School's Out: Experimental Evidence on Limiting Learning Loss Using "Low-Tech" in a Pandemic

Noam Angrist, Peter Bergman, and Moitshepi Matsheng*

November 2020

Abstract

The COVID-19 pandemic closed schools at one point for over 1.6 billion children, with potentially long-term consequences. This paper provides some of the first experimental evidence on strategies to minimize the fallout of the pandemic on learning. We evaluate two low-technology interventions to substitute schooling during this period: SMS text messages and direct phone calls. We conduct a rapid trial in Botswana to inform real-time policy responses, collecting data in multiple waves. We find that phone calls and SMS messages result in cost-effective learning gains of 0.12 standard deviations. We cross-randomize targeted instruction, customizing instruction to a child's learning level using data collected during the trial. We find evidence that targeted instruction can be more effective than non-targeted instruction, especially for SMS messages which have no effect on their own if they are not targeted. Learning gains are robust to a variety of tests, such as randomized problems of the same proficiency and measures of effort on the test. Parents update their beliefs about their child's learning in tandem with progress and they feel greater self-efficacy to support their child's learning. The "low-tech" interventions tested have immediate policy relevance and could have long-run implications for the role of technology and parents as substitutes or complements to the traditional education system.

---

*Correspondence to: noam.angrist@bsg.ox.ac.uk. This trial is registered in the AEA RCT registry at: AEARCTR-0006044. This paper updates an earlier version reporting midline results, with particular thanks to Caton Brewster who co-authored the midline paper. Winnifred Arthur provided excellent research assistance. We thank Jenny Aker, Jim Berry, Alex Eble, Penny Goldberg, Eric Hanusheck, Michael Kremer, Clare Leaver, Susanna Loeb, Todd Rogers, Anna Rudge and participants of the University of Oxford development economics workshop as well as World Bank, RTI, FHI360, and USAID webinars for helpful comments. The intervention and trial were the product of a collaboration between the Botswana Ministry of Basic Education and staff at Young 1ove who nimbly adapted during school closures to collect phone numbers and deliver the interventions. There are nearly a hundred staff who deserve mention and are named on the Young 1ove website. Particular gratitude to Efua Bortsie, Colin Crossley, Thato Letsomo, Rumbidzai Madzuzo, and Tendekai Mukoyi who coordinated and designed the low-tech programs, Patience Derera for carefully compiled cost estimates, Shawn Maruping and Dorothy Okatch for online communications, and Bonno Balopi, Amy Jung, Gaone Moetse, Bogadi Mothlobogwa, Astrid Pineda, Julio Rodriguez and Katlego Sengadi who provided research and implementation support. We thank Madhav Chavan, Samyukta Lakshman, Devyani Pershad, Meera Tendolkar, Usha Rane and the Pratham staff for close guidance on the design of the low-tech interventions. We thank Emily Cupito and Ashleigh Morrell for sharing relevant evidence briefs to inform the low-tech interventions. We thank Dave Evans, Susannah Hares and Matthew Jukes for collaboration on measuring learning via the phone. We thank Mahsa Ershadi, Rebecca Winthrop, and Lauren Ziegler for collaboration on a related survey of parent perceptions. We thank flexible funders and partners who enabled a rapid COVID-19 response, including the Mulago Foundation, the Douglas B. Marshall Foundation, J-PAL Post-Primary Education (PPE) Initiative, and TaRL Africa. We thank Northwestern University's "economics of nonprofits" class led by Dean Karlan, which provided a generous donation. This trial builds on a prior effort to scale up an education intervention called "Teaching at the Right Level" in over 15 percent of schools in Botswana. The coalition supporting scale-up of Teaching at the Right Level in Botswana includes the Botswana Ministry of Basic Education, the Botswana Ministry of Youth Empowerment, Sport and Culture Development, Young 1ove, UNICEF, USAID, Pratham, the Jameel Poverty Action Lab (J-PAL), TaRL Africa, the Brookings Institution, and the People's Action for Learning (PAL) network. The infrastructure built by this coalition prior to COVID-19 enabled this rapid trial and response.

1 Introduction

The COVID-19 pandemic has paralyzed education systems worldwide: at one point, school closures forced over 1.6 billion learners out of classrooms (UNESCO 2020). While smaller in scale, school closures are not unique to COVID-19: teacher strikes, summer breaks, earthquakes, viruses such as influenza and Ebola, and weather-related events cause widespread closures. Closures often result in large learning losses, which have been documented in North America, Western Europe, and Sub-Saharan Africa (Cooper et al. 1996; Slade et al. 2017; Jaume and Willen 2019; Andrabi, Daniels, and Das 2020). To mitigate learning loss in the absence of school, high-income families have access to alternative sources of instruction—books, computers, internet, radio, television, and smart phones—that many low-income families do not (Bacher-Hicks et al. 2020; Chetty et al. 2020; Engzell et al. 2020). Stemming learning loss when schools are closed, particularly in areas where learning resources are lacking in the household, requires outside-school interventions that can substitute instead of complement ongoing instruction. Doing so at scale requires cheap, low-technology solutions that can reach as many families as possible.

In this paper, we provide some of the first experimental estimates on minimizing the fallout of the COVID-19 pandemic on learning. We evaluate two "low-tech" solutions that leverage SMS text messages and direct phone calls to support parents to educate their children. A sample of 4,500 families with primary-school-aged children across nearly all regions of Botswana were randomly assigned to either intervention arm or a control arm. In one treatment arm, SMS text messages provided a few basic numeracy "problems of the week." In a second treatment arm, live phone calls from instructors supplemented these SMS text messages. These calls averaged 15-20 minutes in length and provided a direct walk-through of the learning activities sent via text message. Using high-frequency data collected at week four we also cross-randomized a targeted instruction intervention which sent customized SMS and phone instructions based on student numeracy levels.

Our results show large, statistically significant learning differences between treatment and control groups. For the combined phone and SMS group there was a 0.121 standard deviation (p=0.008) increase in the average numerical operation learned.[2] For households who participated in all sessions, instrumental variables analysis shows learning gains are 0.168 standard deviations (p=0.007). These gains translate to being able to do place value, as well as solving fractions. We find evidence that targeted instruction is more effective on a broad set of competencies, such as learning place

[2] The test used is adapted to the phone from a face-to-face ASER test frequently used in the literature (Banerjee et al. 2007; Banerjee et al. 2010; Banerjee et al. 2017; Duflo et al. 2020).

value, and is three times more effective for learning higher-order competencies such as solving fractions, although it is no better than non-targeted instruction on basic competencies. Targeting seems to be particularly important for SMS messages, with no effect on average, which appears to be driven by limited effects for non-targeted SMS messages, while we find evidence of positive effects for targeted SMS messages.

Together, these results demonstrate that instruction through "low-tech" mobile phones can provide a cost-effective and scalable method to deliver educational instruction outside of the traditional schooling system and to personalize instruction.

We also present several innovations on remote assessment to test the robustness of our learning measures. The learning assessment used was adapted from the Annual Status of Education Report (ASER) into a phone-based assessment and incorporated time limits and a requirement that children explain their work to accurately identify their numeracy levels. We find that learning results are robust to randomized problems that test the same proficiency, a standard reliability test in the psychometric literature (Crocker and Algina 1986). We further disentangle cognitive skills gains from effort effects, which have been shown to affect test scores (Gneezy et al. 2019). In our context, where learning outcomes are measured remotely in the household, effort might be particularly important. We test this hypothesis with a real-effort task. We find that student effort is unaffected by the interventions, suggesting the learning gains observed stem from cognitive skill gains, rather than from effort on the test. It is also possible learning gains are a matter of familiarity with the content in intervention groups which receive exposure to similar material. We test this by including new content not covered during the intervention, but which is related, such as fractions, and find that in the phone and SMS group learning gains translate to fractions. The familiarity hypothesis is also partially tested with randomized problems of the same proficiency, which do not affect results.

We explore parental educational investment mechanisms.[3] Parents exhibit strong demand for the intervention, with over 99 percent of households expressing interest in continuing the program after the first four weeks. Parental engagement in their child's education is high with 92 percent of parents reporting their child attempted to solve the problems sent, with slightly higher engagement in the phone call group of 95 percent. Parents report 8.4 and 15.2 percent greater self-efficacy in supporting their child's learning as a result of the SMS only and phone and SMS interventions, respectively. Parents also update their beliefs about their child's learning level in tandem with their child's learning progress. This suggests that parents are involved and aware of their child's academic progress. We also find that parental return to work post

[3] We use the term "parent" in this paper for consistency with the literature. In practice, we engage "caregivers", 81 percent of whom are parents, 7.6 percent are grandparents, 7.8 percent are aunts or uncles, 2.8 percent are siblings, and less than 1 percent are cousins.

lockdown is unaffected by the interventions, and if anything, is slightly higher, which alleviates the concern that further parental engagement in their child's education might crowd out other activities, such as returning to work.

We also explore partial versus full school substitution mechanisms. The results at endline are particularly striking given schools partially reopened during this period, although with frequent disruptions, whereas at the time of the midline survey schools were fully closed. This suggests that the phone and SMS treatment was effective even when it served as a *partial substitute* to schooling, in addition to as a *full substitute* in the first stage of the trial. However, effects are smaller in the second versus first stage of the trial. We see that the phone and SMS treatment reduced innumeracy by 52 percent relative to 31 percent in the first versus second wave, and a 34 percent reduction relative to 11 for the targeted SMS only group. This could be for a few reasons. First, average treatment effects could stay constant, but engagement could diminish over time. Second, even if engagement persists over time, effects might diminish in size due to habituation or fatigue. Third, the interventions tested might be most effective as pure substitutes, since when schools are closed and virtually no learning takes place, these simple low-cost interventions might matter most. In contrast, when schools reopen, the low-tech interventions now provide a partial rather than a full substitute. Altogether, the results suggest targeted SMS messages and phone calls are most effective as full substitutes, but still improve learning as partial substitutes.

Our work contributes to several literatures. The low-tech interventions tested relate to a growing literature on mobile phone technology and education. Mobile phone SMS messages have been used to supplement adult education programs in Niger and the U.S. (Aker et al. 2012; Aker et al. 2015; Aker and Ksoll 2020), to help parents teach nascent literacy skills to their children in the U.S. (York et al. 2018; Doss et al. 2019), and to help parents monitor their child's effort and progress in school (Kraft and Rogers 2015; Berlinski et al. 2016; Cunha et al. 2017; Siebert et al. 2018; de Walque and Valente 2018; Bergman and Chan 2019; Musaddiq et al. 2019; Gallego et al. 2020; Bergman 2020). See Bergman 2019 for a review. We contribute to this literature by providing evidence on live, direct instruction as well as automated, text-message based instruction, and in a setting where these interventions operate largely as substitutes for schooling rather than as complements.[4] We also contribute novel learning data via phone-based assessment.

We also relate to a literature on targeted instruction. An educational approach called "Teaching at the Right Level" (TaRL), a classroom-based intervention evaluated over 20 years which targets instruction by learning level rather than by age or grade,

[4] The role of technology as a complement or substitute for the traditional schooling system is reviewed in Bettinger et al. (2020).

has been shown to produce cost-effective gains in learning across multiple studies. This approach has worked when delivered by teachers or volunteers (Banerjee et al. 2007; Banerjee et al. 2010; Duflo, Dupas, and Kremer 2011; Banerjee et al. 2017; Duflo et al. 2020) and when using adaptive computer software (Banerjee et al. 2007; Muralidharan, Singh, and Ganimian 2019). We contribute to this literature by testing a particularly low-cost and scalable approach to target instruction using phone-based instruction.

Our results have significant implications for global policy. Recent estimates from the World Bank suggest current school closures could cost up to $10 trillion in net present value (Azevedo et al. 2020). Hanushek and Woessmann (2020) estimate that learning losses due to COVID-19 could yield an average of 1.5 percent lower annual GDP for the remainder of the century. To mitigate this fallout of the pandemic on education, there is global demand for effective solutions to reduce learning loss. Even as schools start to reopen, this reopening is often partial and schooling may be periodically interrupted as new outbreaks occur and as new social distancing guidelines come into effect. Moreover, school closures occur in settings beyond the current pandemic, including summer holidays, public health crises, during adverse weather events, natural disasters, and in refugee and conflict settings. In moments where a substitute for schooling is needed, particularly for families with fewer resources at home, the low-tech solutions tested in this trial have unique potential to reach the masses. While only 15 to 60 percent of households in low- and middle-income countries have internet access, 70 to 90 percent of households own at least one mobile phone (Carvalho and Crawford 2020). Our results provide early evidence that remote instruction by phone and simple SMS messages has high potential to improve children's learning at low cost and at scale both when children are out of school and as schools start to reopen.

