Stemming Learning Loss During the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention in Botswana

Noam Angrist, Peter Bergman, Caton Brewster, and Moitshepi Matsheng*

August 2020

Abstract

The COVID-19 pandemic has closed schools 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 education outcomes. 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 at four- to six-week intervals. We present results from the first wave. We find early evidence that both interventions result in cost-effective learning gains of 0.16 to 0.29 standard deviations. This translates to a reduction in innumeracy of up to 52 percent. We show these results broadly hold with a series of robustness tests that account for differential attrition. We find increased parental engagement in their child's education and more accurate parent perceptions of their child's learning. In a second wave of the trial, we provide targeted instruction, customizing text messages to the child's learning level using data from the first wave. 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. We thank Jenny Aker, Jim Berry, Alex Eble, Michael Kremer, Clare Leaver, Susanna Loeb, Todd Rogers, Anna Rudge and participants of the University of Oxford development economics workshop for helpful comments on the study design and early results. The intervention and rapid trial were the product of a collaboration between the Botswana Ministry of Basic Education and staff at Young 1ove who nimbly adapted during school closure to collect phone numbers and deliver low-tech 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, and Rumbidzai Madzuzo, who coordinated and designed the low-tech programs, Patience Derera for carefully compiled cost estimates, Shawn Maruping and Dorothy Okatch for online communications, and Winnifred Arthur, Bonno Balopi, Amy Jung, Gaone Moetse, Bogadi Mothlobogwa, 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 flexible funders and partners who enabled a rapid COVID-19 response, including the Mulago Foundation, the Marshall Foundation, 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 Sports 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, summers, earthquakes, other 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 2018; 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). 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 on-going instruction. Doing so at scale requires low-technology solutions that can reach as many families as possible.

In this paper, we evaluate two low-tech solutions that leverage text messages and direct phone calls to empower parents to educate their children.[2] A sample of 4,500 families with primary-school-aged children across over half the 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.

This paper provides some of the first experimental estimates on minimizing the fallout of the COVID-19 pandemic on learning.[3] We find large, statistically significant learning differences between treatment and control groups. For the combined phone and SMS group, there was a 52 percent decrease in the share of students who could not do any numerical operations on an ASER test.[4] The gains on average numerical skill are 24 percent, which translates to 0.29 standard deviations. For the SMS-only group, we see positive, statistically significant effects roughly half this size: a 0.16 standard deviation gain on the ASER test. The ASER test was adapted into a phone-based assessment and incorporated time limits and a requirement that children explain their work to accurately identify their numeracy levels. We run a series of robustness tests that account for varying response rates across treatment groups.

[2] These interventions relate to growing literature on mobile phone technology and education. Text messages in particular 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., 2018), and to help parents monitor their child's effort and progress in school (Bergman 2015; Kraft and Rogers, 2015; Cunha et al., 2017; Berlinski et al., 2017; Bergman and Chan, 2017; Siebert et al., 2018; Musaddiq et al., 2019; de Walque and Valente, 2018; Gallego et al., 2020). See Bergman 2019 for a review.

[3] Another ongoing trial we are aware of includes low-cost tutoring in Italy (Carlana and La Ferrara 2020).

[4] 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).

Demand for low-tech interventions was high: 98 percent of households wanted to continue 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 the intensive margin (demand by type of low-tech service). The most demanded service is a combination of phone calls and SMS messages (70 percent) followed by SMS-only (18 percent). Receiving both phone calls and SMS messages increased demand for this combined service; the SMS-only group increased demand for SMS messages only. 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.

Roughly two thirds of families engaged with their child on educational activities per week; the interventions increased this proportion by 10 percent to 18 percent. We find that, instead of crowding out other educational activities, the interventions crowd in more educational interaction as some parents spend more time with their children on educational content.

We also find that parents tended to overestimate their child's performance by 0.25 standard deviations in the absence of any intervention. The phone and SMS interventions close this gap, but around 60 percent of parents still cannot correctly identify their child's learning level after the intervention. This result suggests that while more parental time with their children engaged in education activities increased learning and correct identification of their child's learning level and needs, there was room to help parents further target instruction precisely to their child's level.

Using high-frequency data at week four we implemented a targeted instruction intervention which sent SMS and phone instructions based on student numeracy levels. This targeted instruction approach relates to principles from "Teaching at the Right Level" (TaRL), a classroom-based intervention evaluated over 20 years, revealing that targeting instruction by learning level, rather than by age or grade, produces cost-effective gains in learning when delivered by teachers or volunteers (Banerjee et al. 2007; Banerjee et al. 2010; Dupas, Duflo, 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). This second wave of the trial cross-randomized the two main interventions with content targeted to each student's learning level. Data collection on results from the second wave is ongoing during the month of July.

This trial was designed to be rapid and iterative with multiple waves of data collected at four- to six-week intervals. This approach enabled program adaptation in response to incoming data and generated necessary knowledge on the most effective approach in an environment with significant uncertainty. We present a first wave of results after four weeks, which will be followed by a ten-week follow-up and a longer-term follow-up once schools reopen.

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). 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, 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, and weather shocks. 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.[5] 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 (Center for Global Development 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.

