative longer-term effects. In fact, the different results between the PBA and the in-person assessments appear to be due at least in part to differences in the samples that were assessed at each wave. As we show in Table 2, for all assessment waves, the sample of students who were assessed was not perfectly representative of the full analytic sample. In the case of both the PBA and the in-person assessments, the tested samples were more advantaged than the non-tested samples (formal tests are shown in the "difference" columns). However, the baseline differences between students who were assessed and those who were not assessed on the February and March in-person exams were larger than the differences on the PBA. Students who returned to in-person schooling in February and March of 2021 were higher achieving at baseline than those who did not return. Returning students were also younger, on average, than those who did not return.

The shift in composition of the sample by assessment is important because we find that those students who were least likely to return to school by March 2021 were most likely to benefit from the intervention in the short-term. We present the first piece of evidence for this in Table 7 where we re-estimate the effect of the interventions on the PBA outcome but separately for the sample of students who were and who were not present for the March 2021 in-person assessment. For those who were present for both the PBA and the in-person exam, we observe null effects on PBA math achievement of accountability and statistically significant negative effects of tutoring on the order of -0.09 standard deviations. In contrast, among those students who were present for the PBA but not present at the March in-person endterm exam, we observe large positive effects on the math PBA for both accountability and tutoring, on the order of 0.17 and 0.14 standard deviations respectively, although these estimates are estimated imprecisely due to the small sample size.

We also calculate a student's propensity to return to school and be present for the March in-person assessment in 2021 based on all observable baseline characteristics. In Figure 1, we plot treatment effects on the y-axis against propensity to be present in the in-person exam data on the x-axis, by intervention arm and outcome. These figures show that those students least likely to return—particularly those in the lowest two percentiles—experienced positive effects of both accountability and tutoring on both the February and March in-person assessments. In short, the negative effects we observe four months after the intervention only generalize to children similar to the relatively high-achieving students who returned to in-person schooling by that point in time after pandemic-related shutdowns, and not to those lower-achieving students who had not yet returned. As a reminder, we do not see major signs of differential attrition from the treatment versus control group on the basis of observable characteristics, so the compositional differences in assessment samples implicate external generalizability but should not affect internal validity.

Additionally, as a reminder, the phone-based assessments had greater alignment to the content of the phone-based intervention than did the in-person assessments (76 percent versus between 50 and 57 percent overlap in the content covered). We do not have access to item-level data for the in-person assessments and therefore cannot test this directly, but it is possible that we would not have seen negative effects on the in-person assessment items designed to measure the concepts that were the focus on the phone calls.

Why Did Tutoring Reduce Math Achievement?

Why did phone calls, intended to provide students with individualized academic support, negatively impact student learning, at least among those relatively higher-achieving students who re-enrolled and were assessed after schools reopened? Survey results provide some suggestive evidence that the intervention affected time use in consequential ways. In Table 8 we report results from student and parent survey questions administered via the December 2020 PBA. We show that both arms of the treatment increased student agreement with the phrase “my teacher cares about me” by 0.38 on a scale of one to five for accountability and a similar 0.37 for tutoring (relative to a control group mean of 3.91). Therefore, we find no evidence for the hypothesis that phone calls harmed student perceptions of their teachers or were otherwise poorly received. Although this does not seem to have led to an impact on student enrollment by March 2021, it could potentially signal more positive benefits of the intervention down the road that we are unable to test with the currently available data.

When it came to parent reports of how many hours per week their child was studying at the conclusion of the intervention, the accountability calls increased reported time spent studying by 0.18 hours per week, based on results from the treatment-on-the-treated models accounting for covariates and reported in Table 8. In contrast, there is no evidence that the tutoring intervention increased time studying and suggestive evidence that it may have decreased it by 0.06 hours per week based on the negatively-signed, though insignificant, TOT estimate with controls reported in Table 8. Neither intervention arm had an effect on the reported use of educational television, radio shows, or internet. Accountability decreased the use of books at home to study. The estimate for tutoring's effect on the use of books at home is also negative but not statistically significant. Both intervention arms increased the use of Bridge@Home resources and increased the amount of time spent studying during calls from the teacher. In short, the phone-based mini-tutoring sessions increased the amount of time spent studying by phone while keeping constant the overall amount of time a student spent studying in an average week.

In Figure 2, we plot treatment effects on math scores along the y-axis against treatment effects on reported time spent studying along the x-axis for each outcome, separately for accountability and tutoring, for each randomization block. These figures show a positive relationship between treatment effects on time use and math achievement, suggesting that students who were induced to spend more time studying as a result of the intervention experienced the most positive gains in math while those for whom the intervention decreased time studying experienced the largest negative effects in math. This positive relationship is consistent for both accountability and tutoring though tutoring did not, on average, increase time spent studying. This pattern also holds across the PBA, February and March in-person assessments. In sum, it appears that tutoring, by increasing time spent studying during calls from the teacher without increasing the overall amount of time spent studying per week, may have caused students to substitute the phone-based tutoring for some other, more productive, form of studying such as working with parents or siblings at home or even studying independently.

In April 2021, we administered another phone-based survey to further explore hypothesized mechanisms. We find evidence consistent with a story in which tutoring decreased the amount of time parents (or other adults at home) spent helping children with learning. In Table 9, we show that, according to children, tutoring decreased time their parents helped with studying (between -0.07 and -0.08 on a scale with five answer choices where 1=almost never and

5=almost every day) although these differences are not statistically significant (likely due at least in part to the reduced sample size). The direction of the coefficients for accountability, in contrast, is positive. Similarly, parents from the tutoring group report spending less time helping their children with learning although again none of these effects achieve statistical significance and the magnitude of these effects is sensitive to specification. Interestingly, the accountability intervention appears to have decreased satisfaction with the learning support from school (coefficients are negative though not statistically significant) while increasing parental confidence in their child's progress despite the lack of impact on achievement. Although not definitive, in all these results are consistent with a story in which those parents receiving a less intensive accountability intervention continued supporting their children academically at home because they were less satisfied with the remote help from school and, possibly as a result, were confident in their children's progress because they more closely observed the learning process. We also note that the counterfactual is a relatively high level of child-reported parental learning help (4.16 on a scale of one to five with five representing "almost every day").

DISCUSSION

This field experimental study tested the relative efficacy of two teacher-student phone-call interventions focused on math—accountability checks and mini-tutoring sessions—for promoting academic achievement among students attending low-cost private primary schools in Kenya in the midst of pandemic-related school closures. Overall, we find no evidence that phone calls had an effect on achievement in the short-run phone-based assessments. Similarly, we find no evidence that the accountability checks impacted student math performance four months after the intervention based on in-person assessment results. However, phone-based tutoring appears to have decreased math achievement four months after the intervention, at least among those students who were assessed again in-person because they returned to a Bridge school after reopening. This appeared to be due, at least in part, to the fact that those relatively low-achieving students who were most likely to benefit from the intervention in the short-run were least likely to return to their schools after they reopened and therefore cannot contribute to our estimates of the treatment effects in the long-run. The intervention provided at least short-term benefits to those students who did not return to school. The magnitude of the impact on core numeracy for these children was roughly 17 percent of the typical gap between grade levels 3 and 5—a substantively meaningful improvement for a relatively light-touch intervention but not enough to make up for a full year's worth of lost learning time.

We provide suggestive evidence that the mechanism for how the tutoring intervention reduced math achievement among those students who did return to school was related to time use substitution. Those whose performance benefitted most from the interventions were the same students who seemed to be induced to increase their time studying as a result of the phone calls. Perhaps after engaging with their teacher's instruction by phone, students or their parents were more likely to feel that they had completed their educational activities for the day. In other words, they may have viewed the phone calls as supplanting rather than supplementing learning at home. In the case of Bridge Kenya families, it could be that studying independently or with a family member in-person would have been more productive than working briefly with a teacher by phone. However, we are not able to definitely speak to what children in our treatment groups would have been doing in the absence of the intervention. Additionally, findings may not generalize to contexts with demographically different parent populations. Although the Bridge Kenya is relatively disadvantaged from a global perspective, they are still quite different on observable and likely unobservable dimensions from public school parents given they have opted into Bridge schools. Our findings build on a body of evidence related to educational technology interventions in developing countries that do not achieve the intended improvements in learning – sometimes yielding negative effects – due to unforeseen changes in how the targeted population uses educational time as they engage with the technological tool (e.g., Berlinski & Busso, 2017; Meza-Cordero, 2017; Angrist & Lavy, 2002; Malamud & Pop-Eleches; 2011).

We hope that readers do not conclude, based on our study, that tutoring is ineffective or that it cannot be successfully implemented at a distance. Importantly, the tutoring intervention studied here was aligned in some ways with what researchers know about the features of high-impact tutoring—the intervention relied on full-time teachers, retained the same teacher-student matches throughout the intervention, had low (one-to-one) tutor-tutee ratios, and was aligned with classroom content. However, in other important ways the tutoring intervention diverged from the gold standards of high-impact tutoring—it occurred relatively infrequently (once per week versus the recommended three weekly session minimum), the sessions were relatively short (15 minutes versus the recommended 30-60-minute session length), and it occurred outside of school (while the highest-impact in-person tutoring interventions have occurred during school hours rather than after-school or summertime). In short, our study is not a test of the effectiveness of what researchers and practitioners call "high-dosage" tutoring.