2 Background

2.1 Global Education and COVID-19 Landscape

Over 190 countries closed schools at the height of the COVID-19 pandemic (UNESCO 2020). Estimates of learning loss due to mass school closures reach nearly a full year of schooling adjusted for quality (Azevedo et al. 2020). Even before the pandemic, student learning levels were low and progress was slow, as highlighted by UNESCO and the World Bank (Angrist, Djankov, Goldberg, and Patrinos 2019). According to the World Bank's learning poverty measure, less than 50 percent of students in developing countries could read a story by age 10 (World Bank 2019). To address learning shortfalls, which have been exacerbated by the COVID-19 pandemic, there is a need for approaches that cost-effectively improve learning on a global scale.

2.2 COVID-19 Context in Botswana

Botswana enacted pre-emptive social distancing measures before recording its first COVID-19 case. While the first suspected COVID-19 death occurred in Botswana on March 25th, schools had already been closed, initially for a planned six months starting March 20th. As of early November 2020, around 25 COVID-19 deaths have occurred in Botswana. With around 300,000 tests conducted to date, this statistic is unlikely due to limited testing. While the direct effects of the pandemic have been minimal, the fallout of the pandemic on educational and social services has been severe. Botswana declared a state of emergency on March 31st. Schools reopened on June 17th, were subsequently closed again after a new wave of COVID-19 cases, and have since reopened. Our data reveals that the vast majority of students returned to school (98.8 percent). Similar waxing and waning of school closure is anticipated in the coming months. Even as students return to school, a double-shift system, where half of the students rotate into school in the morning and the other half rotate in the afternoon, drastically reduces time in school for each student. While the government has launched learning programs on national television and radio stations to provide learning content for students, survey data suggests there is high demand among parents and communities for additional remote educational activities for their children.[5] Over 99 percent of parents reported demand for continued remote learning services even if schools reopened, like due to uncertainty around whether schools would remain open, reduced school hours, and disrupted learning.

3 Intervention

A few days before the government announced that schools were closing as a result of the state of emergency, we collected 7,550 phone numbers from primary schools. This response built on an active presence in schools by Young 1ove, one of the largest NGOs in Botswana, which was conducting educational programming in partnership with the Ministry of Basic Education.

Young 1ove's staff, from here on referred to as "facilitators," collected phone numbers in primary schools from students, parents, and teachers in schools where active

[5] In addition, we find access to radio is relatively low. Data from our midline survey shows the only 20 percent of students in the control group are listening to radio in the status quo.

Teaching at the Right Level programming had been in session to enable remote engagement in the pending school closures. These numbers were collected in nearly all schools with an active presence and largely for students in grades 3 to 5. Given high interest for remote support, numbers were collected primarily from students who were *not* participating in prior programming (82 percent of all numbers) as well as those who were (18 percent). After phone collection and verification, facilitators called all numbers to gauge interest from parents in receiving remote learning support via phone. Over 60 facilitators were engaged through training via WhatsApp where voice notes and short briefing scripts were shared on how to conduct the calls.

For parents who opted into remote learning support, we provided two low-tech interventions: (a) one-way bulk SMS texts with multiple numeracy "problems of the week" and (2) SMS bulk texts with live phone call walkthroughs of the problems on a 15-20-minute phone call. Both low-tech interventions were intentionally designed to be simple in order to be digestible via phone by parents, teachers, and students and scalable by governments.

The first intervention was a weekly SMS containing several simple math problems; for example, "Sunshine has 23 sweets. She goes to the shops to buy 2 more. How many does she have altogether?" The SMS was sent at the beginning of each week via a bulk texting platform. The SMS contained a message with 160 to 320 characters that could fit in one or two texts. Figure 1 shows an example weekly message of practice problems focused on place value.

Text messages were sent to parent phone numbers since primary school children rarely have their own phones. In some cases, parents shared the message directly with students and in other cases parents engaged directly with their children to solve the problems. The SMS was one-way and did not require or elicit a two-way response. This was most logistically straightforward given two-way messaging is not always available cheaply and consistently.

The second intervention was a weekly phone call ranging in typical length from 5 to 20 minutes in addition to the weekly SMS, which was sent at the beginning of the week. On the call, the facilitator asked the parent to find the student and put the call on speaker. This arrangement allowed both the parents and student to hear the facilitator at the same time and to engage in learning. The facilitator confirmed that the student had received the SMS message sent and answered any questions related to the task. Furthermore, the facilitator provided the student with a math question to go over and practice. The calls served to provide additional learning support as well as motivation and accountability. Many parents proudly reported to the facilitators during these phone calls that their child had successfully completed the problems of the week. Figure 2 includes a subset of a sample phone call script.

The goal for the calls was to conduct them with both the caretaker and the child simultaneously. This strategy maximized the probability the child received educational support and lowered future barriers to entry for parents to continue engaging in educational activities. It also provided a measure of child protection by ensuring a guardian was present during phone calls with children. Simultaneous calls happened roughly 45 percent of the time. About 37 percent of the time, the facilitators spoke with only the caretaker, using the time to explain how he or she could support their child with the SMS message problems. For the remaining 18 percent of calls, the respondents were unavailable (15 percent), a logistical barrier occurred (2 percent), or they no longer wanted to be part of the program (1 percent).

64 facilitators were assigned a group of about 24 households each to case manage. Each facilitator called for around 6 hours a day to reach all households every week. Facilitators would periodically request ideal scheduling times to call to maximize the probability parents would engage. On average, over 50 percent of total calling time was spent following up or on household logistics (e.g., a parent and child finding a joint space in their household to engage).

Pilots of both interventions were conducted over the course of two weeks prior to launching the interventions to ensure acceptability and feasibility. For example, more elaborate learning activities were initially planned, such as tossing stones into three concentric circles with various place values. However, pilots revealed this level of interaction was deemed too difficult to describe effectively and quickly over the phone in the first series of interactions. To this end, we shifted to conducting simple practice problems similar to those shown in Figures 1 and 2.

A random subset of phone numbers were also cross-randomized with an additional intervention: targeted instruction to each child's learning level. We used data on learning levels from a midline phone-based learning assessment to send tailored text messages to each student in the fifth week. For example, students who knew addition received subtraction problems, whereas students who knew multiplication were sent division problems. This targeted instruction program was adaptive, building on data collected at week four, which (a) enabled us to have near real-time data to target instruction and (b) revealed parents were having a hard time learning the level of their child, thus requiring additional support to target instruction. At each week interval thereafter, we collected weekly data on a "problem of the week" using a phone survey and used this question to further target weekly SMS messages. If the child responded correctly, problems the following week progressed to a higher operation level. If not, the child was given the same level problem. At approximately week twelve, we collected additional survey data and conducted learning assessments to evaluate the targeted instruction component of the intervention.

4 Experimental Design

We collected 7,550 phone numbers in primary schools throughout the country the week before the lockdown was instated. To put this scale in context, there are roughly 44,000 students in one primary grade level across the nation in Botswana. Over a period of two weeks following phone number collection, facilitators called these 7,550 phone numbers to confirm that they were valid numbers, that they belonged to the caregiver of a child in primary school, and, if so, to inform these caregivers about the program and gain consent to participate. We managed to reach and collect follow-up data on roughly 6,375 of the 7,550 phone numbers initially collected. The remaining numbers were either invalid, unreachable, or the child was no longer with the caregiver of the original number given, often due to moving to stay with a different relative. Of the 6,375 phone numbers reached, 4,550 households (about 71 percent) were interested and gave consent to participate in the trial. For this cohort of 4,550 participants, we include a heat map in Figure 3 of the location of the children's schools to demonstrate the distribution of participants across the country.

We randomized the 4,550 phone numbers initially into three groups of equal size: a weekly SMS message followed by a phone call, a weekly SMS message only, and a pure control group. We then cross-randomized 2,250 numbers for a midline assessment and approximately 1,600 phone numbers to receive targeted instruction customized to their learning level using data collected at midline. Randomization was stratified on whether at least one child in the household had previously participated in prior school-based "Teaching at the Right Level" programming, a proxy for having recently made substantial learning gains. Each phone number belongs to a caregiver and household. Roughly 80 percent of the households had one student while 20 percent had multiple students.

Figure 4 provides a timeline of each step from initial phone number collection, piloting and training, program implementation and waves of data collection. Figure 5 provides an overview of the experimental design.

4.1 Data Collection

We conducted two waves of data collection. The endline occurred after 4 months and a midline occurred shortly before the halfway point. The endline survey consists of 17 questions including a learning assessment, parental engagement in educational activities, and parental perceptions of their child's learning and their own self-efficacy. A portion of the survey was conducted with the parent and learning outcomes were collected by directly assessing the child over the phone. The assessment was adapted from the ASER test, which has been adapted for use in over 14 different countries. The ASER test consists of multiple numeracy items, including 2-digit addition (Level 1), subtraction (Level 2), multiplication (Level 3) and division problems (Level 4). A level of 0 on the test is referred to as "beginner" level and indicates the student cannot successfully do any operations which we also refer to as "innumeracy." Figure 6 shows a sample assessment adapted from ASER. The ASER test has consistently been used in the Teaching at the Right Level literature (Banerjee et al. 2017). In order to improve the reliability of the phone-based assessment, we introduced a series of quality-assurance measures: students had a time cap of two minutes per question to minimize the likelihood of family members in the household assisting the child, and we asked each child to explain their work and only marked a problem correct if the child could correctly explain how they solved the problem. While imperfect, these measures provide a level of verification to maximize the likelihood the test captures child learning. We assigned facilitators to phone numbers using an arbitrary match sorted by phone number order. On average, each facilitator was assigned to about 30 phone numbers. Less than 1.5 percent of facilitators that provided weekly intervention calls surveyed the same household, providing for objective assessment. We also conduct several checks to verify the validity of our measures, described further below.

We convert the raw score in terms of standard deviations relative to the control group. In addition to the ASER test, we evaluate the children's ability to answer a simple place value word problem such as "Katlego has 77 apples and organizes them by place value. How many tens does she have?" to capture learning outcomes beyond a core set of mathematical operations.

In addition, we added a series of additional questions to identify mechanisms driving learning gains. This includes a real-effort task in the form of a riddle: "the day before two days from now is Saturday. What day is today?" We also include a higher-order numeracy question to assess whether learning gains translate to material not covered directly in the intervention. In particular, we ask a question on fractions such as "$\frac{3}{8}+\frac{5}{8}=?$" We further conduct a reliability assessment by randomizing five different questions of each proficiency (addition, subtraction, multiplication, division, and fractions) to formally assess the reliability of the learning assessment questions (Crocker and Algina 1986). For example, for a division problem, we have one problem which asks students to divide 68 by 5 and another problem where 38 is divided by 3. Both are two-digit division problems with remainder. If both problems have a similar distribution, as expected given they measure the same latent ability, this increases our confidence in learning estimates.

We also include questions on parental engagement, perceptions, and self-efficacy. We measure learning engagement by asking parents if they recall their child attempting any of the problems sent over the last few weeks. We include a measure of a parent's perception of their child's numeracy level by directly matching their perception of their child's level to their child's actual learning level. If a parent estimates the highest level their child can do is subtraction and their child indeed performs up to subtraction level we code this as "correct." If the parent overestimates or underestimates their child's level we code this as incorrect. We also capture parents' confidence in supporting their child's learning at home and whether they felt their child made progress during the school closure period. We code a dummy for whether parents are "very confident" for both indicators. Additional questions are included on whether the child has returned to school and whether the caregiver has returned to work. Finally, demographic questions were included on the child's age, grade, and gender.

The midline assessment was nearly identical to the endline assessment, including the same phone-based ASER assessment. The midline did not include real-effort tasks, fractions problems, randomized problems or questions around parent self-efficacy and return to work and school. The midline also asked about demand for remote learning services if schools were to reopen.