2 Background

2.1 Global Education and COVID-19 Landscape

Over 190 countries closed schools country-wide at the height of the COVID pandemic (UNESCO 2020). As of July 20th 2020, over 1 billion children continue to be affected by school closures. 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, there was a global learning crisis 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. To address the learning crisis, which has been exacerbated by the COVID-19 pandemic, approaches that can cost-effectively improve learning on a global scale are needed.

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. To date, only one COVID-19 death has occurred in Botswana. With over 50,000 tests conducted, this statistic is unlikely due to limited testing. While the direct effects of the pandemic have been minimal, the fallout of the pandemic on education 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. 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] The role of technology as a complement or substitute for the traditional schooling system is reviewed in Bettinger et al. (2020).

2.3 Botswana Education Context and Programming Prior to COVID-19

A coalition of partners led by the Botswana Ministry of Basic Education is scaling-up an education program called "Teaching at the Right Level (TaRL)" to all primary schools in Botswana. The program involves regrouping students by their actual learning level (e.g. addition, subtraction) rather than their expected learning level, which is often determined by over-ambitious grade-level curricula. This program has been tested across six prior randomized trials over two decades across Kenya, India and Ghana and shows consistent and positive effects ranging from 0.08 to 0.71 standard deviations. A 2017 study shows that the number of students who could read a paragraph or story more than doubled after students received a Teaching at the Right Level intervention (Banerjee et al. 2017). These effects are substantial in the education literature in which 0.20 standard deviation effects size are considered large (Kraft 2019).

The TaRL approach is particularly relevant where students are low and far behind grade-level expectations. This is the case in Botswana. While Botswana has high primary school enrollment at around 90 percent (UNESCO 2014), learning levels are low and stagnating. Among its regional neighbors in East and Southern Africa, Botswana ranks poorly on basic math and literacy proficiency, behind Tanzania and Kenya (SACMEQ 2013). A survey of basic literacy and numeracy in two regions found strikingly low levels of learning: 32 percent of standard 5 students could not do subtraction; 88 percent could not do division; 44 percent could not read a story in English; and a fifth of students could not read a paragraph (Pansiri et al. 2017). Students are falling multiple grade levels behind and are failing to acquire basic skills.

At the time of COVID-19 school closures, the TaRL program in Botswana had reached over 15 percent of primary schools in the country. In 2019, descriptive data show that the intervention resulted in the percentage of students who were innumerate falling from 30 percent to 4 percent and the percentage of students who could do division increasing from 7 percent to 55 percent in one term of implementation (Young 1ove 2019). The coalition scaling up TaRL in Botswana is led by the Ministry of Basic Education and supported by the Ministry of Youth Sports and Culture Development, Young 1ove, UNICEF, USAID as well as TaRL Africa. This scale-up is aligned to the Ministry of Basic Education's Education Training Sector Strategic Plan (ETSSP) for 2015-2020.

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 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 3-5. Given high interest for remote support, numbers were collected both for students participating in prior TaRL programming (13 percent of all numbers) as well as those who were not (87 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 15-20-minute phone call 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 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 below 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).

Figure 1: Intervention SMS Text Message Example

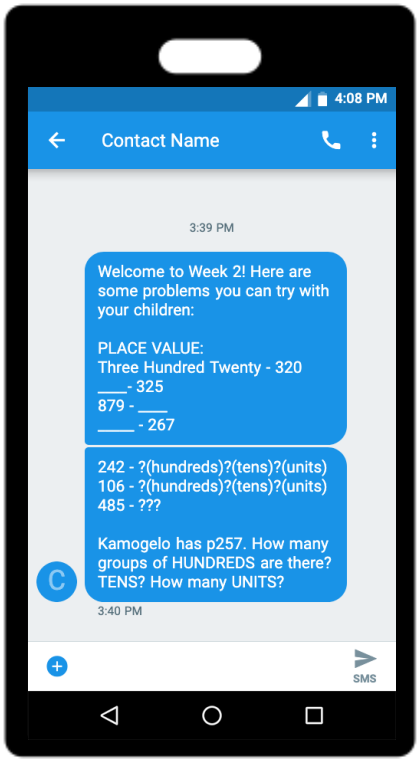


Figure 2: Sample phone call introduction



Sixty-four facilitators were assigned a group of about twenty-four households to case manage. Each facilitator called for around six 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).

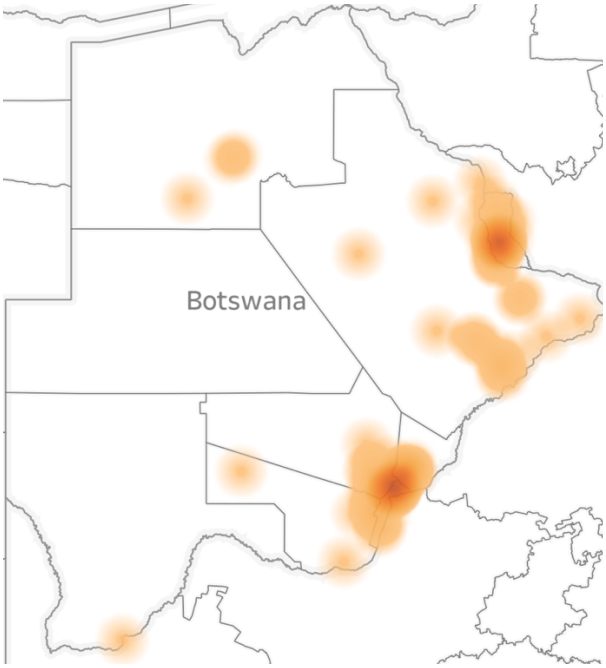
Pilots of both interventions were conducted over the course of two weeks prior to launching the interventions to ensure acceptability and feasibility. For example, more elaborate learning activities were initially planned, such as tossing stones into three concentric circles with various place values. 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.