Of course, most of the differences between the intervention we study and "gold standard" tutoring programs stemmed from logistical and financial needs, as well as the serious constraints policymakers faced when schools were closed due to a public health emergency. Additionally, there are major cost implications for implementing a high-dosage tutoring intervention at scale. For example, based on a back-of-the-envelope calculation, increasing the intensity to a 60-minute weekly phone-based lesson would cost \$12.00 per student, roughly 6 percent of the total annual governmental expenditure on primary education in Kenya (~$200.00 per pupil per year). Therefore, such a program is likely to be cost-prohibitive if implemented universally. Policymakers may be wise to consider programming implemented with greater intensity but to a more targeted sub-population most likely to benefit from such an intervention and to experience lost learning time due to educational disruptions. That said, this would require overcoming practical and political barriers, and designing programs to avoid stigmatizing participants (which may be more possible with remote versus in-person interventions). Another inexpensive program design feature that should be tested is an explicit framing alerting parents and students that the intervention is supplemental and not meant to replace existing home-based supports.

Another reason to be cautious when interpreting the null accountability impacts and the negative tutoring impacts, is that the February and March in-person assessments did not exclusively measure the standards that were the focus of the phone-based interventions but rather covered a broad set of mathematics skills and concepts. More specifically, our partner estimates that, across grades, there was 50 percent alignment between tutoring content and the February in-person exams and 57 percent alignment on the March in-person exams. Unfortunately, we are unable to examine item-level data, but it is possible that the effects of these interventions may vary depending on whether a standard was covered during the intervention. Our findings also suggest that, as students return to schools and to standardized assessments around the world, leaders should be cautious and avoid direct comparisons between their pre-shutdown student populations and the group of students who return to school without accounting for likely compositional shifts.

One puzzle is why our findings differ from the Angrist, Bergman & Matsheng (2021) study which found positive effects of phone-based tutoring in the context of Botswana on the order of 0.12 standard deviations in math achievement. The intervention was similar in that it consisted of short 15-20-minute phone calls focused on math instruction for students in grades three to five (while the intervention under study here targeted 3rd, 5th and 6th grades). However, the contexts in which the two interventions were implemented differed in important ways. The Kenyan intervention targeted students in private schools, whereas the Botswana program was mostly aimed at government schools, which not only affects the student characteristics but also the instruction and materials available during school closures for those students representing the counterfactual. Similarly, our sample is almost two years older on average than the Botswana group (11.6 vs. 9.7), adding to the possibility that the Kenyan sample may be more "selected" than the Botswana group. In fact, Angrist, Bergman & Matsheng (2021) report that on an Uwezo-like scale of math achievement between 0-4, their sample stands at 1.97, on average, compared to an equivalent measure for the Kenyan sample of 3.02 – a difference of 0.65 standard deviations (standardized on the Kenyan sample). In short, it is likely that the Botswana sample was more similar to the students in our Kenya sample who were least likely to return to school post-reopening and most likely to benefit, at least in the short-term, from our intervention.

Beyond context, the Botswana intervention also had a longer duration—12 weeks—versus seven weeks in our case. The intervention relied on volunteers from the not-for-profit organization Young1ove rather than full-time teachers as was the case in Kenya and it was implemented on a smaller scale. There was no need for a complex cascade training model in which details may be more likely to be lost in a "game of telephone." Furthermore, the Botswana intervention included an opt-in process in which the team confirmed phone numbers prior to randomization. In Kenya, not reach all pupils were reached at least in part because the team was working from administrative records that were less than 100 percent accurate due to turnover, number changes, etc. It is also unclear whether the control condition in Botswana was one in which students had access to the relatively rich set of distance learning materials provided to Bridge Kenya students. Finally, researchers have examined the impact of this intervention on short-term PBA outcomes. It is possible that their findings might be different if they studied effects on longer-term in-person exams.

Big picture, we conclude that leaders should not abandon tutoring but should be careful in developing interventions—in general, but especially in the aftermath of Covid school closures—to design programs that closely mirror the features of those programs that have previously demonstrated positive results. Additionally, leaders would be wise to target interventions to populations of students most likely to benefit and consider potential unintended consequences, although these can be difficult to predict from the outset particularly in the midst of demanding circumstances for educational providers such as a pandemic. A final implication of our study is the revealed importance of ongoing rigorous testing and evaluation of educational interventions that seem beneficial, on face, but that may have unintended consequences that would otherwise go undetected in the absence of research. This study therefore represents a rare opportunity to learn from some of the most challenging circumstances in our collective history.

REFERENCES

Abadie, Athey, Imbens and Wooldridge (2017). When should you adjust standard errors for clustering? Statistics Theory.

Aker, J. C., & Ksoll, C. (2019). Call me educated: Evidence from a mobile phone experiment in Niger. Economics of Education Review, 72, 239–257.

Andrabi, T., Daniels, B. & Das, J. (2020). Human capital accumulation and disasters: Evidence from the Pakistan earthquake of 2005. RISE Working Paper 20/039.

Angrist, N., Bergman, P. & Matsheng, M. (2021). School's out: Experimental evidence on limiting learning loss using "low-tech" in a pandemic. NBER Working Paper 28205.

Angrist, N., Bergman, P., Brewster, C. & Matsheng, M. (2020). Stemming learning loss during the pandemic. Center for the Study of African Economies Working Paper.

Angrist, Bergman, Evans, Hares, Jukes & Letsomo (2020). Practical lessons for phone-based assessments of learning. Center for Global Development Working Paper 534.

Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. The Economic Journal, 112(482), 735–765.

Bacher-Hicks, A., Goodman, J., & Mulhern, C. (Forthcoming). Inequality in Household Adaptation to Schooling Shocks. Journal of Public Economics.

Bandiera, O., Buehren, N., Goldstein, M., Rasul, I. & Smurra, A. (2019). The economic lives of young women in the time of Ebola. World Bank Group Working Paper 8760.

Banerjee, Cole, Duflo, Linden (2007) Remedying education: evidence from two randomized experiments in India. Quarterly Journal of Economics, 122(3) 1235–1264.

Banerjee, Banerjee, Berry, Duflo, Kannan, Mukerji, Shotland & Walton (2017). From proof of concept to scalable policies. Journal of Economic Perspectives, 31(4), 73-102.

Battistich, V., Solomon, D., Watson, M., & Schaps, E. (1997). Caring school communities. Educational Psychologist, *32*, 137–151.

Bianchi, N., Lu, Y., & Song, H. (2020) The effect of computer-assisted learning on students' long-term development. Working paper: https://ssrn.com/abstract=3309169.

Berlinski, S., Busso, M., Dinkelman, T., & Martinez, C. (2016). Reducing parent-school information gaps and improving education outcomes. Working Paper.

Berlinski & Busso (2017). Challenges in educational reform. Economics Letters, 156, 172–175.

Bergman, P. (2015). Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment. CESifo Working Paper Series No. 5391.

Bird, K., Castleman, B., Denning, J., Goodman, J., Lamberton, C., & Rosinger, K. O. (2019). Nudging at Scale. NBER Working Paper No. 26158.

Blad, E. (2020). Former Governor Recruits Stuck-at-Home College Students to Combat K-12's 'COVID Slide'. Education Week.

Blitzer, W. (2020). The Future Of Education Is Uncertain Amid Pandemic. The Situation Room.

Bosco et al. (2017): Exploring the high-resolution mapping of gender disaggregated development indicators. Journal of the Royal Society Interface, 14(129). DOI: 10.1098/rsif.2016.0825.

Bray, M. (2021). Shadow education in Africa: Private supplementary tutoring and its policy implications. Hong Kong: Comparative Education Research Centre.

Bridge (2018). Who we are. https://www.bridgeinternationalacademies.com/who-we-are/

Brooks, D. (2020). We Need National Service. Now. The New York Times.

Cabezas, Cuesta, & Gallego (2011). Effects of Short-Term Tutoring on Cognitive and Non-Cognitive Skills. Santiago, Chile: Abdul Latif Jameel Poverty Action Lab (J-PAL).

Carlana, M. & LaFerrara, E. (2021). Apart but connected: Online tutoring and student outcomes during the COVID-19 pandemic. Annenberg Institute EdWorkingPaper No. 21-350.

Castleman, B. L., & Meyer, K. (2016). Can Text Message Nudges Improve Academic Outcomes in College? EdPolicyWorks Working Paper No. 43.

Castleman & Page (2016). Freshman Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence. Journal of Human Resources, 51(2), 389–415.

Chabrier, J., Cohodes, S., & Oreopoulos, P. (2016). What can we learn from charter school lotteries? Journal of Economic Perspectives, 30(3), 57-84.

CNN (2021). Tracking Covid-19's global spread. https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/

Connell, J. P. (1990). Context, self, and action. In D. Cicchetti (Ed.), The self in transition: Infancy to childhood (pp. 61–97). Chicago, IL: University of Chicago Press.

Cook, Dodge, Farkas, Fryer, Guryan, Ludwig, Mayer, Pollack & Steinberg (2015). The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth.

Daily News (2021). UK Government set to give national tutoring programme to Randstad.

Dang, Rogers and Halsey (2008). The growing phenomenon of private tutoring. World Bank Research Observer, 23(2), 161-200.

Doss, C., Fahle, E. M., Loeb, S., & York, B. N. (2019). More Than Just a Nudge. Journal of Human Resources, 54(3), 567–603.

Duflo, Dupas & Kremer (2011). Peer effects, teacher incentives, and the impact of tracking. American Economic Review, 101, 1739-1774.

Dustan, A., Maldonado, S., & Hernandez-Agramonte, J.M. (2019). Motivating bureaucrats with non-monetary incentives when state capacity is weak. Working paper.

Dynarski, S. (2020). The School Year Really Ended in March. The New York Times.

Escueta, Nickow, Oreopoulos & Quan (2020). Upgrading education with technology: Insights from experimental research. Journal of Economic Literature, 58(4), 897-996.

Evans, Hares, Mendez Acosta & Saintis (2021) It's been a year since schools started to close due to Covid-19. Center for Global Development.