4.2 Sample Characteristics and Representativeness

We include a few descriptive statistics to describe how our sample for the low-tech intervention compares to characteristics of other relevant samples in Botswana. Botswana has ten regions total and our sample covers 9 out of all 10 regions. The low-tech sample includes over 103 schools which represents around 15 percent of schools in the country.

We compare our low-tech sample of households to a sample of students who had participated in Teaching at the Right Level in Botswana in the prior two years where face-to-face ASER learning assessments had been implemented. This comparison enables us to explore the representativeness of our sample on learning measures and the comparability of phone and face-to-face assessment. The low-tech sample is similar in the overall number of schools represented, with slightly greater coverage of 103 schools relative to 92 schools. The greater number of schools is likely due to friends and relatives being included as students and parents migrated around the country to relatives' households post school closure. In addition, we compare learning levels in the control group of the first wave of assessment, which captures both the sample composition as well as a measure of how similar phone-based ASER assessments are to face-to-face ASER assessments. By and large we find a similar composition of learning levels, with students who cannot do any operations ranging between 29 to 31 percent in both samples and all learning levels within 2 to 6 percentage points of one another across samples and assessment methods. This comparison shows that our low-tech sample is broadly representative of learning levels across an alternative sample in Botswana and that phone-based assessments capture a similar distribution of learning as face-to-face assessments.

We also compare our low-tech sample to national-level indicators from the Ministry of Basic Education using data on enrollment and gender composition from 2017. We see a similar gender split between 50 to 51 percent in both samples. For schools represented in the low-tech sample we find slightly smaller average enrollments in standards 3 to 5 of 274 students relative to average enrollments of 362 for the national sample. This is likely due to the low-tech sample having representation weighted towards remote villages relative to the national distribution. We also compare study schools on the Primary School Leave Examinations (PSLE) from the Botswana Examinations Council. We find similar distributions of learning: the percentage of students who score an A, B, and C is 16, 21 and 41 percent in study schools, respectively, and 14, 17 and 36 for all primary schools in the nation.

In addition, we collect simple descriptive data on child age, grade, and gender in surveys. Around 50 percent of our sample is female; the average age of students is 9.7; 28.5 percent of students are in grade 3, 39.1 percent in grade 4 and 32.4 percent in grade 5. We also capture the identity of the caregiver at the household whose number was provided and who is providing instructional support to their child. We find that 81 percent of caregivers are parents, 7.6 percent are grandparents, 7.8 percent are aunts or uncles, 2.8 percent are siblings, and less than 1 percent are cousins.

For a subsample of parents, we also measure parental education level and additional characteristics.[6] These measures suggest the sample of parents in the trial have relatively low literacy. 32 percent had only completed Form 5, which means they did not attend university. 18 percent had started university but did not finish, and 16 percent did not finish Form 5 and thus did not complete a high school degree. The average age of parents or caregivers participating in the randomized trial was 35 and 68 percent of parents were female and 32 percent male.

4.3 Balance and Attrition

[6] A subset of 222 parents were asked a series of additional questions in a survey conducted via Whatsapp in partnership with the Brookings Institution, including parental education level (Winthrop et al. 2020). This subset of parents is not necessarily representative of the entire sample. However, they were the most responsive parents, suggesting they had reliable internet access and likely represent an upper bound of the most literate parents.

Table 1 shows endline survey response rates. We successfully followed up with 64.9 percent of households. Nearly all students were assessed for learning outcomes, with a similar response rate of 63.8 percent for place value questions and a 62.2 percent response rate for all operations questions. Columns 1, 2, and 3 show that there are no statistically significant response rate differences between treatment groups relative to the control group or each other. This suggests analysis on endline outcomes is unbiased across study groups among respondents.

Table 2 reports balance across a series of indicators including student grade, age, and sex, as well as the identity of the caregiver at the household whose phone number was identified for providing support to the child at the household (e.g., if they were a parent). Though we have limited baseline covariates, we see no statistically significant differences between groups. In addition, we link administrative data from schools for students who participate in the trial and for the subset of households for which we are able to link them to a specific school. We examine differences across school-level pass rates on the primary school leaving examination in prior years. Again, we find no statistically significant differences. These tests reveal that randomization appears to have been successful in generating comparable treatment and control groups.

5 Empirical Strategy

We estimate treatment effects of the SMS only and phone and SMS intervention using the following specification:

$$Y_{ij} = \alpha_0 + \beta_1 SMS_j + \beta_2 PhoneSMS_j + \delta_s + \varepsilon_{ij}$$

where Y_{ij} is an outcome for child i in randomly assigned household j. SMS is an indicator variable coded to one for the SMS message only treatment group and zero otherwise, and SMSPhone is an indicator variable coded to 1 if a household received both an SMS and a phone call and zero otherwise. δ_s is a strata indicator, which indicates whether a child participated in education programming immediately prior to the intervention. We include only one child identified for instruction in each household level j, which is determined by the caregiver's phone number and is the unit of randomization. We use this specification to measure the impact of each intervention on students' learning level, engagement, and parents' perceptions of their child's level and self-efficacy.

We also estimate the effect of targeted instruction with the following specification:

$$Y_{ij} = \alpha_0 + \beta_1 Targeted_j + \beta_2 NotTargeted_j + \delta_s + \varepsilon_{ij}$$

Given randomization and equivalent treatment and control groups, each specification estimates causal effects of the intervention.

6 Results

The results presented are primarily intention-to-treat effects on whether households were randomly assigned to treatment. Monitoring data suggests that in the phone group active weekly participation by parents was on average between 85 to around 60 percent. This suggests treatment-on-the-treated effects for those who actively participate are likely to be larger than our reported intention-to-treat effects. We estimate treatment-on-the-treated effects directly in section 7.4.

6.1 Learning

Our results show large, statistically significant learning differences between treatment and control groups after four months. In Table 3, we see that for the combined phone and SMS group, there was 0.121 standard deviation (p=0.008) increase in the average numerical operation. These gains translate to broader competencies, such as gains in place value of .114 standard deviations (p=0.009) as well as higher-order competencies, such as solving fractions with gains of 0.075 standard deviations (p=0.100). As we show in section 6.2, these results are robust to a number of validity checks. We find no significant effects on average for SMS messages only across all learning proficiencies.

We also find that targeted instruction performs similarly for basic numerical competencies with effects of .076 standard deviations (p=0.097) relative to non-targeted interventions with a 0.070 standard deviation effect (p=0.130). However, targeted interventions are more effective on a broader set of competencies, with effects on knowing place value of 0.098 standard deviations (p=0.026) relative to 0.026 standard deviations (p=0.572) for non-targeted instruction. Targeted interventions are also substantially more effective on higher-order competencies with 0.093 standard deviation gains (p=0.041) on solving fractions relative to 0.029 standard deviation gains (p=0.527) for non-targeted instruction. Figure 7 summarizes these effects in terms of standard deviations. Additional exploratory results shown in Appendix Figure 1 suggest that targeting is particularly important for SMS messages. We find evidence of limited effects for non-targeted SMS messages across all learning proficiencies, while SMS messages that are targeted appear to be more effective with average level learning gains of 0.079 standard deviations (p=0.157) as well as gains on broader competencies such as place value with 0.071 standard deviation gains (p=0.204) and higher-order competencies, such as fractions, with 0.080 standard deviation gains (p=0.147).[7]

Altogether, the results suggest that combined phone and SMS "low-tech" interventions can generate substantial learning gains, and that targeted interventions are more effective on a broader range of competencies, with particular importance for SMS messages. Learning gains in the phone and SMS group translate into 31 percent reductions in innumeracy.

To put these effect sizes in context, Kraft (2020) provides benchmarks based on a review of 1,942 effect sizes from 747 RCTs evaluating education interventions with standardized test outcomes. In this review, 0.10 is the median effect size. To this end, we find effect sizes that are around or above the median effect size with a relatively short, cheap, and scalable intervention.[8]

6.2 Robustness of Learning Results

We run a series of robustness tests on treatment effects on learning. The learning assessment used was adapted from the Annual Status of Education Report (ASER) which has been used consistently in the literature (Banerjee et al. 2007; Banerjee et al. 2010; Banerjee et al. 2017; Duflo et al. 2020) and is currently implemented routinely in 14 countries. We adapt this assessment into a phone-based assessment and incorporate time limits and a requirement that children explain their work to accurately identify their numeracy levels. We discuss practical steps to implement learning measurement via phone in Angrist et al. (2020a).

We conduct a series of robustness tests for phone-based learning measurement. First, we randomize problems which test the same proficiency, a version of a reliability test used in the psychometric literature called "parallel forms reliability" in constructing learning assessments (Crocker and Algina 1986). For example, for a subtraction problem, a random set of students will receive the question "83 - 45" whereas another random set of students will receive the question "72 - 18" to test the subtraction with borrowing proficiency. We randomize 5 problems for each proficiency including for addition, subtraction, multiplication, division, and fractions. Table 4 shows results.

[7] Moreover, we find that targeting phone calls is effective for learning higher-order competencies such as fractions with 0.106 standard deviation gains (p=0.056). These gains are relative to 0.045 standard deviations (p=0.420) for non-targeted phone calls. However, for more basic competencies, such as addition and subtraction reflected in "average levels", explicit targeting for phone calls is worse, with 0.072 standard deviation gains (p=0.193), relative to non-targeted instruction, with 0.168 standard deviation gains (p=0.002). This suggests that in the absence of explicit targeting, instructors might be more inclined to focus on basic proficiencies. With explicit targeting, this might provide a nudge to move on to higher competencies. Alternatively, this might suggest that instructors might be able to implicitly target using expert knowledge, and for certain competencies do so better than explicit targeting.

[8] Of note, the learning gains observed might be driven by either learning gains, minimizing learning loss, or a combination of both.

We find that each random problem across all proficiencies is not statistically significantly different relative to a base random problem.[9] Figure 8 shows results for one proficiency, addition, revealing that on average about 15 percent of students can do simple addition. Across all random problems there is some variance around this mean, but all confidence intervals overlap. These tests reveal that the phone-based learning assessment has a high level of internal reliability.

We further disentangle cognitive skills gains from effort effects, which have been shown to affect test scores (Gneezy et al. 2019). In our context, where learning outcomes are measured remotely in the household, effort might be particularly important. We test this hypothesis with a real-effort task. We ask "The day before two days from now is Saturday. What day is today?" This question largely requires one to spend time to think about the question and exert effort or motivation to answer it, rather than capture any substantive numerical proficiency. As shown in column (1) in Table 5, Around 29 percent of students are able to answer this question in the control group and we find that answering this question correctly is unaffected by any of the interventions. Column (2) through (4) contrast the lack of significant effect on effort with significant effects on learning. We see that the average level increases substantially. We observe that learning occurs across the learning spectrum, with fewer innumerate students and more students who can do division. Figure 9 summarizes the contrast in effort and learning effects. These results show that learning gains due to the intervention are largely a function of cognitive skill, rather than effort on the test.

It is also possible that learning gains are a matter of familiarity with the content in the intervention groups which receive exposure to similar material. The familiarity hypothesis is partially tested by randomizing problems of the same proficiency, since this exogenously varies the question asked to minimize overlap with any particular question asked during the intervention itself, and does not affect results. We also test this by including content not covered during the intervention, but which is related, such as fractions, and, as noted earlier, we find that in the phone and SMS group learning gains translate substantially to being able to solve fractions problems as shown in Figure 7.[10]

7 Mechanisms

[9] Relatedly, we find no difference in treatment effects by the random question received for each proficiency. Results available on request.

[10] We also examine the extent to which gains in the ASER test translate to fractions independent from the intervention as an additional assessment of validity of each respective learning measure. We would expect that as students are able to move up from addition to division they would also be more likely to be able to solve fractions. We find this is the case, with each step in the ASER assessment corresponding to 26 percentage points higher likelihood of being able to solve fractions.

7.1 Engagement and Demand

We explore parental educational engagement mechanisms. In Table 6, we see that parental engagement in their child's education is high with 92.1 percent of parents reporting their child attempted to solve any of the problems in the SMS only group, and slightly higher engagement of 95.2 percent in the phone call group. In the phone call treatment we have particularly granular data on week-on-week engagement defined as spending any time on the phone with the instructor. In Figure 10, we see that weekly engagement starts at around 85 percent and declines over time to 60 percent.[11] In addition, we find that while engagement overall declines over time, the *type* of engagement changes, with more parents spending longer on the phone. We see an increase in the number of minutes spent on educational content on the phone, with fewer lessons spanning less than ten minutes and more longer phone calls spanning more than ten minutes. This reveals that while there is slightly less engagement over time, the remaining engagement that does exist, which is still high at 60 percent in the final week, is also more intensive.