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 collected 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 to three groups of equal size: a weekly SMS message followed by a phone call, a weekly SMS message only, and a pure control group. Randomization was stratified on whether at least one child in the household had previously participated in TaRL. Each phone number represents a caregiver or a household. Roughly 80 percent of the households had one student while the remaining 20 percent had multiple students. We conduct a four-week evaluation on a random subset of all households covering roughly 2,250 students.

Figure 3: Distribution of Schools of Student Participants across Botswana



Notes: this density map shows the relative distribution of schools linked to students in our sample. Darker regions correspond to higher concentrations of schools for study participants.

After week four, this random subset was cross-randomized with an additional intervention: targeted instruction. We used data on the learning levels from the week-four assessment to send tailored text messages to each student in the fifth week. For example, students who know addition received subtraction problems to push them to a higher level of learning, whereas students who knew multiplication were sent division problems. This targeted instruction program is adaptive, building on data collected at week four, which (a) revealed parents were having a hard time learning the level of their child, thus requiring additional support to target instruction effectively and (b) enabled us to have near real-time data to target instruction. At each week interval thereafter, we collect weekly data on a "problem of the week" using a phone survey and use this question to further target weekly SMS messages. If the child responds correctly, problems the following week progress to a higher operation level. If not, the child is given the same level problem. At week ten, we will collect survey data and conduct assessments to evaluate these additional components of the intervention.

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.

Figure 4: Intervention and Evaluation Timeline

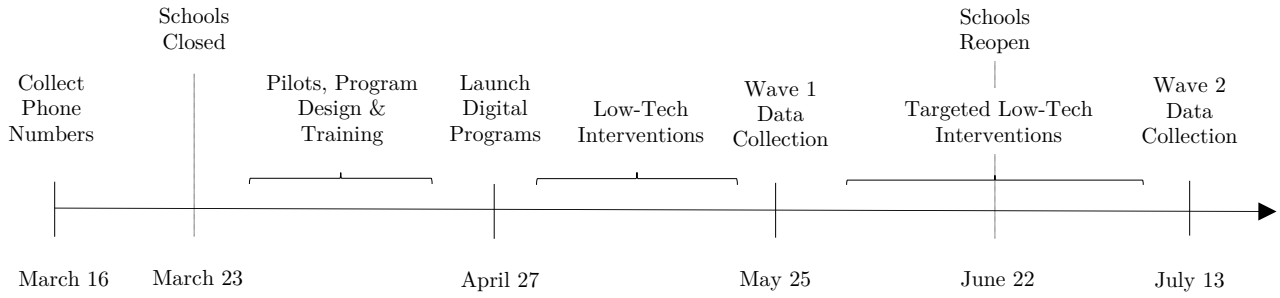
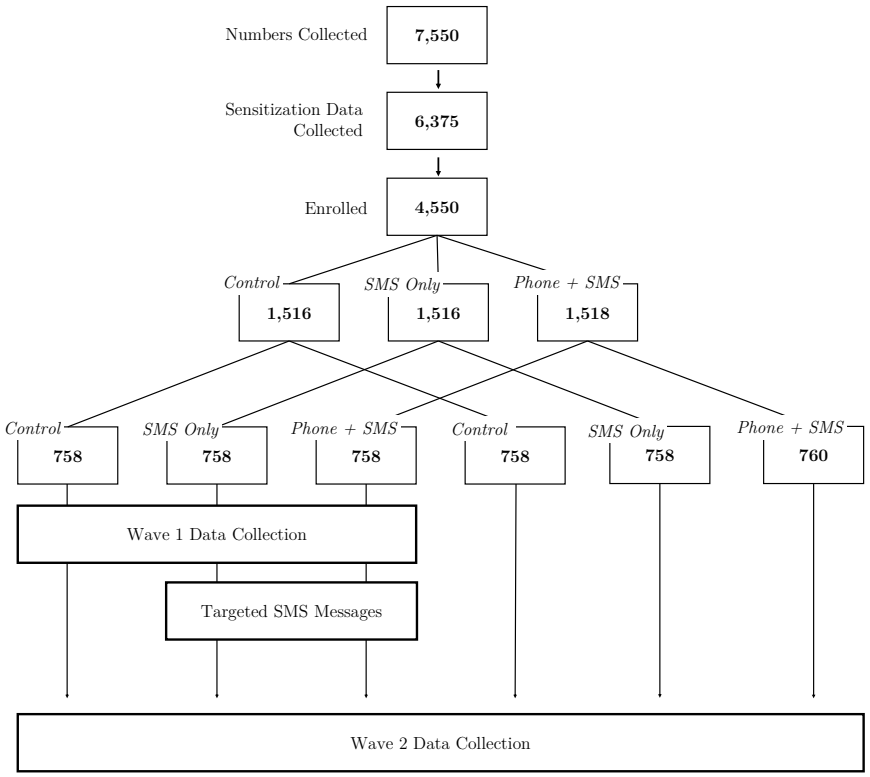