Finn, J., & Rock, D. (1997). Academic success among students at risk for school failure. Journal of Applied Psychology, *82*, 221–234.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, *74*, 59–109.

Fryer, R. (2014). Injecting Charter School Best Practices Into Traditional Public Schools: Evidence From Field Experiments. Quarterly Journal of Economics, 129, 1355-1407.

Fryer, R. (2016). The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments. NBER Working Paper #22130.

Fryer, R. (2016). The 'Pupil' Factory. NBER Working Paper No. 22205.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. Journal of Educational Psychology, *95*, 148–162.

Ganimian, A., Vegas, E. & Hess, F. (2020). Realizing the promise: How can education technology improve learning for all? Brookings Institution.

Gill, Lerner, Meosky (2017). Reimagining Accountability in K–12 Education. https://behavioralpolicy.org/articles/reimagining-accountability-in-k-12-education/

Gill, B. (2020). Using Transparency To Create Accountability When School Buildings Are Closed and Tests Are Canceled. Education Next.

Gill, Walsh, Wulsin, Matulewicz, Severn, Grau, Lee, and Kerwin (2015). Inside Online Charter Schools. Cambridge, MA: Mathematica Policy Research.

Harris, D. (2009). Toward Policy-Relevant Benchmarks for Interpreting Effect Sizes. Educational Evaluation and Policy Analysis, 31(1), p. 3-29.

Heinrich, C., Robert, M. and Whitten, G. (2010). Supplemental Education Services Under No Child Left Behind. Educational Evaluation and Policy Analysis, 32(2), 273-298.

Heinrich et al. (2014). Improving the Implementation and Effectiveness of Out-of-School-Time Tutoring. Journal of Policy Analysis and Management, 33(2), 471-494.

Hill, A. & Jones, D. (2018). A Teacher Who Knows Me: The Academic Benefits of Repeat Student-Teacher Matches. Economics of Education Review, 64, 1-12.

Hurwitz, L. B., Lauricella, A. R., Hanson, A., Raden, A., & Wartella, E. (2015). Supporting Head Start parents. Early Child Development and Care, 185(9), 1373–1389.

International Telecommunication Union (2018). Statistics. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx

Jayachandran, S. (2014). Incentives to teach badly: after-school tutoring in developing countries. Journal of Development Economics.

Jayachandran, S. (2014). Incentives to teach badly: After-school tutoring in developing countries. Journal of Development Economics.

Johnston, J., & Ksoll, C. (2017). Effectiveness of Interactive Satellite-Transmitted Instruction. Stanford CEPA Working Paper 17–08: https://eric.ed.gov/?id=ED579066

Jukes et al. (2017). Improving literacy instruction in Kenya through teacher professional development and text messages support. Journal of Research on Educational Effectiveness, 10(3), 449–481.

Kenya National Bureau of Statistics (2018). National ICT Survey Report. https://ca.go.ke/wp-content/uploads/2018/02/National-ICT-Survey.pdf

Kim, E., Goodman, J. & West, M. (2021). Kumon in: The recent, rapid rise of private tutoring centers. Brown University Annenberg EdWorkingPaper No. 21-367.

Kirksey, J. & Gottfried, M. (2018). Familiar faces: Can having similar classmates from last year link to better school this year attendance? The Elementary School Journal, 119(2).

Kraft, M. & Dougherty, S. (2013). The effect of teacher-family communication on student engagement. Journal of Research on Educational Effectiveness, 6(3), 199-222.

Kraft, M. & Goldstein, M. (2020). Getting tutoring right to reduce Covid-19 learning loss. Brookings Brown Center Chalkboard.

Kraft & Monti-Nussbaum (2017). Can schools enable parents to prevent summer learning loss? The ANNALS of the American Academy of Political and Social Science, 674(1), 85–112.

Ksoll, Aker, Miller, Perez & Smalley (2015). Learning without teachers? Evidence from a randomized experiment of a mobile phone-based adult education in Los Angeles.

Lerner & Tetlock, (1999). Accounting for The Effects of Accountability. Psychological Bulletin, 125(2):255-75

Malamud, Cueto, Cristia & Beuermann (2019). Do children benefit from internet access? Experimental evidence from Peru. Journal of Development Economics, 138, 41–56.

Mayer, S., Kalil, A., Oreopoulos, P., & Gallegos, S. (2015). Using Behavioral Insights to Increase Parental Engagement. NBER Working Paper No. 21602.

Mays, J. & Newman, A. (2020). Virus Is Twice as Deadly for Black and Latino People Than Whites in N.Y.C. The New York Times.

Mbiti, Muralidharan, Romero, Schipper, Manda & Rajani (2019). Inputs, Incentives, and Complementarities in Education. Quarterly Journal of Economics, 134(3), 1627-1673.

Meza-Cordero, J. (2017). Learn to play and play to learn: Evaluation of the one laptop per child program in Costa Rica. Journal of International Development, 29(1), 3–31.

Muhumuza, R. (2020). In African nations, it's doubly hard for kids to distance-learn. The Christian Science Monitor.

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

Musu, L. (2018). The Digital Divide: Differences in Home Internet Access. *NCES Blog*.

Naik, Chitre, Bhalla & Rajan (2020). Impact of use of technology on student learning outcomes: Evidence from a large-scale experiment in India. World Development, 127, 104736.

Nickow, Oreopoulos & Quan (2020). The impressive effects of tutoring on preK-12 learning: A systematic review and meta-analysis of the experimental evidence. NBER Working Paper No. 27476.

Oreopoulos, P. (2020). Scale up tutoring to combat COVID learning loss for disadvantaged students. Scientific American.

Oreopoulos, P., & Petronijevic, U. (2018). Student Coaching: How Far Can Technology Go? Journal of Human Resources, 53(2), 299–329.

Owsley, N. (2017). Getting the Message: Using Parental Text Messaging Learner Attendance. Master Thesis: University of Cape Town.

Parker, K., Horowitz, J. M., & Brown, A. (2020). About Half of Lower-Income Americans Report Household Job or Wage Loss Due to COVID-19. Pew Research Center.

Robinson, Kraft, Loeb & Schueler (2021). Accelerating student learning with high-dosage tutoring. EdResearch for Recovery Brief. Brown University Annenberg Institute.

Roschelle, Cheng, Hodkowski, Neisler & Haldar (2020). Evaluation of an online tutoring program in elementary mathematics. https://files.eric.ed.gov/fulltext/ED604743.pdf

Rodriguez-Segura, D. (2021). EdTech in developing countries: A systematic review. Forthcoming at the World Bank Research Observer.

Rodriguez-Segura, D. & Schueler, B. (2021). Can we measure learning over the phone? Evidence from Kenya. Working Paper.

Romero, Chen & Magari (2021). Cross-age tutoring: Experimental evidence from Kenya. Economic Development and Cultural Change. doi: https://doi.org/10.1086/713940

Sacerdote, B. (2012). When the saints go marching out. American Economic Journal: Applied Economics, 4(1), 109-135.

Santos & Novelli (2017). The effect of the Ebola crisis on the education system's contribution to post-conflict sustainable peacebuilding in Liberia. Research Consortium on Education & Peacebuilding.

Schueler, B. (2018). Making the Most of School Vacation: A Field Experiment of Small Group Math Instruction. Education Finance and Policy*, 15(2)*, 310-331.

Shores, K. & Steinberg, M. (2017). The impact of the great recession on student achievement: Evidence from population data.

Song, Y., Loewenstein, G., & Shi, Y. (2018). Heterogeneous effects of peer tutoring: evidence from rural Chinese middle schools. Research in Economics.

Twaweza. (2014). Uwezo data-Household data. https://www.twaweza.org/go/uwezodatasets

UNESCO (2021). Education: From disruption to recovery. https://en.unesco.org/covid19/educationresponse

U.S. DOE (2021). ED Covid-10 Handbook: Roadmap to reopening safely and meeting all students' needs Volume 2. https://www2.ed.gov/documents/coronavirus/reopening-2.pdf

Uwezo (2020). Are our children learning? The status of remote-learning among school-going children in Kenya during the Covid-19 crisis.

Vakis, R., & Farfan, G. (2018). Envío de mensajes de texto para incrementar la motivación y satisfacción docente. Policy report: Evidencias MineduLAB No. 04.

Wang & Holcombe (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. American Educational Research Journal, 47, 633–662.

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. Journal of Educational Psychology, *90*, 202–209.

Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of parents, teachers, and peers. Journal of Educational Psychology, *90*, 202–209.