We further explore demand mechanisms in Table 6. Parents exhibit strong demand for the intervention, with over 99 percent of households expressing interest in continuing the program after four weeks. While the interventions do not affect demand on the extensive margin (desire for any low-tech service), likely since demand was already nearly 100 percent, they do affect demand by type of low-tech service. The most demanded service is a combination of phone calls and SMS messages (69.3 percent) followed by SMS-only (17.6 percent). Receiving both phone calls and SMS messages increased demand for this combined service substantially by 17.7 percentage points; the SMS-only group increased demand for SMS messages only by 7.7 percentage points. Results are also shown in Appendix Figure 2. This finding suggests that receiving an intervention, even when not the preferred intervention at the outset (as was the case for SMS-only), can increase subsequent demand. One reason this might be true is that families in the treated groups observed the benefits of their intervention only and thus demanded more of the intervention.

7.2 Parent Perceptions and Self-Efficacy

Previous research has shown that parents often misperceive their child's effort and learning, which can impede parents' support for their child's learning (Banerjee et al. 2010; Dizon-Ross 2019; Bergman 2020). Direct engagement by parents in their child's

[11] As a benchmark, phone-based response rates have been found to typically range around 50 percent or below. A World Bank survey in Sierra Leone during the Ebola response had a 51 percent response rate across three rounds (World Bank 2016).

learning might update parent beliefs and ameliorate misperceptions. It might also instill a sense of self-efficacy and enable greater parental investment (Hoover-Demspey and Sandler 1997).

We find that parents update their beliefs about their child's learning level in tandem with their child's learning progress. In Table 7 we see that in the SMS group, students learn, but only a little, and parent beliefs update marginally positively in tandem. In the phone and SMS group, students learn a substantial amount, and parent beliefs update significantly. This suggests that parents' beliefs are malleable and that they are involved and aware of their child's academic progress. We also find that parents are slightly more accurate than they were before in the phone and SMS group. This points to two important insights: first, more intensive involvement in a child's learning might be important for belief updating; second, despite substantial engagement in their child's education, parent's accuracy of beliefs updated only partially. This demonstrates the difficulty in identifying a child's learning level, which is consistent with the literature and reveals that experience might need to be supplemented with information to maximize belief updating (Banerjee et al. 2010; Hanna et al. 2014; Dizon-Ross 2019; Bergman 2020).

Figure 11 summarizes results on parent beliefs as well as reports results on self-efficacy and perceptions. Parents report 4.9 (p=0.021) and 8.6 (p<0.001) percentage points greater self-efficacy in supporting their child's learning in the SMS only and phone and SMS group, respectively. This level of self-efficacy is high. We also find parent confidence that their child made progress on their learning which ranges from 6.6 (p=0.002) to 10.5 (p<0.001) percentage points. Altogether, these results reveal that parental investments can play an important role in their child's education.

7.3 Other Outcomes and Potential for Crowd Out Effects

Parents who engage more in their children's learning might in turn displace other activities, such as returning to work when lockdowns were lifted. In Table 8, we find no evidence of such crowd out effects. Rather, we find a slight increase in return to work, with a reduction in parents who remain out of any type of work by 2.9 percentage points (p=0.092) in the phone and SMS group from a comparison of 19 percent unemployment in the control group. Any positive effect on employment could be for a number of reasons, including confidence in their child's learning progression and therefore more comfort returning to work or general self-efficacy translating into labor market outcomes. We do not focus on explaining these effects and instead the goal is to test concerns that further parental engagement in their child's education might crowd-out other activities, such as returning to work. The latter does not seem to be the case.

In terms of students returning to school, we find that there is no margin to affect school return: nearly all children (98 percent) return to school. This might be driven by the relatively short school closure period in Botswana. In other contexts, however, where school re-entry is lower, it is possible remote education might also mediate return to school.

7.4 Full versus Partial Substitution of Schooling

We explore the interaction between the low-tech interventions and the degree of school closure. The endline after four months of the intervention reveals that effects have longer-term persistence in the phone and SMS group. These results are particularly striking given schools partially reopened during this period, whereas in the first wave of data collection schools were fully closed. This suggests that the phone and SMS treatment was effective even when it served as a *partial substitute* to schooling, in addition to as *full substitutes* in the first stage of the trial. However, effects are slightly smaller in the second versus first stage of the trial. We see reductions in the phone and SMS treatment innumeracy of 52 percent relative to 31 percent in wave 1 versus wave 2 and of 34 percent relative to 11 percent for the targeted SMS only group. This translates to 0.121 standard deviation gains at endline relative to 0.235 standard deviation gains at midline for the phone and SMS group, and .024 standard deviation gains at endline relative to 0.120 standard deviations for the SMS only group at midline.

The difference in treatment effects between midline and endline could be for a few reasons. First, even if effects persist over time, they might diminish in size due to habituation or fatigue. Second, the interventions tested might be most effective as pure substitutes, since when schools are closed and virtually no learning takes place in the control group, these simple low-cost interventions might matter most. In contrast, when schools reopen, learning in the control group increases in line with typical learning trajectories in school, indicating that the low-tech interventions now provide a partial rather than full substitute.

An alternative explanation is that engagement simply drops over time, with average treatment effects staying constant, but intention-to-treat effects reducing due to lower engagement. We explore the degree to which effects between midline and endline vary due to a drop-off in engagement rather than a partial versus full substitution mechanism. We conduct a treatment-on-the-treated analysis in the phone and SMS groups where we have detailed week-by-week data on engagement. Engagement starts at 85 percent in the first week and decreases by the final week, although it still remains high at around 60 percent. We code a continuous treatment variable for the number of sessions attended and instrument this endogenous variable with treatment assignment.

In Table 9, we see that the standard deviation effects diminish between midline and endline comparing column (4) with .235 standard deviation gains and (1) with .121 standard deviation gains. However, for each session a household participates in, we see in column (2) learning gains are .028 (p=0.008). This translates into .0168 standard deviations gained in column (3) for households who participate in all sessions. These results reveal, first, that effects are more similar at endline and midline than they first appear, and, second, that results do not fully converge suggesting that engagement alone does not explain the difference between midline and endline results. Rather, it appears the intervention is more effective as a full substitute at midline, although it is still effective as a partial substitute at endline.

Altogether, the results suggest targeted SMS messages and phone calls are effective both as full substitutes when schools are closed as well as partial substitutes even as schools reopen. Moreover, a drop-off in engagement over time is to be expected even in our case where engagement between 60 to 85 percent is high relative to the literature. To this end, further experimentation to keep engagement high in low-tech interventions is likely to have high returns.

7 Cost-effectiveness

Both low-tech interventions are relatively low cost.[12] We estimate an upper bound on costs which includes programmatic costs, personnel time, as well as fixed costs to collect phone numbers, set up new infrastructure, conduct training, and collect routine monitoring data.[13] A portion of these costs are fixed costs, suggesting likely even lower costs at scale when considering economies of scale and that running costs largely consist of marginal costs rather than once-off fixed costs to set up the intervention.

For the SMS-only treatment arm, the total cost after four months was about $7,825 USD. For phone calls, the marginal cost above the bulk text message was $28,775. This equates to $5 per child reached in the SMS group and $19 dollars per child reached in the phone and SMS group. Given average treatment effects in the phone and SMS group of 0.12 standard deviations, this translates to 0.63 standard deviation

[12] Of note, more complex iterations of the low-tech interventions we tested could include two-way SMS text messages, providing airtime for parents to make calls to a hotline or call center, or interactive voice response (IVR). While these low-tech options are marginally higher cost, the binding constraint in Botswana to running these interventions was logistics rather than cost.

[13] We do not include costs to parents since no direct costs were incurred. This is because facilitators called parents directly, rather than parents calling facilitators and SMS messages were one-way, which drove costs to parents to zero. Moreover, in the current context where workplaces were largely closed there are minimal opportunity costs in terms of time.

gains for the phone and SMS group per $100 USD. For those that engage in all sessions of the program with a treatment effect of 0.17 standard deviations, this translates into .89 standard deviations gained per $100 USD.

These estimates are cost-effective relative to the literature. As a comparison, conditional cash transfers in Malawi yielded less than 0.1 standard deviation per $100; remedial tutoring in India yielded around a standard deviation per $100 (Kremer, Brannen, and Glennerster 2013).[14] Another relevant cost-effectiveness comparison are tutoring programs.[15] A recent review by Nickow, Oreopoulos, and Quan (2020) shows that tutoring programs have been consistently effective across 96 randomized trials with a pooled effect of 0.37 standard deviations. The phone call intervention in our trial compares closely to one of these tutoring programs which yielded .19 to .31 standard deviation learning gains and cost $2,500 per child (Cook et al. 2015). The phone and SMS intervention yields similar effects and is substantially cheaper.

These comparisons show that the low-tech interventions tested are cost-effective relative to other popular and cost-effective interventions in the education literature. We also translate results into a policy-relevant unit: high-quality years of education. We draw on the methodology proposed by Angrist et al. (2020b) to express learning gains in terms of Learning-Adjusted Years of Schooling (LAYS). The interventions tested translate into up to 1.1 years of schooling in a high-quality education system per $100.

8 Policy Implications

Our results have both immediate and long-term policy relevance. Over 1.6 billion children were out of school at the height of the pandemic (UNESCO 2020). These short-term shocks can have long-run consequences, with estimates suggesting school closures due to COVID-19 could cost up to 10 trillion in net present value (Azevedo et al. 2020).

Our results suggest that low-tech solutions can cost-effectively improve learning during school closures, with learning gains of 0.12 to 0.17 standard deviation gains via phone and SMS. We find that targeted instruction is similarly effective on basic numerical competencies, and is substantially more effective on a broader set of competencies, such as place value, and is three times more effective for higher-order competencies such as solving fractions. These learning gains translate into up to 1.1 years of high-quality schooling for $100 per child.

[14] We use estimates in terms of standard deviation gains per $100 in line with a cost-effectiveness review for education interventions by Kremer, Brannen, and Glennerster (2013) with detailed estimates also reported on the Jameel Poverty Action Lab (J-PAL) website.
[15] Carlana and La Ferrara (2020) evaluate remote tutoring with college students in Italy during covid-19 with results pending.

These low-tech solutions are cheap and feasible to deliver at scale. Both rely on phones and do not require internet access. While only 15 percent to 60 percent of households in low- and middle-income countries have internet access, 70 percent to 90 percent of households own at least one mobile phone (Carvalho and Crawford 2020). This high rate of access means these low-tech solutions have the potential to reach the masses in an era of unprecedented school closures, especially for low-resource families with limited access to the internet and alternative sources of learning at home.

Many governments have dedicated funding for Information and Communications (ICT), often including tablets and computers for education (World Bank 2018). For example, the Ministry of Basic Education in Botswana recently allocated a new line item towards ICT solutions in the most recent budget speech. The World Bank has highlighted countries with large-scale education technology projects such Kenya, Uruguay, Thailand, Peru, Rwanda, Turkey, India, and Argentina (Trucano 2013). The Government of Kenya reportedly spent over $600 million on computers and tablets (Odhiambo 2019). Many governments already invest in ICT approaches and could leverage existing budget allocations to scale low-tech solutions to promote learning.

A policy or scale-up that builds on the insights from this trial could include any combination of national SMS schemes and targeted phone call campaigns. One-way SMS text messages are feasible to implement at large scale in most countries using bulk texting platforms. SMS messages could be targeted by teachers or a central platform using prior data on student learning assessments. Direct phone calls at scale might consist of weekly teacher phone calls to the bottom 5 to 10 percent of their class to support students furthest behind.

Of note, the implementation of the program is feasible to deploy relatively quickly. Within six weeks, we conceived and rolled out a program to over half of the regions of Botswana, including phone number collection, pilots, and program design, setting up texting and calling infrastructure, and training. A caveat is that we had a team of over 60 facilitators who could be readily deployed to make phone calls. In the absence of this ready workforce, more time, cost and effort would have been needed to recruit staff. While this might be challenging, many governments already employ teachers at national scale who could conduct direct phone calls.