Figure 5: Experimental Design



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

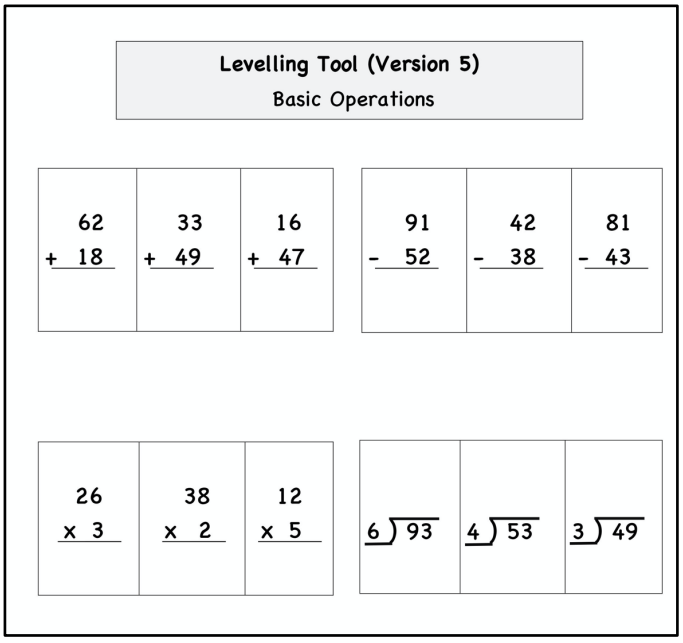
4.1 Data Collection

We conducted a first wave of evaluation over two weeks with a random subset of around 2,250 out of all 4,550 phone numbers in the trial. 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. Only three households of the 2,250 in our sample were evaluated by the same facilitator that provided their weekly intervention calls providing for objective assessment.

The survey consists of 11 questions related to engagement in educational activities, parental perceptions of their child's learning, and student learning outcomes. While the survey was conducted with the parent, student learning outcomes were collected by directly assessing the child or children 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 below 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 the following 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.

The ASER test is measured on a 0-4 scale. We use the raw score as an outcome as well as convert it to standard deviations relative to the control group. We also show the results for passing each section of the assessment: addition, subtraction, multiplication, and division. In addition to the ASER test, we evaluate the children's ability to answer a simple place value word problem to capture learning outcomes more basic than mathematical operations. We present these results as a binary outcome for whether the child successfully answered the world problem.

Figure 6: Sample of ASER test used in Botswana



We have no direct measure of baseline learning; however, we did collect data on whether or not the child had participated in TaRL before, a proxy for a child's learning level. In Term 1 from February to March 2020 prior to the pandemic closing schools, 66 percent of students in TaRL had learned at least one new operation and innumeracy rates dropped from 32 percent to 11 percent. In the results that follow, we show that prior participation in TaRL is a good proxy for whether a child was at a higher learning level at the start of the low-tech interventions.

We measure learning engagement in two ways: first, we asked parents if they spent any time with their child on any educational activities in the past week. This outcome is an indicator for a parent recalling that they spent a non-zero amount of time on educational activity and zero otherwise. Among the treatment groups, we also ask parents if they recall their child attempting any of the problems sent via SMS in the last week.

We include four measures of a parent's perception of their child's numeracy level. First, we create two indicator variables: one for whether the parent specified their child's level and overestimated it, and another for whether the parent specified their child's level and underestimated it. We code a response of "I don't know," as zero in both instances as the parent neither over nor underestimated their child's level. Second, we create an indicator for whether the parent specified their child's level correctly; we again code "I don't know" as zero. Lastly, we show an indicator for whether a parent reported not knowing their child's numeracy level.

4.2 Sample Characteristics and Representativeness

We include a few descriptive statistics to describe how our sample for the low-tech intervention compares to national-level characteristics in Botswana. Table 1 includes these comparisons.

First, we compare the low-tech sample of households to the sample of students who had participated in Teaching at the Right Level in Botswana in the prior two years. We find that 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. This is further reflected in more regions represented in the low-tech sample, from 4 in prior TaRL implementations to 9 for the low-tech sample. In addition, we compare learning levels, 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, with students who cannot do any operations (labelled as "Beginner") 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, while not a formal validation test, 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.

Second, we compare our low-tech sample to national-level indicators from the Ministry of Education using data on enrollments and gender composition from 2017. Botswana has ten regions total, thus the low-tech sample which covers 9 regions has some representation in nearly all regions in the

nation. The low-tech sample includes over 103 schools which represents aronud 15 percent of schools in the country. We further 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 future waves of data collection, we will provide additional descriptive characteristics of our sample, such as the parent's level of education.