World Bank (2020). The COVID-19 andemic: Shocks to Education and Policy Responses. Washington, DC.: https://openknowledge.worldbank.org/handle/10986/33696

World Bank (2017). Individuals using the internet (% of population). https://data.worldbank.org/indicator/IT.NET.USER.ZS

Table 1. Characteristics of the Control and Treatment Arm Groups, February 2020 Roster and November 2020 Call List

	Control	Accountability Treatment				Tutoring Treatment				
	Overall	Overall	Account. - Control	Compliers	Compliers - Non-Compliers	Overall	Tutoring - Control	Compliers	Compliers - Non-Compliers	Tutoring - Account. Diff
	Mean (SD)	Mean (SD)	Diff.	Mean (SD)	Diff.	Mean (SD)	Diff.	Mean (SD)	Diff.	Diff.
Female	0.50	0.49	0.00	0.49	-0.06**	0.48	-0.02	0.49	0.03	0.00
	(0.50)	(0.50)	(0.01)	(0.50)	(0.03)	(0.50)	(0.01)	(0.50)	(0.02)	(0.01)
Age	11.61	11.61	-0.02	11.58	-0.21**	11.50	-0.10	11.50	-0.03	-0.10**
	(1.69)	(1.64)	(0.07)	(1.64)	(0.10)	(1.66)	(0.08)	(1.66)	(0.12)	(0.05)
Placement Score	15.37	15.90	0.21	15.68	-1.67**	14.95	-0.23	14.82	-1.35	-0.79
	(10.45)	(10.55)	(0.51)	(10.61)	(0.62)	(10.94)	(0.63)	(10.96)	(0.82)	(0.61)
Baseline Math Score	-0.03	-0.06	-0.04	-0.03	0.10	0.03	0.03	0.03	0.01	0.06
	(0.97)	(0.96)	(0.05)	(0.96)	(0.08)	(1.00)	(0.06)	(1.00)	(0.08)	(0.05)
Baseline Kiswahili Score	0.00	-0.01	-0.02	0.00	0.16	-0.03	-0.07	-0.02	0.15*	-0.07
	(1.00)	(1.03)	(0.06)	(1.01)	(0.13)	(1.02)	(0.06)	(1.02)	(0.09)	(0.06)
Baseline English Score	-0.01	-0.11	-0.10*	-0.09	0.13	-0.04	-0.06	-0.04	-0.05	0.05
	(1.01)	(0.98)	(0.05)	(0.97)	(0.08)	(0.98)	(0.05)	(0.97)	(0.09)	(0.05)
Community Size	275,416	241,016	-58,000	247,674	90,111**	205,433	-70,000	210,864	33,535	1,550
	(441,116)	(285,273)	(59,737)	(296,218)	(44,911)	(312,979)	(66,092)	(325,616)	(27,494)	(52,095)
Female Literacy	0.82	0.86	0.02	0.86	0.01	0.87	0.05	0.87	0.01	0.03
	(0.18)	(0.13)	(0.03)	(0.13)	(0.01)	(0.09)	(0.03)	(0.09)	(0.01)	(0.02)
Poverty	0.34	0.33	0.01	0.33	-0.03**	0.35	0.02	0.35	-0.01	-0.01
	(0.19)	(0.15)	(0.02)	(0.16)	(0.01)	(0.17)	(0.03)	(0.17)	(0.01)	(0.02)
Cell Tower Distance	0.35	0.36	0.06	0.35	-0.02	0.57	0.27	0.60	0.05	0.16
	(0.27)	(0.69)	(0.10)	(0.66)	(0.02)	(1.14)	(0.19)	(1.19)	(0.09)	(0.19)
School Performance	-0.03	-0.01	0.00	0.01	-0.01	0.06	0.03	0.06	0.01	0.02
	(0.49)	(0.34)	(0.04)	(0.33)	(0.01)	(0.43)	(0.04)	(0.44)	(0.01)	(0.04)
Manager Attendance	0.88	0.85	-0.03	0.85	0.03	0.92	0.04	0.92	0.04	0.08
	(0.27)	(0.31)	(0.07)	(0.30)	(0.04)	(0.21)	(0.06)	(0.20)	(0.03)	(0.05)
Teacher Attendance	0.84	0.86	0.01	0.86	0.00	0.82	-0.02	0.81	-0.01	-0.04
	(0.15)	(0.14)	(0.03)	(0.14)	(0.01)	(0.17)	(0.04)	(0.17)	(0.02)	(0.03)
Pupil Attendance	0.49	0.54	0.04	0.54	0.01	0.47	-0.01	0.48	0.01	-0.06*
	(0.15)	(0.14)	(0.03)	(0.14)	(0.01)	(0.14)	(0.03)	(0.15)	(0.02)	(0.03)
Enrollment	275.51	269.77	-8.98	266.38	-14.83*	262.98	-11.44	266.11	20.69***	-2.82
	(58.47)	(78.31)	(16.32)	(76.96)	(8.15)	(79.14)	(14.63)	(80.52)	(6.1)	(18.98)
N of schools	35	35	-	35	-	35	-	35	-	-
N of students	2,847	2,779	-	2,434	-	2,693	-	2,392	-	-

Table 2. Characteristics of Each Assessment Sample Relative to Original Experimental Sample

Sample	(1) Roster February 2020	(2) PBA December 2020		(3) In-Person Exam February 2021		(4) In-Person Exam March 2021		(5) Phone Survey April 2021	
	Mean (SD)	Mean (SD)	Difference	Mean (SD)	Difference	Mean (SD)	Difference	Mean (SD)	Difference
Female	0.49	0.50	0.01	0.49	0.00	0.49	0.00	0.51	0.01
	(0.50)	(0.50)	(0.01)	(0.50)	(0.01)	(0.50)	(0.01)	(0.50)	(0.02)
Age	11.58	11.53	-0.06	11.52	-0.17***	11.53	-0.13***	9.95	-0.06
	(1.66)	(1.66)	(0.04)	(1.66)	(0.05)	(1.67)	(0.04)	(0.98)	(0.05)
Placement Score	15.41	15.05	-0.52*	14.66	-2.35***	14.51	-2.70***	13.40	-3.90***
	(10.65)	(10.75)	(0.29)	(10.93)	(0.32)	(10.94)	(0.31)	(11.98)	(0.50)
Baseline Math Score	-0.02	0.02	0.05**	0.03	0.15***	0.04	0.18***	0.08	0.19***
	(0.98)	(0.98)	(0.02)	(0.99)	(0.03)	(0.98)	(0.03)	(0.97)	(0.04)
Baseline Kiswahili Score	-0.01	0.02	0.06**	0.02	0.12***	0.03	0.15***	0.13	0.25***
	(1.02)	(1.01)	(0.03)	(1.00)	(0.04)	(1.00)	(0.03)	(0.97)	(0.05)
Baseline English Score	-0.05	0.00	0.07***	-0.01	0.15***	0.00	0.16***	0.04	0.29***
	(0.99)	(0.98)	(0.02)	(0.97)	(0.04)	(0.97)	(0.03)	(1.00)	(0.05)
Community Size	241,270	261,753	29,547***	256,607	48,074**	254,251	38,677**	251,538	27,904*
	(355,349)	(376,719)	(10,253)	(375,839)	(18,865)	(374,600)	(15,685)	(384,104)	(16,912)
Female Literacy	0.85	0.85	0.01**	0.85	0.00	0.85	0.00	0.84	-0.01*
	(0.14)	(0.14)	(0.00)	(0.14)	(0.01)	(0.14)	(0.01)	(0.14)	(0.01)
Poverty	0.34	0.33	-0.01***	0.34	0.00	0.34	0.00	0.35	0.00
	(0.17)	(0.17)	(0.00)	(0.18)	(0.01)	(0.18)	(0.01)	(0.18)	(0.01)
Cell Tower Distance	0.43	0.42	0.00	0.41	-0.06	0.42	-0.01	0.44	0.00
	(0.78)	(0.8)	(0.01)	(0.73)	(0.08)	(0.76)	(0.05)	(0.80)	(0.05)
School Performance	0.01	0.01	0.00	0.02	0.05	0.01	0.02	0.03	0.04*
	(0.43)	(0.41)	(0.01)	(0.42)	(0.03)	(0.42)	(0.02)	(0.43)	(0.02)
Manager Attendance	0.88	0.89	0.01	0.89	0.02	0.88	0.00	0.88	0.00
	(0.27)	(0.26)	(0.01)	(0.26)	(0.02)	(0.26)	(0.01)	(0.27)	(0.01)
Teacher Attendance	0.84	0.84	0.01***	0.84	0.02	0.84	0.01	0.84	0.00
	(0.15)	(0.15)	(0.00)	(0.15)	(0.01)	(0.15)	(0.01)	(0.15)	(0.01)
Pupil Attendance	0.50	0.51	0.01***	0.51	0.03***	0.51	0.03***	0.50	0.01*
	(0.15)	(0.15)	(0.00)	(0.15)	(0.01)	(0.15)	(0.01)	(0.15)	(0.01)
Enrollment	269.54	273.07	5.10***	271.86	7.27**	270.72	3.52	271.13	6.37*
	(72.6)	(73.05)	(1.44)	(72.43)	(3.74)	(71.42)	(3.11)	(72.99)	(3.36)
N of schools	105	105		105		105		105	
N of students	8,319	2,552		5,665		5,527		2,066	3,012

Table 3. Characteristics of Control and Treatment Arm Groups, by Assessment Sample