Our results further reveal the potential for parents to play a larger role in their child's education. Prior research has shown the parents serve as effective complements to school inputs, providing motivation and accountability to the traditional schooling system. We find that parents, with light additional support, can partially substitute schooling by serving as at-home teachers. This includes parents in both rural and urban communities and with limited to no formal teacher training. This suggests potential for greater parent-teacher interaction around a child's education. Many schemes exist to facilitate parent and teacher interaction in school systems worldwide already, such as report cards and parent-teacher associations (PTAs). Our results suggest these built-in interactions in low- or middle-income country contexts – which often focus on providing information on the child's performance – might be substantially enhanced with simple learning content that parents can directly engage their child in at home.

Our results also have implications for school closure beyond the current pandemic. School closures occur during annual school holidays, other public health crises, natural disasters, during weather-related shocks and in refugee and conflict settings. To this end, methods to substitute school when schools are closed are needed. They might also add-value as complements when schools are open.

9 Conclusion

This paper provides some of the first experimental estimates on minimizing the fallout of the COVID-19 pandemic on learning. We find that low-tech phone calls and SMS interventions have large and cost-effective effects on household engagement in education and learning during full and partial school closures. We find up to 0.12 to 0.17 standard deviation gains. In terms of cost-effectiveness, we estimate up to 1.1 years of high-quality schooling can be gained for $100. We also find that targeted interventions outperform non-targeted interventions, in particular for SMS interventions. This finding suggests that mobile phones provide a cheap and scalable way to target instruction, an approach shown to produce cost-effective learning gains in classroom-based models. We find learning gains are robust to a variety of novel phone-based robustness tests, including randomized problems across the same proficiency and differentiating effort from cognitive skills with real-effort tasks. We further find that gains persist in the phone and SMS treatment across multiple waves of assessment, even as schools partially reopen.

In terms of mechanisms, we find high parental engagement in educational activities with their children, high demand, and greater self-efficacy to support their child's learning, as well as partial gains in accurate perceptions of their child's level. This finding reveals that parental investments in education can improve their child's learning outcomes even in a low-literacy context.

Our results reveal promise for low-tech interventions that are relatively cheap and easy to scale. Of note, while our results are promising, follow-on trials will be important to adapt these low-tech interventions across contexts. More advanced low-tech interventions could be tested such as two-way texts which might enable hyper-adaptive intervention. Follow-on research might also explore scalability through government implementation by teachers, other low to medium-tech learning options such as radio and television, as well as disentangling the mechanisms surrounding take-up, which is a first-order issue for out-of-school interventions. Follow-on trials could be implemented in a rapid and adaptive approach, generating real-time data to optimize and improve interventions as they scale. Future trials might also explore the implications of these low-tech interventions as both complements and substitutes of the traditional schooling system depending on whether schools are closed or open.

The results in this trial have immediate implications for global policy during the current school disruptions, revealing cost-effective and scalable approaches to stem learning loss during the pandemic. Moreover, school closures occur in settings beyond the current pandemic, including summer holidays, public health crises, during adverse weather events, natural disasters, and in refugee and conflict settings. In moments where a substitute for schooling is needed, particularly for families with fewer resources at home, the low-tech solutions tested in this trial have unique potential to reach the masses. To this end, the results from this trial have long-run implications for the role of technology and parents as substitutes or complements to the traditional schooling system.

References

Aker, Jenny C., Christopher Ksoll, and Travis J. Lybbert. 2012. "Can mobile phones improve learning? Evidence from a field experiment in Niger." *American Economic Journal: Applied Economics* 4, no. 4: 94-120.

Aker, Jenny C., and Christopher Ksoll. 2020. "Can ABC lead to sustained 123? The medium-term effects of a technology-enhanced adult education program." *Economic Development and Cultural Change* 68, no. 3: 1081-1102.

Andrabi, Tahir, Benjamin Daniels, and Jishnu Das. 2020. "Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005." RISE Working Paper.

Angrist, Noam, Simeon Djankov, Pinelopi Koujianou Goldberg, and Harry Anthony Patrinos. 2019. "Measuring Human Capital." The World Bank.

Angrist, Noam, Peter Bergman, David K. Evans, Susannah Hares, Matthew CH Jukes, and Thato Letsomo. 2020a. "Practical lessons for phone-based assessments of learning." *BMJ Global Health* 5, no. 7: e003030.

Angrist, Noam, David K. Evans, Deon Filmer, Rachel Glennerster, F. Halsey Rogers, and Shwetlena Sabarwal. 2020b. "How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric." The World Bank.

Azevedo, João Pedro, Amer Hasan, Diana Goldemberg, Syedah Aroob Iqbal, and Koen Geven. 2020. "Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes." The World Bank.

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern. 2020. "Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time." *Journal of Public Economics*, forthcoming.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying education: Evidence from two randomized experiments in India." *The Quarterly Journal of Economics* 122, no. 3 (2007): 1235-1264.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India." *American Economic Journal: Economic Policy* 2, no. 1: 1-30.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From proof of concept to scalable policies: Challenges and solutions, with an application." *Journal of Economic Perspectives* 31, no. 4: 73-102.

Bergman, Peter, and Eric W. Chan. 2019. “Leveraging parents through low-cost technology: The impact of high-frequency information on student achievement.” *Journal of Human Resources*: 1118-9837R1.

Bergman, Peter. 2019. “How behavioral science can empower parents to improve children's educational outcomes.” *Behavioral Science & Policy* 5, no. 1: 52-67.

Bergman, Peter. 2020. “Parent-child information frictions and human capital investment: Evidence from a field experiment.” *Journal of Political Economy,* forthcoming.

Berlinski, Samuel, Matias Busso, Taryn Dinkelman, and Claudia Martinez. 2016. “Reducing parent-school information gaps and improving education outcomes: Evidence from high frequency text messaging in Chile.” Manuscript.

Bettinger, Eric, Robert W. Fairlie, Anastasia Kapuza, Elena Kardanova, Prashant Loyalka, and Andrey Zakharov. 2020. “Does edtech substitute for traditional learning? Experimental estimates of the educational production function” National Bureau of Economic Research Working Paper No. w26967.

Carlana, Michela and Eliana La Ferrara. 2020. “Apart but Connected: Online Tutoring to Mitigate the Impact of COVID-19 on Educational Inequality.” Presented at Virtual Development Economics Seminar Series.

Carvalho, Shelby and Lee Crawfurd, “School's Out: Now What?” *Center for Global Development* (blog), 25 March, 2020. https://www.cgdev.org/blog/schools-out-now-what

Chetty, Raj, John N. Friedman, Nathaniel Hendren, and Michael Stepner. 2020. “How did COVID-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data.” National Bureau of Economic Research Working Paper No. w27431.

Cook, Philip J., Kenneth Dodge, George Farkas, Roland G. Fryer, Jonathan Guryan, Jens Ludwig, Susan Mayer, Harold Pollack, and Laurence Steinberg. 2015. “Not too late: Improving academic outcomes for disadvantaged youth.” *Institute for Policy Research Northwestern University Working Paper WP-15* 1.

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse. 1996. “The effects of summer vacation on achievement test scores: A narrative and meta-analytic review.” *Review of Educational Research* 66, no. 3: 227-268

Crocker, Linda M., and James Algina. *Introduction to Classical and Modern Test Theory.* Fort Worth: Holt, Rinehart, and Winston, 1986

Cunha, Nina, Guilherme Lichand, Ricardo Madeira, and Eric Bettinger. 2017. “What is it about communicating with parents.” CEPA Working Paper.

De Walque, Damien, and Christine Valente. 2018. “Incentivizing school attendance in the presence of parent-child information frictions.” The World Bank.

Dizon-Ross, Rebecca. 2019. “Parents' beliefs about their children's academic ability: Implications for educational investments.” *American Economic Review* 109, no. 8: 2728-65.

Doss, Christopher, Erin M. Fahle, Susanna Loeb, and Benjamin N. York. 2019. “More Than Just a Nudge Supporting Kindergarten Parents with Differentiated and Personalized Text Messages.” *Journal of Human Resources* 54, no. 3: 567-603.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. “Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya.” *American Economic Review* 101, no. 5: 1739-74.

Duflo, Annie, Jessica Kiessel, and Adrienne Lucas. 2020. “External Validity: Four Models of Improving Student Achievement.” National Bureau of Economic Research No. w27298.

Engzell, Per, Arun Frey, and Mark D. Verhagen. 2020. “Learning inequality during the COVID-19 pandemic.” SocArXiv. October 29.

Gallego, Francisco A., Ofer Malamud, and Cristian Pop-Eleches. 2020. “Parental monitoring and children's internet use: The role of information, control, and cues.” *Journal of Public Economics* 188: 104208.

Gneezy, Uri, John A. List, Jeffrey A. Livingston, Xiangdong Qin, Sally Sadoff, and Yang Xu. 2019. “Measuring success in education: the role of effort on the test itself.” *American Economic Review: Insights* 1, no. 3: 291-308

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. “Learning through noticing: Theory and evidence from a field experiment.” *The Quarterly Journal of Economics* 129, no. 3: 1311-1353.

Hanushek, Eric and Ludger Woessmann. 2020. “The economic impacts of learning losses.” OECD Education Working Papers, No. 225, OECD Publishing, Paris.

Hoover-Dempsey, Kathleen V., and Howard M. Sandler. 1997. “Why do parents become involved in their children’s education?.” *Review of educational research* 67, no. 1: 3-4

Jaume, David, and Alexander Willén. 2019. “The long-run effects of teacher strikes: evidence from Argentina.” *Journal of Labor Economics* 37, no. 4: 1097-1139.

Kraft, Matthew A., and Todd Rogers. 2015. “The underutilized potential of teacher-to-parent communication: Evidence from a field experiment.” *Economics of Education Review* 47: 49-63.

Kraft, Matthew A. 2020. “Interpreting effect sizes of education interventions.” *Educational Researcher* 49, no. 4: 241-253.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. “The challenge of education and learning in the developing world.” *Science* 340, no. 6130: 297-300.

Ksoll, Christopher, Jenny C. Aker, Danielle Miller, Karla Perez, and Susan L. Smalley. 2015. "Learning without Teachers? Evidence from a Randomized Experiment of a Mobile Phone-Based Adult Education Program in Los Angeles." CGD Working Paper.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review* 109, no. 4: 1426-60.

Musaddiq, Tareena, Alexa Prettyman, and Jonathan Smith. 2019. "School Messaging Platforms and Student Attendance." Policy Brief.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. 2020. "The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence." *National Bureau of Economic Research Working Paper Series* w27476.

Odhiambo, Moses. "State to spend Sh8.4bn more on school tablets" *The Star,* February 16, 2019. https://www.the-star.co.ke/news/2019-02-16-state-to-spend-sh84bn-more-on-school-tablets/

Siebert, Stanley, Ho Lun Wong, Xiangdong Wei, and Xiang Zhou. 2018. "Student feedback, parent-teacher communication, and academic performance: Experimental evidence from rural China." IZA working paper N. 11347.

Slade, Timothy S., Benjamin Piper, Zikani Kaunda, Simon King, and Hibatalla Ibrahim. 2017. "Is 'summer' reading loss universal? Using ongoing literacy assessment in Malawi to estimate the loss from grade-transition breaks." *Research in Comparative and International Education* 12, no. 4: 461-485.

Trucano, Michael. "A new wave of educational efforts across Africa exploring the use of ICTs." *World Bank blog.* 04 April, 2013. https://blogs.worldbank.org/edutech/new-wave-in-africa-maybe

UNESCO. 2020. Retrieved from https://en.unesco.org/covid19/educationresponse.

Winthrop, Rebecca, Mahsa Ershadi, Noam Angrist, Efua Bortsie and Moitshepi Matsheng. 2020. "A historic shock to parental engagement in education: parent perspectives in Botswana during COVID-19." The Brookings Institution.

World Bank. 2018. "Information and Communications for Development 2018: Data-Driven Development." The World Bank.

World Bank 2016. "Sierra Leone - High Frequency Cell Phone Survey on the Socio-Economic Impacts of Ebola 2014-2015." World Bank Microdata Library.