Table 1: Sample Characteristics Comparison

	Low-Tech Sample: Control Group	2018-2020 TaRL Sample: Baseline
Learning Assessment	Phone ASER	Face-to-Face ASER
Beginner	0.29	0.31
Addition	0.21	0.25
Subtraction	0.12	0.17
Multiplication	0.25	0.20
Division	0.13	0.07
Regions	9	4
Schools	103	92
	Study Schools in Sample	National School Data
Demographics		
Female	0.50	0.49
Enrollment Std 3-5	274	362
Male Enrollment Std 3-5	136	185
Female Enrollment Std 3-5	137	177
Primary School Leaving Exam Scores		
Percent A	0.16	0.14
Percent B	0.21	0.17
Percent C	0.41	0.36
Regions	9	10
Schools	103	755

Notes: For the low-tech sample and phone-based assessment, we include descriptive statistics for the control group only. For the TaRL sample, we include learning levels for baseline levels only. School-level data is from the Ministry of Basic Education and Botswana Examinations Council from 2017.

4.3 Attrition

Appendix Table 4 reports survey response rates. At the week-four interval we successfully followed up with 52 percent to 49 percent of students in the phone & SMS group and SMS-only group, respectively, with no statistically significant difference between the two groups. The control group was underrepresented by a statistically significant 7 to 9 percentage points. This is likely due to two factors: in the control group there was limited contact in between survey rounds making follow-up mechanically more difficult with less up-to-date contact and scheduling information; in addition, there is potential for response bias with control parents who received no intervention less likely to pick up the phone, potentially due to less trust or interest. We conducted an intensive follow-up at week five with a focus on the control group. This follow-up had a high response rate and yielded a total response rate in the control group of 64 percent, higher than the treatment groups. We address this response bias in several ways. First, we note that if we compare the SMS-only and SMS & Phone groups to each other, excluding the control group, the response rates are not different between these two groups. Second, we construct bounds on the treatment effects using a series of cut-offs to ensure robustness of our results. While our main results include only data collected in week 4, Appendix Table 3 presents our main outcomes when using different treatment bounds. First, we use a subset of the highest performers from the additional week-5 control group sample to impute scores from non-respondents in week 4. Second, we use a random subset of the additional week-5 control group sample to impute scores from non-respondents in week 4 students. Finally, we use all data from both week 4 and week 5 data collection and present the results. Appendix Table 4 presents a summary of response rates depending on which treatment bounds are used.

5 Empirical Strategy

Given randomization of the treatment and control groups, we use the following specification to estimate the causal effect of each intervention:

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

where Y_{ij} is an outcome for child i in 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 the TaRL education program immediately prior to the intervention. Standard errors are clustered at the 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.

We also estimate a specification in which we interact each treatment with the indicator for a student's participation in TaRL before the intervention:

$$Y_{ij} = \alpha_0 + \beta_1 SMS_j + \beta_2 SMSPhone_j + \beta_3 SMS_j * TaRL_{ij} + \beta_4 SMSPhone_j * TaRL_{ij} + \beta_5 TaRL_{ij} + \varepsilon_{ij}$$

This specification estimates heterogeneous effects by a proxy for baseline learning level. While we do not have direct measures of baseline learning, students who have participated in TaRL previously are likely to be at higher learning levels at the time of the low-tech intervention. We also run a similar specification to explore heterogeneous effects by gender:

$$Y_{ij} = \alpha_0 + \beta_1 SMS_j + \beta_2 SMSPhone_j + \beta_3 SMS_j * Female_{ij} + \beta_4 SMSPhone_j * Female_{ij} + \beta_5 Female_{ij} + \delta_s + \varepsilon_{ij}$$

6 Results

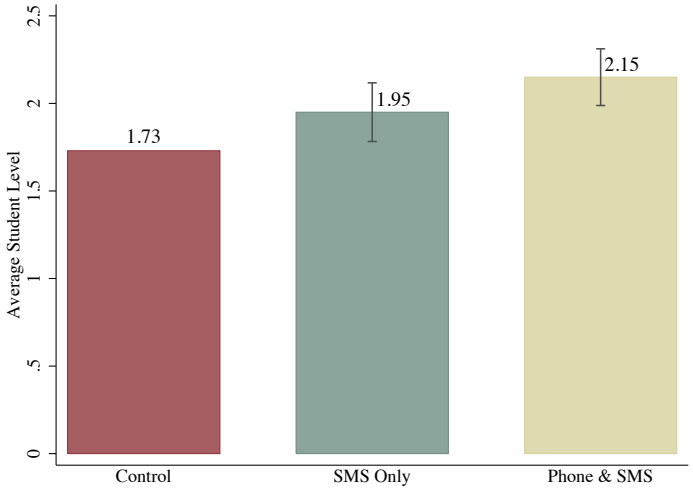
The results presented here are all intention-to-treat effects on whether households were randomly assigned to treatment. Monitoring data suggests that in the phone calling group active weekly participation by parents was on average between 79 to 86 percent and live interaction with both the parent and child on the call varied between 33 percent and 58 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 report intention-to-treat estimates in this paper since these are most likely to be the policy-relevant quantity assuming partial active participation at scale. However, if future efforts increase active participation, the effects could be larger.