Sample	PBA December 2020				In-Person Exam February 2021				In-Person Exam March 2021				Phone Survey April 2021			
	Control	Account. - Control	Tutoring - Control	Tutoring - Account. Diff	Control	Account. - Control	Tutoring - Control	Tutoring - Account. Diff	Control	Account. - Control	Tutoring - Control	Tutoring - Account. Diff	Control	Account. - Control	Tutoring - Control	Tutoring - Account. Diff
	Mean (SD)	Diff.	Diff.	Diff.	Mean (SD)	Diff.	Diff.	Diff.	Mean (SD)	Diff.	Diff.	Diff.	Mean (SD)	Diff.	Diff.	Diff.
Female	0.51	0.00	-0.02	-0.01	0.51	-0.02	-0.01	0.01	0.50	-0.02	-0.01	0.01	0.51	-0.01	0.00	0.01
	(0.50)	(0.02)	(0.02)	(0.02)	(0.50)	(0.02)	(0.01)	(0.02)	(0.50)	(0.02)	(0.02)	(0.02)	(0.50)	(0.03)	(0.02)	(0.03)
Age	11.53	0.01	-0.03	-0.03	11.55	0.02	-0.10	-0.13*	11.57	0.00	-0.10	-0.10	-0.10	-0.02	-0.05	-0.03
	(1.63)	(0.09)	(0.09)	(0.09)	(1.69)	(0.10)	(0.10)	(0.07)	(1.70)	(0.09)	(0.09)	(0.06)	(0.06)	(0.08)	(0.09)	(0.06)
Placement Score	14.85	0.53	-0.20	-0.72	14.70	0.27	-0.35	-0.55	14.51	0.36	-0.42	-0.76	-0.76	0.99	-0.17	-1.05
	(10.69)	(0.72)	(0.80)	(0.73)	(10.74)	(0.67)	(0.72)	(0.62)	(10.79)	(0.67)	(0.71)	(0.62)	(0.62)	(0.94)	(1.10)	(0.91)
Baseline Math Score	0.01	0.01	-0.02	-0.03	0.04	-0.06	0.00	0.06	0.06	-0.08	-0.03	0.05	0.05	-0.07	0.05	0.12
	(0.98)	(0.06)	(0.07)	(0.06)	(0.98)	(0.06)	(0.07)	(0.05)	(0.98)	(0.05)	(0.06)	(0.05)	(0.05)	(0.11)	(0.10)	(0.11)
Baseline Kiswahili Score	0.04	0.04	-0.10	-0.14**	0.03	-0.01	-0.05	-0.04	0.04	-0.01	-0.07	-0.06	-0.06	0.03	0.07	0.05
	(0.99)	(0.07)	(0.08)	(0.06)	(0.98)	(0.07)	(0.06)	(0.06)	(0.98)	(0.07)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.09)
Baseline English Score	0.03	-0.04	-0.07	-0.01	0.05	-0.12*	-0.09	0.05	0.05	-0.11*	-0.10*	0.04	0.04	-0.11	-0.05	0.08
	(0.98)	(0.06)	(0.06)	(0.06)	(0.99)	(0.06)	(0.06)	(0.05)	(0.98)	(0.06)	(0.06)	(0.05)	(0.05)	(0.10)	(0.09)	(0.10)
Community Size	304,846	-85,000	-75,000	13,027	292,707	-77,000	-84,000	-1,254	294,548	-79,000	-84,000	813	813	-85,000	-61,000	27,402
	(471,424)	(67,407)	(72,573)	(59,124)	(462,380)	(67,394)	(70,164)	(56,919)	(463,765)	(66,204)	(70,462)	(56,038)	(56,038)	(65,238)	(70,074)	(60,673)
Female Literacy	0.82	0.02	0.05	0.03	0.81	0.04	0.05*	0.02	0.82	0.02	0.04	0.02	0.02	0.02	0.05	0.03
	(0.17)	(0.03)	(0.03)	(0.02)	(0.18)	(0.03)	(0.03)	(0.02)	(0.18)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Poverty	0.33	0.02	0.02	0.00	0.35	0.00	0.01	0.00	0.34	0.01	0.01	0.00	0.00	0.02	0.02	0.00
	(0.19)	(0.03)	(0.03)	(0.03)	(0.20)	(0.03)	(0.03)	(0.02)	(0.20)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Cell Tower Distance	0.34	0.08	0.28	0.18	0.36	0.01	0.16	0.14	0.35	0.05	0.21	0.15	0.15	0.10	0.25	0.13
	(0.27)	(0.13)	(0.2)	(0.21)	(0.29)	(0.13)	(0.15)	(0.20)	(0.28)	(0.13)	(0.17)	(0.20)	(0.20)	(0.14)	(0.19)	(0.21)
School Performance	-0.04	0.02	0.04	0.02	0.00	0.00	0.02	0.02	-0.02	0.01	0.02	0.02	0.02	0.01	0.03	0.02
	(0.47)	(0.04)	(0.04)	(0.04)	(0.47)	(0.04)	(0.04)	(0.04)	(0.48)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Manager Attendance	0.89	-0.04	0.02	0.07	0.87	-0.01	0.05	0.06	0.88	-0.04	0.05	0.09	0.09	-0.05	0.03	0.08
	(0.25)	(0.07)	(0.06)	(0.05)	(0.28)	(0.07)	(0.06)	(0.05)	(0.27)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)	(0.06)	(0.06)
Teacher Attendance	0.84	0.00	-0.03	-0.04	0.84	0.01	-0.03	-0.05*	0.84	0.00	-0.03	-0.04	-0.04	-0.01	-0.03	-0.03
	(0.14)	(0.03)	(0.04)	(0.03)	(0.15)	(0.03)	(0.04)	(0.03)	(0.14)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
Pupil Attendance	0.50	0.04	-0.01	-0.06*	0.49	0.05	-0.01	-0.06**	0.50	0.04	-0.02	-0.06**	-0.06	0.04	-0.01	-0.05*
	(0.15)	(0.03)	(0.03)	(0.03)	(0.15)	(0.03)	(0.03)	(0.03)	(0.15)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Enrollment	279.62	-5.52	-14.46	-6.45	277.76	-6.44	-11.60	-3.03	276.48	-8.77	-11.71	-0.73	-0.73	-11.48	-15.95	-2.65
	(58.88)	(16.67)	(17.07)	(18.35)	(58.65)	(16.24)	(16.39)	(17.41)	(57.79)	(16.05)	(16.09)	(17.78)	(17.78)	(17.18)	(17.73)	(19.39)
N of schools	35	105	105	-	35	105	105	-	35	105	105	-	35	105	105	-
N of students	872	2,552	2,552	-	1,962	5,665	5,665	-	1,905	5,527	5,527	-	3,622	2,066	2,066	-

Table 4. The Effect of Teacher Phone Calls on Math Achievement, Intent-to-Treat Estimates

	(1)				(2)			(3)		
	Control Mean (SD)	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff
					PBA December 2020					
Overall Math Score	-0.01	0.04	-0.03	-0.07	0.05	-0.02	-0.06	0.04	-0.04	-0.07
	(0.99)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)
Numeracy	0.01	0.06	-0.02	-0.08*	0.06	-0.02	-0.07	0.07	-0.02	-0.09*
	(0.98)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Curriculum-aligned	0.01	0.07	-0.02	-0.08*	0.07	-0.02	-0.07	0.07	-0.02	-0.09*
	(0.98)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Ability to do addition	0.92	0.03**	0.00	-0.03**	0.03**	0.00	-0.02*	0.03**	0.00	-0.03**
	(0.27)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ability to do word problem	0.26	-0.01	-0.04**	-0.02	-0.01	-0.04**	-0.03*	-0.01	-0.04**	-0.02
	(0.44)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N of schools	35	105	35	35	105	35	35	105	35	35
N of students	872	2,552	2,552	1,680	2,551	2,551	1,680	2,552	2,552	1,680
					In-Person Exam February 2021					
Overall Math Score	0.00	-0.12*	-0.11	0.01	-0.08	-0.10	-0.02	-0.12*	-0.11	0.01
	(1.00)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)	(0.07)	(0.07)	(0.06)
N of schools	35	105	35	35	105	35	35	105	35	35
N of students	1,644	4,835	4,835	3,191	4,831	4,831	3,190	4,835	4,835	3,191
					In-Person Exam March 2021					
Overall Math Score	0.00	-0.11	-0.16**	-0.05	-0.06	-0.14*	-0.07	-0.10	-0.15**	-0.05
	(1.00)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)	(0.06)	(0.07)	(0.08)	(0.07)
N of schools	35	105	35	35	105	35	35	105	35	35
N of students	1,472	4,514	4,514	3,042	4,511	4,511	3,041	4,504	4,504	3,033
Unconstrained			x							
Covariates						x				
Weights to match full sample									x	

Table 5. The Effect of Teacher Phone Calls on Math Achievement, Treatment-on-the-Treated Estimates

	(1)				(2)			(3)		
	Control Mean (SD)	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff
	PBA December 2020									
Overall Math Score	-0.01	0.04	-0.04	-0.08	0.05	-0.03	-0.08	0.04	-0.04	-0.08
	(0.99)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Numeracy	0.01	0.07	-0.03	-0.10	0.07	-0.04	-0.11*	0.07	-0.03	-0.10
	(0.98)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Curriculum-aligned	0.01	0.07	-0.03	-0.10	0.07	-0.04	-0.11*	0.07	-0.03	-0.10
	(0.98)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Ability to do addition	0.92	0.03**	0.00	-0.03**	0.03**	0.00	-0.03**	0.03**	0.00	-0.03**
	(0.27)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ability to do word problem	0.26	-0.01	-0.04**	-0.03	-0.02	-0.05***	-0.03	-0.01	-0.04**	-0.03
	(0.44)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N of schools	35	70	70	35	70	70	35	70	70	35
N of students	872	1,775	1,649		1,774	1,648		1,775	1,649	
	In-Person Exam February 2021									
Overall Math Score	0.00	-0.11*	-0.11	0.00	-0.06	-0.10	-0.04	-0.10	-0.10	0.00
	(1.00)	(0.06)	(0.07)	(0.06)	(0.05)	(0.07)	(0.06)	(0.06)	(0.07)	(0.06)
N of schools	35	70	70	35	70	70	35	70	70	35
N of students	1,644	3,252	3,227		3,249	3,223		3,252	3,227	
	In-Person Exam March 2021									
Overall Math Score	0.00	-0.11	-0.17**	-0.06	-0.03	-0.17**	-0.14	-0.10	-0.15**	-0.05
	(1.00)	(0.07)	(0.07)	(0.07)	(0.06)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
N of schools	35	70	70	35	70	70	35	70	70	35
N of students	1,472	3,071	2,915		3,069	2,912		3,064	2,911	
Unconstrained			x							
Covariates						x				
Weights to match full sample									x	