York, Benjamin N., Susanna Loeb, and Christopher Doss. 2018. "One Step at a Time: The Effects of an Early Literacy Text-Messaging Program for Parents of Preschoolers." *Journal of Human Resources* 54(3):537–66.

Figure 1: Intervention SMS Text Message Example

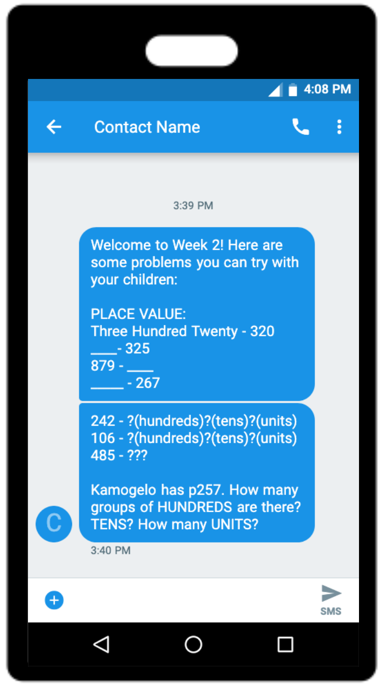
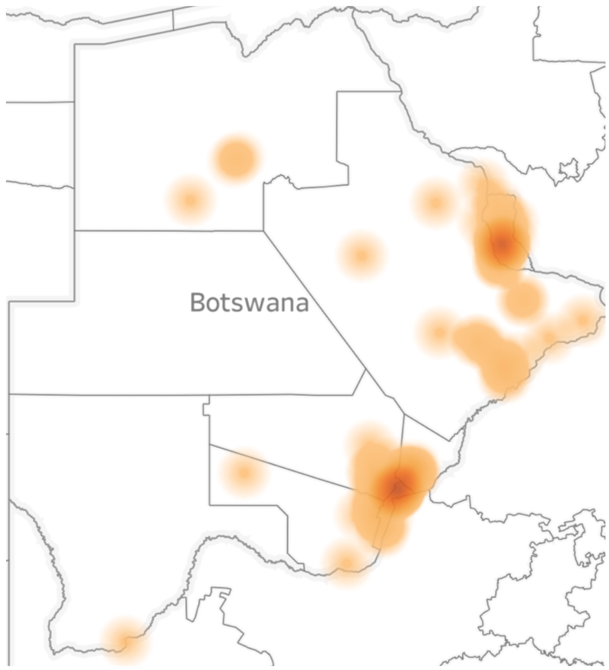


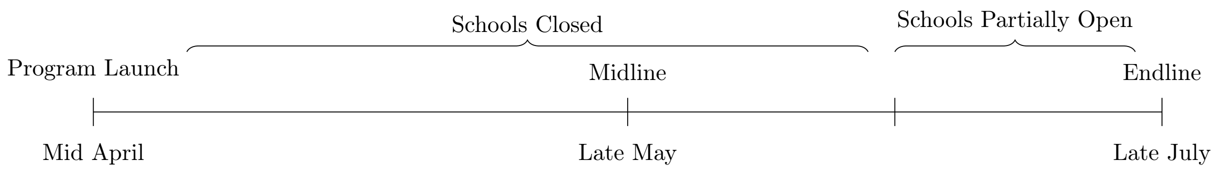
Figure 2: Sample phone call introduction

Figure 3: Distribution of Schools of Student Participants across Botswana



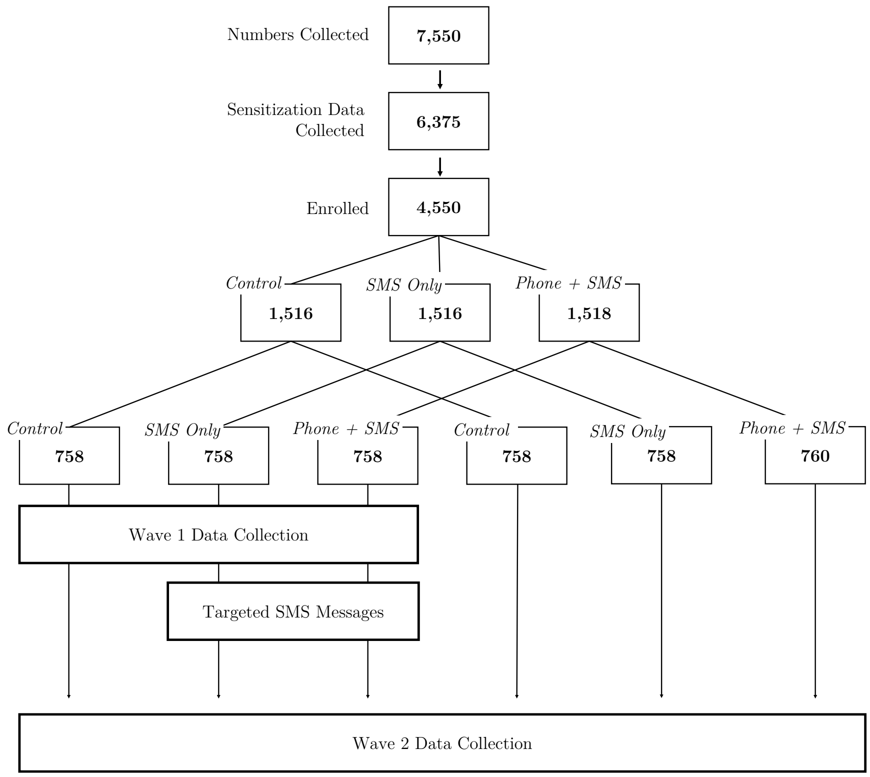
Notes: this density map of schools in Botswana shows the relative distribution of schools linked to students in our sample. Darker regions correspond to higher concentrations of schools for study participants. The sample in the study includes nearly all regions in Botswana (9 out of 10).

Figure 4: Intervention and Evaluation Timeline



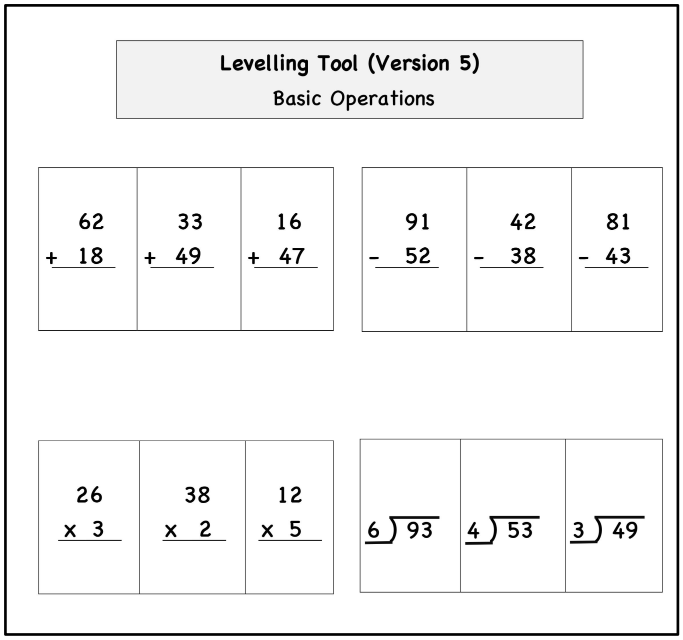
Notes: All dates refer to the year 2020.

Figure 5: Experimental Design



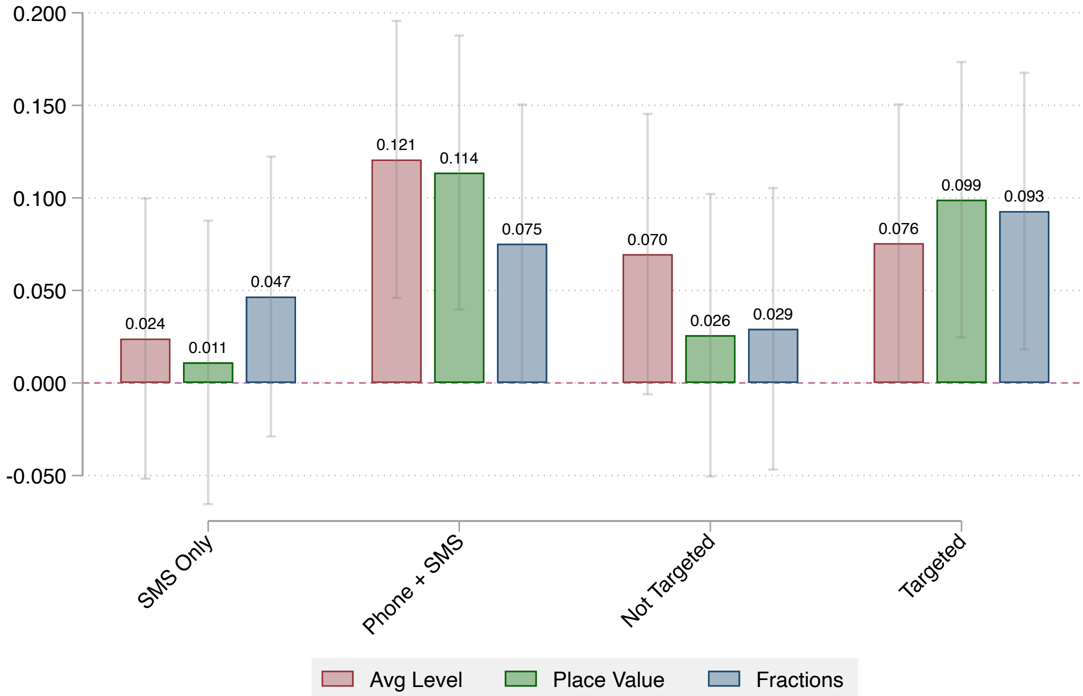
Notes: Counts represent the quantity of phone numbers. Each phone number corresponds to one household.

Figure 6: Sample of ASER test used in Botswana



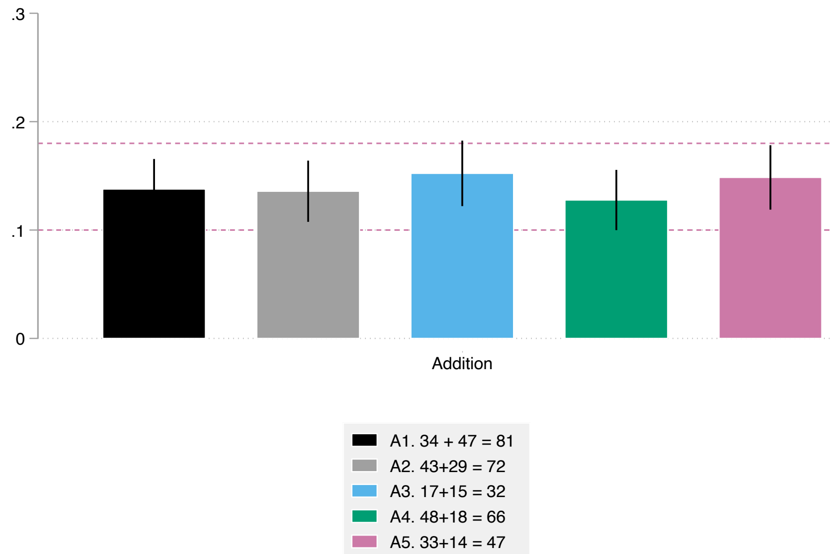
Notes: The ASER assessment was pioneered in India and has since been adapted to 14 countries all over the world. This includes a related assessment called Uwezo in East Africa and a global coordinating body called the People's Action for Learning (PAL) network.