6.1 Learning Effects of the SMS and the Phone Interventions

We find large, statistically significant impacts on learning in our first wave of results. Figure 7 presents results on average learning. For the combined phone plus SMS group, there is a 24 percent level gain on the ASER test score. In standard deviations, this equates to a 0.29 standard deviation increase. For the SMS-only group, we see positive, statistically significant effects about half this size: a 13 percent improvement on the ASER test score, which translates into 0.16 standard deviation gains.[6] The control mean learning level is 1.73, which reflects that the average student is, roughly speaking, not able to successfully do subtraction, while the phone and SMS interventions close this gap to around a level 2 and above which equates to subtraction-level and above. The lowest level a student can be in is "beginner," which indicates the student cannot successfully complete any operation, which we refer to as innumerate. Figure 8 highlights the impact of the interventions on those at beginner level. 29 percent of students are at the beginner level in the control group, which is reduced by over half (52 percent) in the phone plus SMS group and by over a third (34 percent) in the SMS-only group.

[6] The control standard deviation is 1.44 and the phone and SMS treatment effect is 0.421, translating into a 0.29 standard deviation gain. For the SMS-only treatment effect of 0.225, this translates into a 0.16 standard deviation gain.

Figure 7: Average Numerical Skill Level



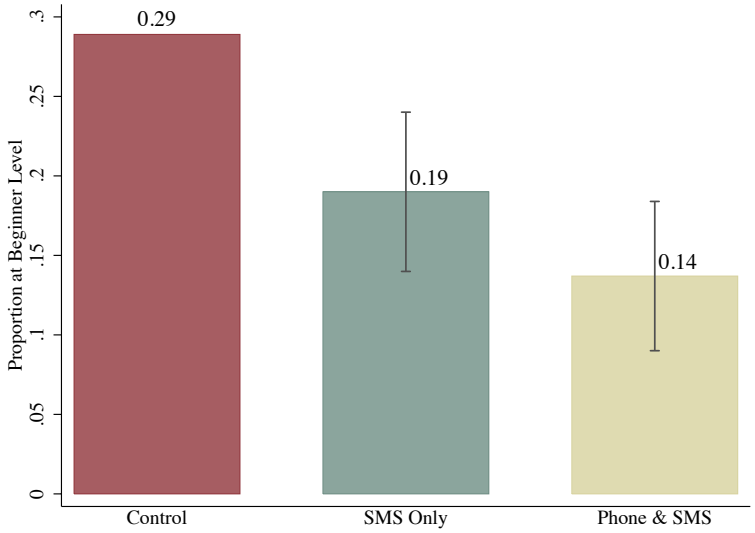
Notes: 90 percent confidence interval bars on SMS Only and Phone & SMS show treatment effects relative to the control group.

Appendix Table 1 shows that the difference between the phone plus SMS group and the SMS-only group is statistically significant at the 5 percent level. Note that there is no statistically significant differential attrition between these two groups, which provides further evidence that the phone plus SMS intervention leads to learning gains. Students were also assessed on place value competency. The SMS-only intervention had no impact on this skill, but the phone plus SMS intervention increased scores from 74 percent in the control group to 86 percent in the treatment group (see Appendix Table 1). Appendix Figure 1 shows the results for each level: beginner, addition, subtraction, multiplication and division. Of note, most numeracy gains are on lower levels, such as fewer innumerate beginner students and more subtraction students, rather than on higher levels such as multiplication and division. This result is expected since the content of the weekly SMS problems in weeks one through three included only place value and basic operations. The results by level reveal that the intervention worked as designed and suggests a likely mechanism behind learning gains is specific to the content of the intervention rather than solely motivation or accountability.

Both low-tech interventions yield large, statistically significant learning gains. The phone & SMS messages intervention impacts are nearly twice as large and push students' numeracy skills farther from the beginner level, but the SMS intervention is lower cost to implement. To put these interventions in context, Kraft (2019) provides benchmarks based on a review of 1,942 effect sizes from 747 RCTs evaluating education interventions with standardized test outcomes. In this review, 0.05 is considered small, 0.10 is the median effect size, and anything above 0.20 is considered large.

Next we explore heterogeneous effects by degree of prior numeracy learning support. Appendix Table 2 and Figure 9 show effects by students' prior participation in TaRL, a proxy for extra learning support and in turn baseline learning level at the start of the low-tech interventions. In our data in the control group, 13 percent of students who were in TaRL are innumerate compared to 34 percent of students who were not in TaRL. Students who were in TaRL also have a 0.40 standard deviation higher assessment scores. While we do not have a test-score measure of baseline learning, prior TaRL participation appears to be a reasonable proxy for higher baseline learning levels at the start of the low-tech interventions.