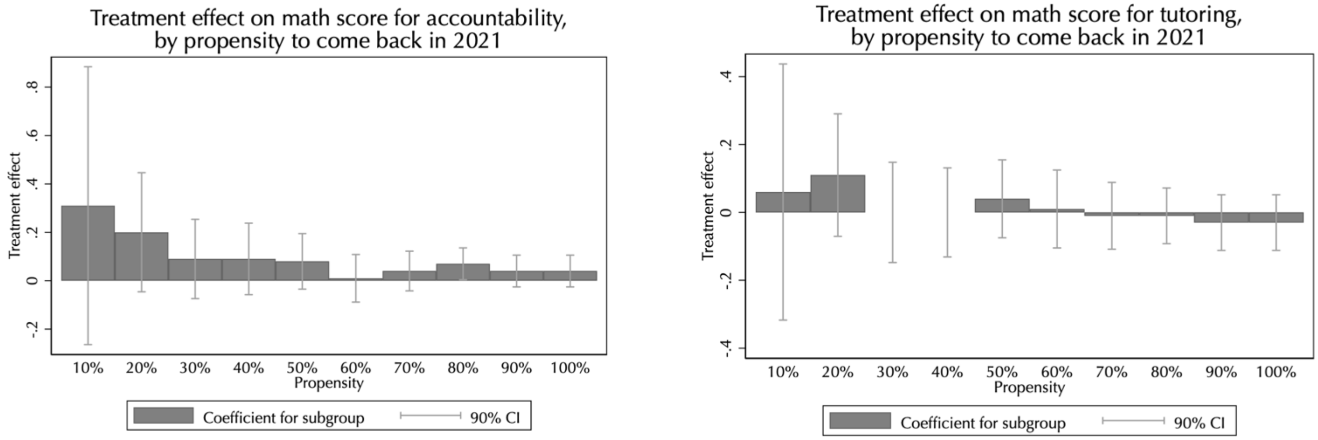
Table 6. The Effect of Teacher Phone Calls on Enrollment and on Missing a Value for an Outcome

	Control Overall	Full Sample December 2020		PBA December 2020		In-Person Exam February 2021		In-Person Exam March 2021		Phone Survey April 2021	
	Intent-to-Treat										
	Mean (SD)	Account.	Tutoring	Account.	Tutoring	Account.	Tutoring	Account.	Tutoring	Account.	Tutoring
Enrollment	0.32	0.01	0.01	0.00	0.02	-0.01	0.01	0.00	0.00	0.00	0.00
	(0.47)	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Missing	-	-	-	0.02	-0.02	-0.02	0.00	0.00	-0.01	0.00	-0.01
				(0.01)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
	Treatment-on-the-Treated										
	Mean (SD)	Account.	Tutoring	Account.	Tutoring	Account.	Tutoring	Account.	Tutoring	Account.	Tutoring
Enrollment	0.32	0.01	0.02	0.00	0.02	-0.01	0.01	0.00	0.00	0.00	0.00
	(0.47)	(0.03)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Missing	-	-	-	0.02	-0.02	-0.03	0.00	-0.01	-0.01	-0.01	-0.01
				(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
N of students		8,319		8,319		8,319		8,319		3,012	

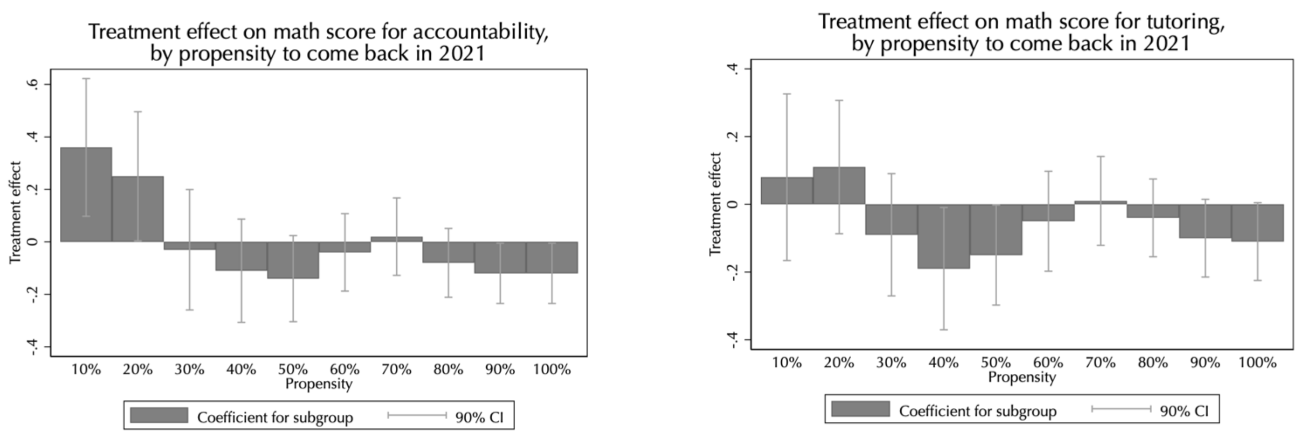
Table 7. The Effect of Teacher Phone Calls by Assessment Sample, Intent-to-Treat Estimates

	March In-Person Exam Sample		Not in March In-Person Exam Sample	
	Account.	Tutoring	Account.	Tutoring
Overall Math PBA Score	-0.01	-0.09*	0.17	0.14
	(0.04)	(0.5)	(0.10)	(0.11)
Hours studying per week	0.01	-0.05	0.42*	0.24
	(0.09)	(0.11)	(0.20)	(0.21)
N of students	1,936		616	

PBA (December 2020)



In-Person Exam (February 2021)



In-Person Exam (March 2021)

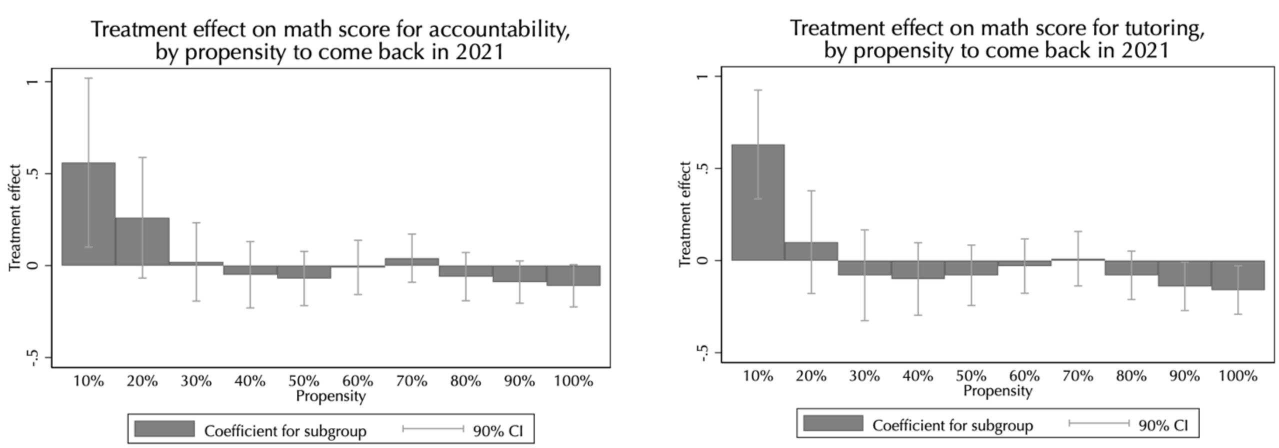
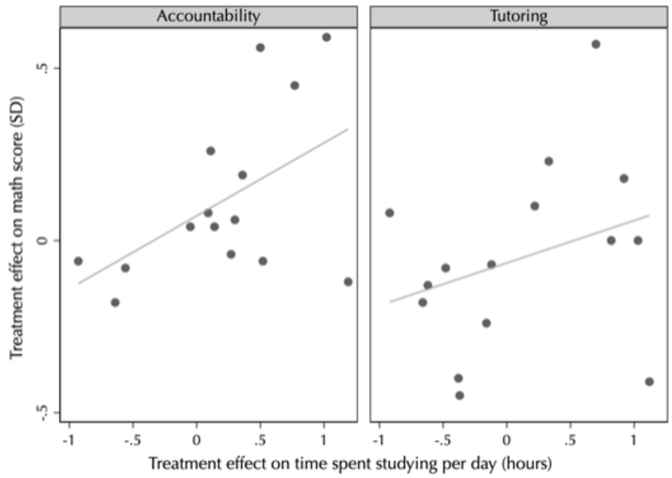


Figure 1. Treatment effects by propensity to be in the March 2021 in-person assessment sample.