Figure 7: Learning Treatment Effects



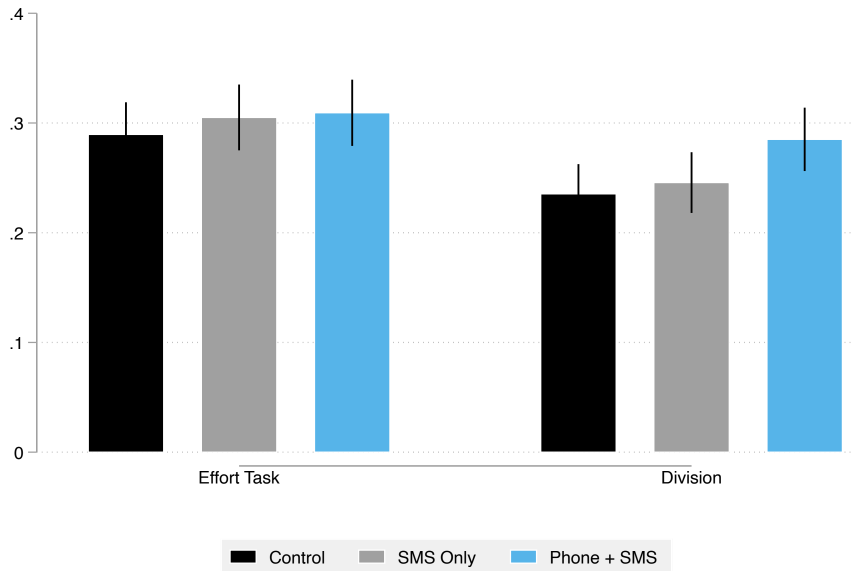
Notes: This figure shows treatment effects relative to the control group with 90 percent confidence interval bars. Effects are expressed in terms of standard deviations for comparable units. Each color bar represents a distinct learning question. "Average Level" reports skill on the ASER 0 to 4 scale corresponding to no operations, addition, subtraction, multiplication, and division. "Place Value" refers to a distinct place value problem, and "Fractions" refers to a distinct question asking students to solve a fractions problem. Each group "SMS Only", "Phone + SMS", "Not Targeted", and "Targeted" refer to randomized treatment groups pooled across the designated category.

Figure 8: Reliability of Learning Measure - Random questions



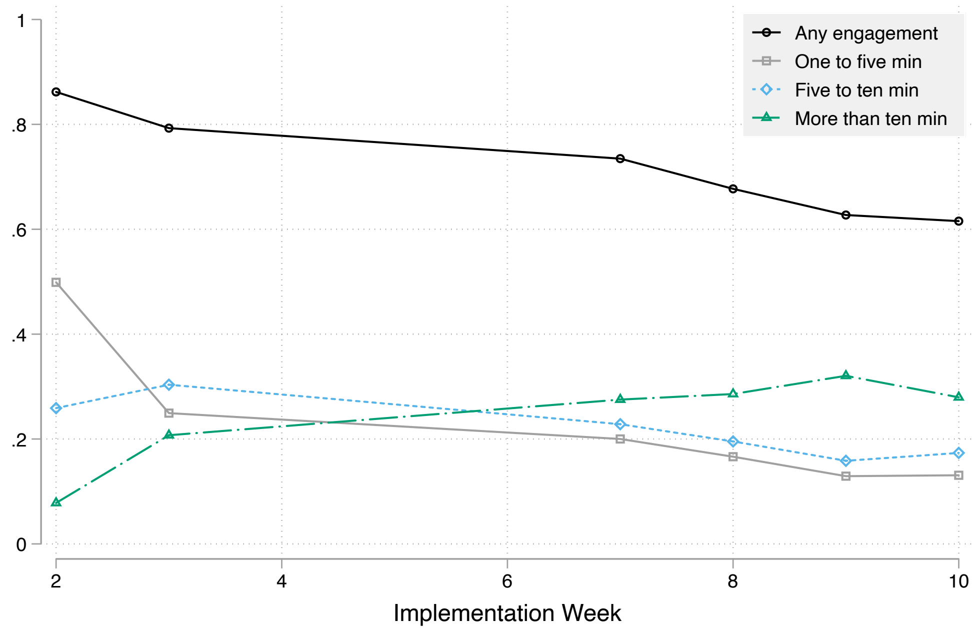
Notes: This figure shows the average percent of students who can do addition across random groups of students who received different problems of the same proficiency with 90 percent confidence interval bars. The proficiency shown is two-digit addition with carryover. Each proficiency (e.g., subtraction, multiplication, division, fractions) is estimated in Table 4 and similar figures are available on request.

Figure 9: Effort or Cognitive Skill



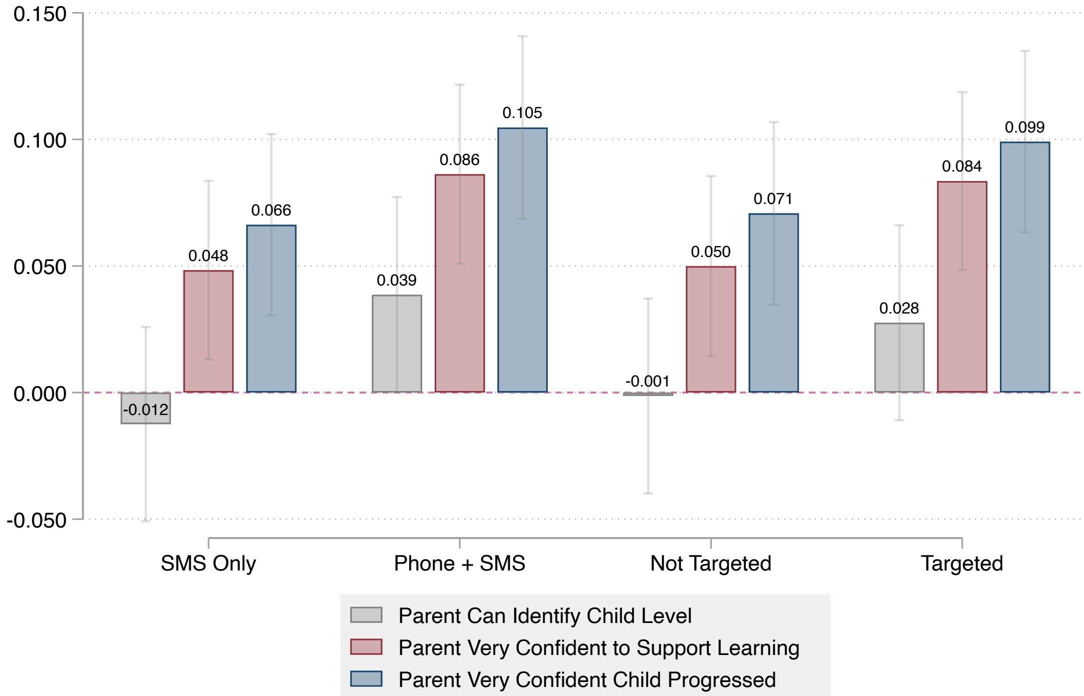
Notes: This figure shows the share of students who can do division and the share of students who correctly answer the real-effort task across treatment groups with 90 percent confidence interval bars.

Figure 10: Week on Week Engagement in the Phone and SMS Treatment



Notes: This figure shows the average percent of households who picked up the phone and engaged in a given week in the phone and SMS treatment group. Data collection occurred between 4 and 6 hence missing values for the intervention. The number of minutes refers to time spent on content instruction (not logistics).

Figure 11: Parent Beliefs, Perceptions, and Self-Efficacy



Notes: This figure shows treatment effects on parent accuracy of their child's learning level, their self-efficacy to support their child's learning, and their belief that their child made progress in learning in general, across treatment groups with 90 percent confidence interval bars.

Table 1: Attrition

	(1) Phone Call Response	(2) Place Value Response	(3) Avg Level Response
Panel A			
SMS Only	-0.004	-0.010	-0.008
	(0.017)	(0.018)	(0.018)
	[0.811]	[0.565]	[0.647]
Phone + SMS	0.004	-0.004	-0.002
	(0.017)	(0.017)	(0.018)
	[0.819]	[0.821]	[0.911]
Panel B			
Not Targeted	0.001	-0.006	-0.002
	(0.017)	(0.017)	(0.018)
	[0.949]	[0.726]	[0.903]
Targeted	-0.001	-0.008	-0.008
	(0.017)	(0.017)	(0.018)
	[0.939]	[0.651]	[0.654]
Control Mean	0.649	0.638	0.622
Strata Fixed Effects	Yes	Yes	Yes
Observations	4550	4550	4550
p-val: SMS = Phone	0.640	0.727	0.730
p-val: Targeted	0.889	0.918	0.744

Notes: This table reports attrition on endline survey response rates for three indicators: whether households picked up the phone to respond to the survey, if their child conducted a learning assessment for the place value question, and if their child conducted a learning assessment across four basic numeracy options: addition, subtraction, multiplication and division (for which we report the average level on a scale of 0-4). Standard errors are in parentheses and p-values are in square brackets.

Table 2: Balance

	(1) Child Grade	(2) Child Female	(3) Child Age	(4) Parent	(5) School Pass Rate
Panel A					
SMS Only	0.000	0.014	0.018	0.012	-0.001
	(0.034)	(0.022)	(0.067)	(0.014)	(0.006)
	[0.999]	[0.531]	[0.784]	[0.393]	[0.859]
Phone + SMS	0.033	0.027	0.016	0.010	0.002
	(0.034)	(0.022)	(0.064)	(0.014)	(0.006)
	[0.336]	[0.235]	[0.808]	[0.497]	[0.713]
Panel B					
Not Targeted	0.001	0.032	0.001	0.008	0.004
	(0.034)	(0.022)	(0.064)	(0.014)	(0.006)
	[0.970]	[0.158]	[0.994]	[0.585]	[0.496]
Targeted	0.032	0.009	0.034	0.014	-0.003
	(0.034)	(0.022)	(0.067)	(0.014)	(0.006)
	[0.354]	[0.688]	[0.618]	[0.323]	[0.643]
Control Mean	4.030	0.505	9.680	0.807	0.796
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	3014	2987	3008	4523	2394
p-val: SMS = Phone	0.338	0.571	0.967	0.862	0.585
p-val: Targeted	0.381	0.312	0.619	0.657	0.251

Notes: This table reports balance on survey responses for multiple demographic characteristics (student grade, age and sex), the identity of the household caregiver in each treatment (parent or another caregiver such as grandparent, aunt or uncle, cousin or sibling) and baseline school-level pass rates for schools we are able to link to students in the sample using administrative data from the Botswana Examinations Council (BEC) on the Primary School Leaving Examination (PSLE). Standard errors are in parentheses and p-values are in square brackets.

Table 3: Learning

	(1) Avg Level	(2) Place Value	(3) Fractions
Panel A			
SMS Only	0.024	0.009	0.047
	(0.046)	(0.046)	(0.046)
	[0.602]	[0.837]	[0.309]
Phone + SMS	0.121	0.114	0.075
	(0.046)	(0.044)	(0.046)
	[0.008]	[0.009]	[0.100]
Panel B			
Not Targeted	0.070	0.026	0.029
	(0.046)	(0.045)	(0.046)
	[0.130]	[0.572]	[0.527]
Targeted	0.076	0.098	0.093
	(0.046)	(0.044)	(0.045)
	[0.097]	[0.026]	[0.041]
Control Mean	1.974	1.774	1.605
Strata Fixed Effects	Yes	Yes	Yes
Observations	2815	2881	2751
p-val: SMS = Phone	0.033	0.017	0.528
p-val: Targeted	0.896	0.098	0.160

Notes: This table reports results on student learning assessment using three learning constructs in terms of standard deviations. Average level refers to how a child scores on four basic numeracy options: no operations correct, addition, subtraction, multiplication and division (for which we report the average level on a scale of 0-4). Place value refers to a distinct place value question. Fractions refers to a distinct question to solve a higher-order fractions problems. Each panel reports separate models which pool treatment groups by category. Standard errors are in parentheses and p-values are in square brackets.

Table 4: Robustness Check: Random Problem

	(1) Addition	(2) Subtraction	(3) Multiplication	(4) Division	(5) Fractions
Random Problem 2	-0.002	0.024	0.017	-0.039	0.017
	(0.020)	(0.024)	(0.028)	(0.025)	(0.026)
	[0.938]	[0.316]	[0.530]	[0.124]	[0.501]
Random Problem 3	0.014	0.007	-0.004	-0.008	-0.023
	(0.021)	(0.024)	(0.028)	(0.026)	(0.027)
	[0.512]	[0.765]	[0.895]	[0.765]	[0.400]
Random Problem 4	-0.011	0.036	-0.044	0.005	-0.008
	(0.020)	(0.024)	(0.027)	(0.026)	(0.026)
	[0.599]	[0.145]	[0.101]	[0.858]	[0.753]
Random Problem 5	0.010	0.005	-0.011	0.002	-0.032
	(0.021)	(0.024)	(0.027)	(0.026)	(0.027)
	[0.631]	[0.849]	[0.681]	[0.951]	[0.228]
Observations	2815	2815	2815	2815	2751
F-test: equivalence across all problems	0.715	0.458	0.139	0.307	0.498

Notes: This table reports results from a regression estimating differences in average proficiency across four randomly assigned problems relative to a base random problem for the following proficiency: addition, subtraction, multiplication, division and fractions. For example, for a subtraction problem, a random fifth of students will receive the question "83 - 45" whereas another random fifth of students will receive the question "72 - 18" to test the subtraction with borrowing proficiency, and so forth, across five random problems total for each proficiency. Standard errors are in parentheses and p-values are in square brackets.