Figure 8: Proportion of Innumerate Students

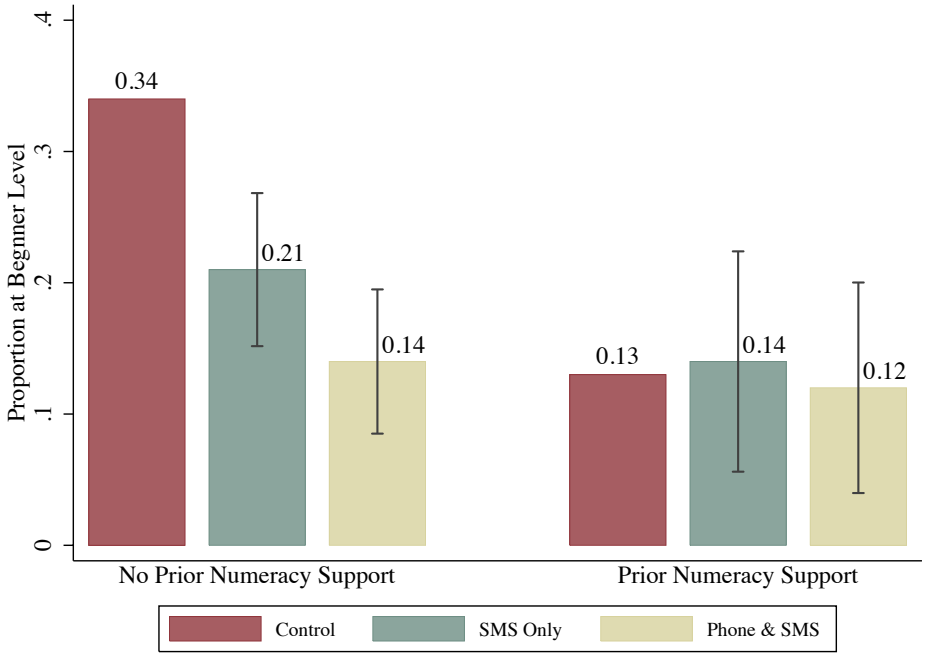


Notes: 90 percent confidence interval bars are shown on SMS Only and Phone & SMS and show treatment effects relative to the control group.

We see that the effects of both interventions are twice as large for students who did not receive prior numeracy support as for students who received prior support: the phone plus SMS intervention increases students' levels by half a point compared to a quarter of a point, respectively. Similarly, the SMS-only intervention increases learning by 0.29 points for students with no prior support relative to 0.15 points for students with prior support. These results suggest that the low-tech learning interventions have the potential to close gaps in learning between higher and lower-performing students; in particular, the phone plus SMS intervention has large benefits for lower-performing students. These results also reveal that these low-tech interventions improve learning outcomes even (and mostly) for students who did not have prior relationships with the facilitators who call them, suggesting the low-tech interventions do not require pre-existing relationships to work, and are mediated more by pre-existing learning gaps.

We also estimate heterogeneous treatment effects by gender in Appendix Table 2. We find that girls generally perform better by over 0.30 average levels, but do not benefit more from the low-tech interventions. Rather it appears boys, who also start with lower numeracy levels, benefit more from low-tech support. This reinforces the notion that the effectiveness of the interventions is largely driven by closing prior learning gaps on the most basic numeracy levels. Of note, the content of the intervention was basic during the initial few weeks (place value and basic operations), targeting those starting at lower learning levels. The next evaluation wave will reveal effects effects when all learning levels received targeted content starting from week 5.

Figure 9: Proportion of Innumerate Students by Prior Numeracy Support



Notes: 90 percent confidence interval bars on SMS Only and Phone & SMS show treatment effects relative to the control group.

Our learning results are short-term. In future waves of data collection, we plan to test for persistence and account for measurement error via phone. We expand on lessons learned on measuring learning via phone in Angrist et al. (2020). Of note, the learning gains observed might be driven by either learning gains, minimizing learning loss, or a combination of both. We are not able to disentangle between the two in this first wave of data collection. In follow-on data collection we will map learning trajectories in the control group to give an indication of the degree of learning loss in the status quo. We will also aim to detect the degree to which learning gains are a function of cognitive skill or effort on the test, building on prior evidence suggesting that effort alone can result in higher learning outcomes on tests (Gneezy et al. 2019).

6.2 Robustness of Learning Results

We run a series of robustness tests for our main learning outcomes. Below we summarize these analyses. We collected data at the week four interval which was balanced between SMS and phone treatments at 50.7 to 52.6 follow-up, but the control group was underrepresented by 7 percentage points. This enables us to test for differences between the two treatment groups, which had similar response rates. For comparisons with the control group, we run a series of robustness tests and bounding exercises.

We conducted a follow-up survey attempt with the control group in week 5 to achieve balance between treatment and control groups. In fact, we reached even more students in the control group with 12 percentage points more control students represented compared to the treatment groups. The week four and week five data enable us to derive a series of plausible bounds on either end of a

balanced sample across treatment groups. First, as we have already shown above, we presented comparisons of the SMS intervention to the phone plus SMS intervention, for whom response rates were similar. These results show that the latter resulted in increased learning. Second, we include a random subsample of week 5 data to balance the control group response rate. Third, we conduct a "worse-case" scenario similar in spirit to Horowitz and Manski (2000). We equalize the response rates across groups drawing only from the top of the test-score distribution among control participants evaluated in week 5, such that any observed differences between treatment and control reflect a conservative lower bound of the likely impact. Fourth, we show results including all of the week 5 data. Appendix Table 4 shows details of response across treatment groups and survey weeks.