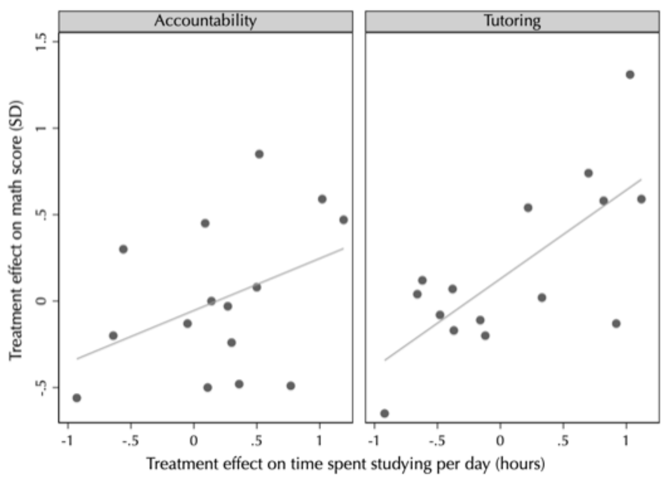
Table 8. The Effect of Teacher Phone Calls on on December 2020 Survey Outcomes, Treatment-on-the-Treated Estimates

				(1)			(2)			(3)		
	Control Mean (SD)	Min.	Max.	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff
My teacher cares about me	3.91	1	5	0.41***	0.38***	-0.03	0.41***	0.37***	-0.04	0.41***	0.37***	-0.04
	(1.51)			(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)
Hours studying per week	2.60	0	8	0.15*	0.00	-0.15	0.18**	-0.06	-0.24**	0.15*	0.00	-0.15
	(2.05)			(0.08)	(0.11)	(0.09)	(0.07)	(0.1)	(0.08)	(0.08)	(0.11)	(0.09)
Calls received from teacher	2.71	0	9	1.25***	1.32***	0.07	1.18***	1.31***	0.13	1.25***	1.3***	0.05
	(2.71)			(0.12)	(0.14)	(0.13)	(0.11)	(0.16)	(0.13)	(0.12)	(0.14)	(0.13)
Used Bridge@Home	0.62	0	1	0.10***	0.09***	-0.01	0.10***	0.08***	-0.02	0.1***	0.09***	-0.01
	(0.49)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Used educational TV/radio shows, internet	0.21	0	1	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.00
	(0.41)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Used books at home to study	0.64	0	1	-0.04**	-0.02	0.02	-0.04**	-0.02	0.02	-0.04**	-0.02	0.02
	(0.48)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Studied during calls from teacher	0.28	0	1	0.09***	0.07***	-0.02	0.08***	0.07***	-0.01	0.09***	0.07***	-0.02
	(0.45)				(0.02)			(0.02)			(0.02)	
N of schools					70			70			70	
N of students					1,775			1,774			1,775	
Unconstrained					x							
Covariates								x				
Weights to match full sample											x	

PBA (December 2020)



In-Person Exam (February 2021)



In-Person Exam (March 2021)

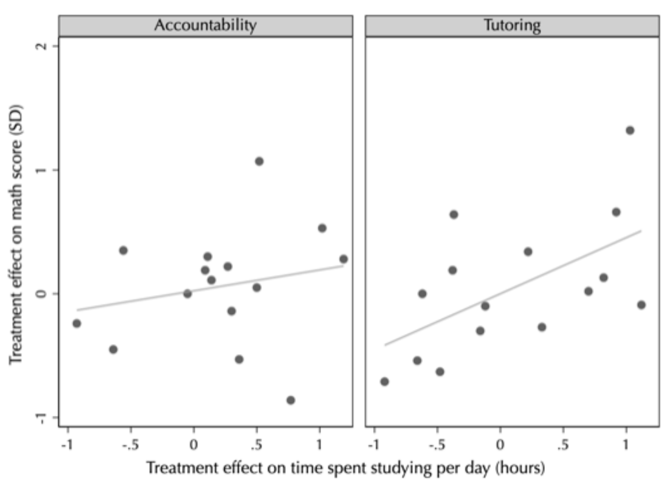


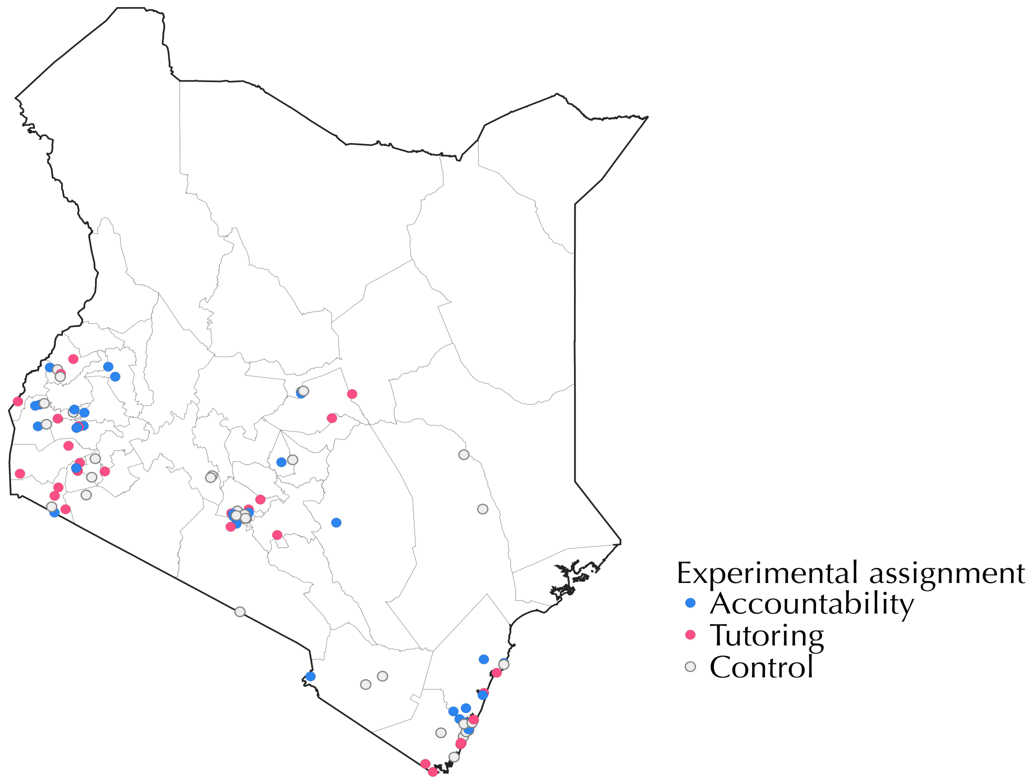
Figure 2. Treatment effects on time spent studying (x-axis) by treatment effects on math (y-axis)

Table 9. The Effect of Teacher Phone Calls on April 2021 Survey Questions Related to Remote Learning Period, Treatment-on-the-Treated Estimates

				(1)			(2)			(3)		
	Control Mean (SD)	Min.	Max.	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff
How often my parent helped with studying	4.16	1	5	0.02	-0.07	-0.09	0.09	-0.07	-0.16	0.06	-0.08	-0.14
	(1.24)			(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Adequacy of learning support from school	3.16	1	5	-0.04	0.01	0.05	-0.02	0.05	0.07	-0.05	0.02	0.07
	(1.08)			(0.1)	(0.09)	(0.09)	(0.1)	(0.08)	(0.09)	(0.1)	(0.09)	(0.09)
Hours I (parent) spent helping child with learning	5.19	0	26	-0.14	-0.37	-0.23	0.03	-0.06	-0.09	-0.1	0.00	-0.56
	(8.84)			(0.47)	(0.41)	(0.44)	(0.46)	(0.42)	(0.44)	(0.87)	(0.72)	(0.79)
Confidence in child's academic progress	3.74	1	5	0.1	0.01	-0.09	0.15*	0.01	-0.14	0.11	0.02	-0.09
	(0.88)			(0.08)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)
N of schools	35			70		70	35		70	70		70
N of students	265			512		506	496		491	495		490
Unconstrained					x							
Covariates								x				
Weights to match full sample											x	

Note: This sample includes families with third grade students only.

Appendix



Appendix Figure A1. Map of schools by experimental assignment.

Appendix Table A1. The Effect of Teacher Phone Calls on December 2020 Survey Outcomes, Intent-to-Treat Estimates

				(1)			(2)			(3)		
	Control Mean (SD)	Min.	Max.	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff
My teacher cares about me	3.91	1	5	0.38***	0.37***	0.00	0.38***	0.37***	0.01	0.37***	0.36***	0.00
	(1.51)			(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
Hours studying per week	2.60	0	8	0.13	0.03	-0.09	0.12	0.02	-0.09	0.13	0.02	-0.09
	(2.05)			(0.08)	(0.10)	(0.10)	(0.08)	(0.1)	(0.10)	(0.09)	(0.11)	(0.1)
Calls received from teacher	2.71	0	9	1.14***	1.22***	0.07	1.12***	1.19***	0.03	1.14***	1.21***	0.05
	(2.71)			(0.12)	(0.14)	(0.13)	(0.11)	(0.14)	(0.13)	(0.12)	(0.14)	(0.13)
Used Bridge@Home	0.62	0	1	0.09***	0.08***	-0.01	0.09***	0.08***	-0.02	0.09***	0.08***	-0.01
	(0.49)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Used educational TV/radio shows, internet	0.21	0	1	0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.01
	(0.41)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Used books at home to study	0.64	0	1	-0.04**	-0.02	0.02	-0.04**	-0.02	0.01	-0.04**	-0.02	0.02
	(0.48)			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Studied during calls from teacher	0.28	0	1	0.08***	0.06***	-0.02	0.08***	0.07***	-0.02	0.08***	0.06***	-0.02
	(0.45)				(0.02)			(0.02)			(0.02)	
N of schools					105			105			105	
N of students					2,552			2,551			2,552	
Unconstrained					x							
Covariates								x				
Weights to match full sample											x	

Appendix Table A2. The Effect of Teacher Phone Calls on April 2021 Survey Questions Related to Remote Learning Period, Intent-to-Treat Estimates

				(1)			(2)			(3)		
	Control Mean (SD)	Min.	Max.	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff	Account.	Tutoring	Tutoring - Account. Diff
How often my parent helped with studying	4.16	1	5	0.04	-0.09	-0.1	0.08	-0.11	-0.18*	0.07	-0.09	-0.13
	(1.24)			(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)
Adequacy of learning support from school	3.16	1	5	-0.07	0.03	0.1	-0.04	0.05	0.08	-0.08	0.04	0.13
	(1.08)			(0.1)	(0.09)	(0.09)	(0.09)	(0.09)	(0.1)	(0.1)	(0.09)	(0.1)
Hours I (parent) spent helping child with learning	5.19	0	26	-0.14	-0.36	-0.04	-0.06	-0.15	0.13	-0.15	0.00	-0.15
	(8.84)			(0.47)	(0.42)	(0.45)	(0.48)	(0.46)	(0.48)	(0.84)	(0.74)	(0.86)
Confidence in child's academic progress	3.74	1	5	0.10	0.01	-0.11*	0.11	0.01	-0.12*	0.1	0.01	-0.10
	(0.88)			(0.08)	(0.07)	(0.06)	(0.08)	(0.08)	(0.06)	(0.08)	(0.07)	(0.07)
N of schools	35			105		70	35		70	35		70
N of students	265			771		506	749		491	748		490
Unconstrained					x							
Covariates								x				
Weights to match full sample											x	

Note: This sample includes families with third grade students only.