Table 5: Robustness Check: Effort on the Test

	Effort	Learning		
	(1)	(2)	(3)	(4)
	Effort Task	Avg Level	Innumerate	Division
SMS Only	0.016	0.030	-0.010	0.011
	(0.021)	(0.057)	(0.013)	(0.020)
	[0.448]	[0.602]	[0.460]	[0.594]
Phone + SMS	0.021	0.150	-0.029	0.050
	(0.021)	(0.057)	(0.012)	(0.020)
	[0.335]	[0.008]	[0.022]	[0.013]
Control Mean	0.290	2.459	0.093	0.235
Strata Fixed Effects	Yes	Yes	Yes	Yes
Observations	2732	2815	2815	2815
p-val: SMS = Phone	0.839	0.033	0.121	0.053

Notes: This table reports results of differences across treatment groups relative to a control on a real-effort task. Effort is contrasted with results on learning, including average learning level as well as learning gains broken down by the lower end (innumerate) and the upper end (learning division). Standard errors are in parentheses and p-values are in square brackets.

Table 6: Mechanisms: Engagement and Demand

	Engaged	Demand		
	(1)	(2)	(3)	(4)
	Did Problems	Phone and SMS	SMS Only	None
SMS Only	0.921	-0.027	0.077	-0.005
	(0.009)	(0.030)	(0.026)	(0.005)
	[0.000]	[0.363]	[0.003]	[0.322]
Phone + SMS	0.952	0.177	-0.102	0.003
	(0.007)	(0.026)	(0.021)	(0.007)
	[0.000]	[0.000]	[0.000]	[0.639]
Control Mean	0.000	0.693	0.176	0.009
Observations	3405	1478	1478	1478
p-val: SMS = Phone	0.005	0.000	0.000	0.139

Notes: This table reports results of differences across treatment groups relative to a control on engagement questions at endline and demand at midline. We code engagement at zero for the control group since by definition there were no problems sent to respond to. For demand, we report demand at midline since this question was asked at the halfway point, with particular emphasis on demand for the interventions even if schools were to re-open. The observation count is lower for demand since a random subset of households received the midline. Standard errors are in parentheses and p-values are in square brackets.

Table 7: Mechanisms: Parent Beliefs and Investments

	(1) Reported Level	(2) Correct Level	(3) Self Efficacy	(4) Student Progressed
Panel A				
SMS Only	0.025	-0.012	0.049	0.066
	(0.050)	(0.023)	(0.021)	(0.022)
	[0.621]	[0.594]	[0.023]	[0.002]
Phone + SMS	0.153	0.039	0.086	0.105
	(0.050)	(0.023)	(0.021)	(0.022)
	[0.002]	[0.099]	[0.000]	[0.000]
Panel B				
Not Targeted	0.050	-0.001	0.050	0.071
	(0.051)	(0.023)	(0.022)	(0.022)
	[0.323]	[0.957]	[0.020]	[0.001]
Targeted	0.125	0.028	0.084	0.099
	(0.049)	(0.023)	(0.021)	(0.022)
	[0.012]	[0.239]	[0.000]	[0.000]
Control Mean	2.500	0.398	0.566	0.492
Strata Fixed Effects	Yes	Yes	Yes	Yes
Observations	2957	2650	3127	3127
p-val: SMS = Phone	0.009	0.029	0.071	0.075
p-val: Targeted	0.128	0.217	0.115	0.194

Notes: This table reports treatment effects relative to a control group on parent accuracy of their child's learning level, their self-efficacy to support their child's learning, and their belief that their child made progress in learning in general, across treatment groups. Standard errors are in parentheses and p-values are in square brackets.

Table 8: Mechanisms: Potential for Crowd Out

	(1) Return Full-Time Work	(2) Did not Return to Work
Panel A		
SMS Only	-0.010	-0.000
	(0.020)	(0.018)
	[0.622]	[0.994]
Phone + SMS	0.031	-0.029
	(0.020)	(0.017)
	[0.116]	[0.092]
Panel B		
Not Targeted	-0.002	-0.010
	(0.020)	(0.018)
	[0.925]	[0.565]
Targeted	0.021	-0.018
	(0.019)	(0.017)
	[0.276]	[0.296]
Control Mean	0.735	0.190
Strata Fixed Effects	Yes	Yes
Observations	2990	2990
p-val: SMS = Phone	0.037	0.088
p-val: Targeted	0.234	0.640

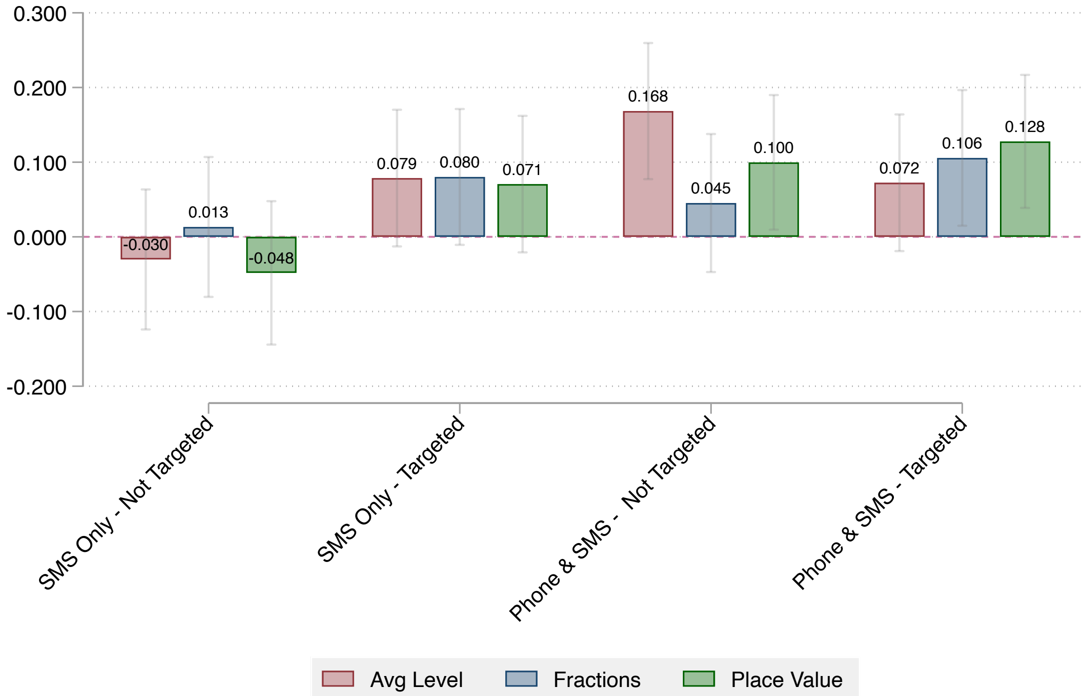
Notes: This table shows treatment effects on parent labor market outcomes in the form of returning to work post lockdown across treatment groups. Options for return to work included: returned to work full-time, returned to work part-time, retired, or unemployed. Standard errors are in parentheses and p-values are in square brackets.

Table 9: Mechanisms: Partial vs Full School Substitution

	Endline ITT	Endline TOT		Midline
	(1)	(2)	(3)	(4)
	Avg Level	Avg Level	Avg Level	Avg Level
SMS Only	0.024			0.120
	(0.046)			(0.070)
	[0.600]			[0.088]
Phone + SMS	0.121			0.235
	(0.046)			(0.070)
	[0.008]			[0.001]
Phone + SMS - Per Session		0.028		
		(0.010)		
		[0.007]		
Phone + SMS - All Sessions			0.167	
			(0.062)	
			[0.007]	
Observations	2815	1878	1878	1127

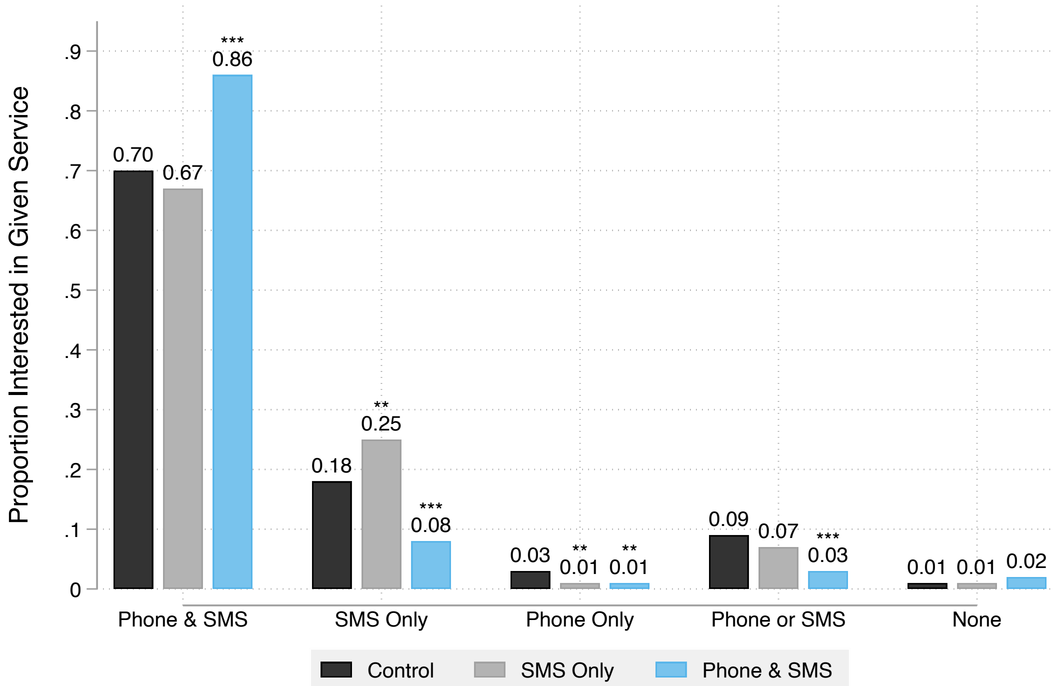
Notes: This table shows treatment effects in terms of standard deviations. Column (1) reports intention-to-treat (ITT) effects at endline. Column (2) reports treatment-on-the-treated (TOT) using instrumental variables estimation with random assignment to the Phone and SMS group as an instrument for a continuous measure of participation per session in the Phone and SMS group. Column (3) reports extrapolated treatment-on-the-treated (TOT) estimates in the Phone and SMS group if households attended all sessions. We do not have similarly rich week-by-week implementation data in the SMS group to conduct a meaningful TOT analysis. The observation count is lower in Columns (2)-(4) than Column (1) since we exclude the SMS group in the regression. Column (4) has a lower observation count since we report midline intention-to-treat effects where only a random subset of households were surveyed. Standard errors are in parentheses and p-values are in square brackets.

Appendix Figure 1: Learning Treatment Effects by Subgroup



Notes: This figure shows treatment effects relative to the control group with 90 percent confidence interval bars. Effects are expressed in terms of standard deviations for comparable units. Each color bar represents a distinct learning question. “Average Level” reports skill on the ASER 0 to 4 scale corresponding to no operations, addition, subtraction, multiplication, and division. “Place Value” refers to a distinct place value problem, and “Fractions” refers to a distinct question asking students to solve a fractions problem. Each group “SMS Only - Not Targeted”, “SMS Only – Targeted”, “Phone + SMS – Not Targeted”, “Phone + SMS – Targeted” refer to randomized treatment groups across the designated category.

Appendix Figure 2: Demand for Low-tech Services



Notes: This figure reports the proportion of households interested in a given service (e.g., demand) at midline by treatment groups and by service. We report demand at midline since this question was asked at the halfway point, with particular emphasis on demand for the interventions even if schools were to reopen. Stars denote statistical significance in relation to the control mean. *** p<0.01, ** p<0.05, * p<0.1.