Bridge @Home Teacher Calls - Sample Transcript (Accountability)

Bridge SMS + Teacher Calls programme connects pupils with their classroom teachers. Teacher calls can take many forms. Below, you can find a sample transcript to guide your calls to parents and pupils.

Sample Transcript

Step 1: Introduction

Guardian

- Introduce yourself and your reason for calling
- Ask to speak with the participating parent/guardian
- Explain you are calling to check in on @Home learning
- Ask the parent to put their phone on speaker and invite the pupil(s) to participate in the call

Teacher: Hello. My name is Teacher Francis calling from Baba Dogo school. I am the Class 5 teacher of your son Ibrahim.
Parent: Hello Teacher Francis.
Teacher: I am calling to check on Ibrahim's @Home learning this week. Can you put your phone on speaker and ask Ibrahim to join?
Parent: Yes. Let me get him…[1 minute later] The phone is on speaker and Ibrahim is here.
Teacher: Hello Ibrahim, this is Teacher Francis calling. I am calling to check in on your maths work this week.
Pupil: Hello Teacher Francis.

Step 2: Gather Relevant information

Parents
Student(s)

- Learn whether the parent received the weekly SMS this week
- Assess whether the pupil has completed the weekly SMS problem
- Cold-call the pupil to learn whether they answered correctly. Use the correction procedure
- Assess whether the pupil has completed any quizzes this week

Teacher: Did you receive the problems sent through SMS this week?
Parent: Yes, we did.
Teacher: Ibrahim, did you complete the practice problems?
Pupil: Yes, I did.
Teacher: What was the topic of the problems this week?
Pupil: Adding fractions with the same denominator.
Teacher: The first question was 2/6 + 3/6. What was your answer?
Pupil: 1/6.
Teacher: Incorrect. The correct answer is 5/6.
Pupil: OK.
Teacher: The second question was 5/12 + 1/12. What was your answer?
Pupil: 6/24.
Teacher: Incorrect. The correct answer is 6/12 or 1/2.
Pupil: Thank you.
Teacher: Now we will discuss how you have performed with quizzes this week. How many mobile interactive quizzes in math have you completed this week?
Pupil: I have completed 4 interactive maths quizzes.
Teacher: How did you score on each quiz that you completed this week?
Pupil: I scored 80% on 3 quizzes and 40% on 1 quiz.
Teacher: Thank you Ibrahim.

Step 3: Conclusion

- Preview the following week's call (will you be calling again, will there be any change, etc.)
- Recommend that the pupil takes mobile interactive quizzes for additional practice
- Confirm that the current contact number is the best way to reach the parent
- If possible, schedule a day and time for next week's call
- Thank the parent and pupil; end the call

Guardian

Teacher: Next week, I will call again on the same day and at a similar time. Is this still the correct number to reach you?
Parent: Yes, it is.
Teacher: Next week, we will review the week's SMS problems. I would also like you to take at least 5 maths quizzes using the mobile quiz service. You can take quizzes on Operations with Fractions for more practice. Do you know how to use the mobile quiz service?
Parent: No, I do not.
Teacher: You received the WhatsApp link by SMS from Bridge. Click on the link to begin using the quizzes.
Parent and Pupil: Thank you Teacher Francis.
Teacher: Thank you. We will speak again next week. Goodbye

Bridge @Home Teacher Calls - Sample Transcript (Accountability Plus Tutoring)

Bridge SMS + Teacher Calls programme connects pupils with their classroom teachers. Teacher calls can take many forms. Below, you can find a sample transcript to guide your calls to parents and pupils.

Sample Transcript

Step 1: Introduction

Guardian

- Introduce yourself and your reason for calling
- Ask to speak with the participating parent/guardian
- Explain you are calling to support the parent and pupil(s) on a maths topic
- Ask the parent to put their phone on speaker and invite the pupil(s) to participate in the call

Teacher: Hello. My name is Teacher Francis calling from Baba Dogo school. I am the Class 5 teacher of your son Ibrahim.
Parent: Hello Teacher Francis.
Teacher: I am calling to support Ibrahim with the weekly SMS problem that you received this week. Can you put your phone on speaker and ask Ibrahim to join?
Parent: Yes. Let me get him…[1 minute later] The phone is on speaker and Ibrahim is here.
Teacher: Hello Ibrahim, this is Teacher Francis calling. I am calling to support you with your maths work this week.
Pupil: Hello Teacher Francis.

Step 2: Gather Relevant information

Parents
Student(s)

- Learn whether the parent received the weekly SMS this week
- Assess whether the pupil has completed the weekly SMS problem
- Cold-call the pupil to learn whether they answered correctly. Use the correction procedure
- Answer questions that parents and/or pupil(s) might have while attempting the SMS problem
- Ask whether the pupil is completing other @Home assignments like mobile interactive quizzes

Teacher: Did you receive the problems sent through SMS this week?
Parent: Yes, we did.
Teacher: Ibrahim, did you complete the practice problems?
Pupil: Yes, I did.
Teacher: What was the topic of the problems this week?
Pupil: Adding fractions with the same denominator.
Teacher: The first question was 2/6 + 3/6. What was your answer?
Pupil: 1/6.
Teacher: Incorrect. The correct answer is 5/6.
Pupil: OK.
Teacher: The second question was 5/12 + 1/12. What was your answer?
Pupil: 6/24.
Teacher: Incorrect. The correct answer is 6/12 or 1/2. Did you have any questions when you completed the problems? It seems like you struggled to answer the questions correctly.
Pupil: No, I did not have any questions.
Teacher: Now we will discuss how you have performed with quizzes this week. How many mobile interactive quizzes in math have you completed this week?
Pupil: I have completed 4 interactive maths quizzes.
Teacher: How did you score on each quiz that you completed this week?
Pupil: I scored 80% on 3 quizzes and 40% on 1 quiz.
Teacher: Thank you Ibrahim. Did you have any questions on the quizzes that you took this week?
Pupil: No.

Step 3: Delivering Content

- Begin with a **demonstration** of the topic in the SMS problems. Focus on 2-3 key concepts that are necessary for an understanding of the topic.
- Review an additional **example** of the topic with the pupil.
- Assign additional **"problems of the day"**.

Student(s)

Teacher: Now, I will explain more about the topic of the week. This week's topic is adding fractions with the same denominator. What is a fraction?
Pupil: A fraction is a part of a whole.
Teacher: Good. A fraction is a part of a whole. What is a denominator?
Pupil: The number on the bottom of a fraction.
Teacher: Good. The denominator is the number on the bottom of the fraction. What is the number at the top of a fraction?
Pupil: Numerator.
Teacher: Good. The numerator is the number at the top of the fraction. When we add fractions with the same denominator, which number do we add?
Pupil: The numerator.
Teacher: Good. When we add fractions with common denominators, we add the numerators together. Do we add the denominators together?
Pupil: No.
Teacher: Good. The denominator stays the same. Let's use an example. If we were adding one fourth and two fourths, how would we solve it? First, we would add the numerators together. One plus two equals three. The new numerator is three. The denominator stays the same. The new denominator is four. So one fourth plus two fourths equals three fourths. Do you understand?
Pupil: Yes. Thank you.
Teacher: Now, I will assign 5 additional practice problems. Are you ready to write the problems down?
Pupil: Yes.
Teacher: One. One eighth plus four eighths. Two. Two ninths plus one ninth. Three. Five twelfths plus six twelfths. Four. Two thirds plus zero thirds. Five. One seventh plus six seventh. Work on these five problems independently.
Pupil: OK teacher Francis.

Step 4: Conclusion

- Preview the following week's call (will you be calling again, will there be any change, etc.)
- Recommend that the pupil takes mobile interactive quizzes for additional practice
- Confirm that the current contact number is the best way to reach the parent
- If possible, schedule a day and time for next week's call
- Thank the parent and pupil; end the call

Guardian

Teacher: Next week, I will call again on the same day and at a similar time. Is this still the correct number to reach you?
Parent: Yes, it is.
Teacher: Next week, we will review the week's SMS problems, and I will lead a demonstration on the new topic. I would also like you to take at least 5 maths quizzes using the mobile quiz service. You should focus on quizzes on Operations with Fractions for more practice. Do you know how to use the mobile quiz service?
Parent: No, I do not.
Teacher: You received the WhatsApp link by SMS from Bridge. Click on the link to begin using the quizzes.
Parent and Pupil: Thank you Teacher Francis.
Teacher: Thank you. We will speak again next week. Goodbye

<u>Example SMS Message: Grade 3</u>

Welcome to week 1!
NUMBER PATTERNS
5, 9, ?, 17
34, ?, 48, 55
?, 21, 13, 5
The first 3 numbers of a pattern are 11, 22, 33. What is the next number in a pattern?
11, 22, 33, ?
See whether numbers are getting bigger or smaller. Find the difference between each number.

11+11=22
22+11=33
33+11=?

For more practice, try taking Maths mobile quizzes this week in the following topic: Whole Numbers.