Appendix Table 3 shows robustness tests and reveals smaller average treatment effects, but similar overall patterns. The phone and SMS intervention is the most effective, and the SMS-only intervention less effective, but both produce statistically significant learning gains on basic numeracy such as fewer beginner or innumerate students. For instance, both the SMS and the phone plus SMS intervention reduce the likelihood a student is at the beginner level, even under our worst-case bounds described above. The phone and SMS treatment effect is a 46 percent reduction in innumeracy (relative to 52 percent for the unbounded results) and the SMS treatment effect is a 21 percent reduction in innumeracy (relative to 34 percent for unbounded results).

In future waves of data collection, we will formalize this analysis further, including additional validation tests as well as comparing waves of data collection.

6.3 Learning Engagement

It is possible that prompting families to engage with their child about numeracy might crowd out parental engagement, who turn their attention elsewhere believing their child is now in capable hands, rather than increasing engagement. Both interventions, however, increase overall learning engagement. Table 2 shows two measures of engagement described below.

First, we asked families whether they recall spending time engaging with their child on educational activities. We discretize the variable to minimize the influence of outliers. Column 1 of Table 2 shows that roughly two thirds of families did so; the phone plus SMS and SMS-only intervention increased this engagement by 7 percentage points and 12 percentage points respectively. The 5 percentage point difference between the two treatments is not statistically significant below the 10 percent level. The overall change in engagement still holds if we use finer discretization of the data.

Second, within the treatment groups only, we asked families whether their child attempted the SMS message problems of the week. Engagement overall is high: 84 percent of children attempted the problem. Engagement is higher in the phone plus SMS group, which increased engagement with the problems by 10 percentage points. This is likely because facilitators walked families through the problems during the live calls.

Overall, it appears that there is high engagement with both the SMS-only and the phone plus SMS message interventions. If anything, the interventions seem to increase parents' overall educational activities with families, potentially through a motivation or accountability effect.

Table 2: Engagement

	(1) At least some time spent on educational activities	(2) Student attempted problems from intervention SMS
Phone & SMS	0.07**	0.10***
	(0.03)	(0.02)
SMS Only	0.12***	
	(0.03)	
Observations	1,671	1,164
Comparison Group	Control	SMS Only
Comparison Group Mean	0.67	0.84
Phone & SMS v. SMS Only P-Value	0.13	0.00

Notes: 1. *** p<0.01, ** p<0.05, * p<0.1. 2. Robust standard errors clustered at household level in parentheses. 3. Column (2) was only collected from treatment arms, thus the comparison mean is the mean for the SMS Only arm and no p-value is presented.

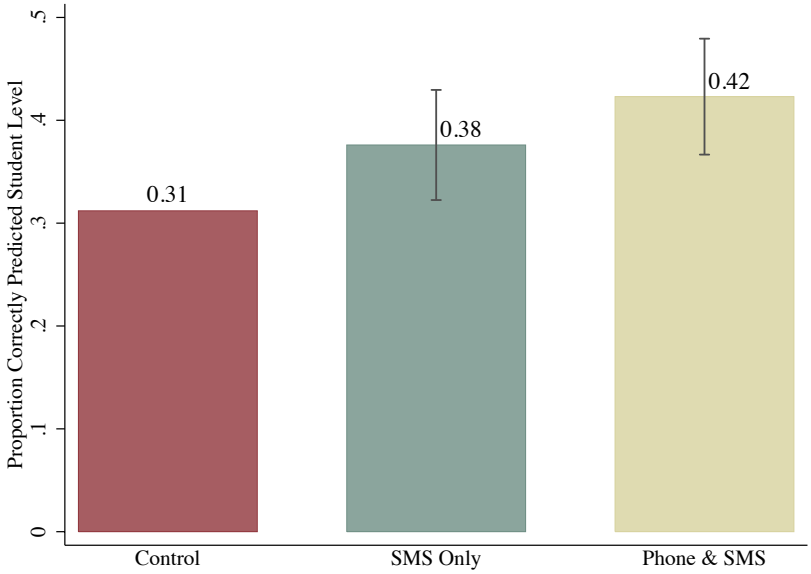
6.4 Parent Perceptions of Their Child's Ability and Learning

Previous research has shown that parents can misperceive their child's effort and learning, which can impede parents' engagement with their child's learning (Banerjee et al. 2010; Bergman 2019; Dizon-Ross 2019). Providing information to parents about their child's learning can ameliorate these misperceptions. We find results consistent with this mechanism.

Figure 10 shows the share of parents, by treatment group, who correctly estimate their child's learning level. Only one third of parents in the control group correctly estimate their child's level. Appendix Table 5 shows the interventions primarily shift parents' perceptions from "I don't know"-- reducing this response by more than 10 percentage points-- to correctly estimating their child's level. The phone plus SMS intervention increases the latter by 11 percentage points (statistically significant at the 5 percent level) and the SMS-only intervention increases correct perceptions by 6 to 7 percentage points (statistically significant at the 10 percent level).

Thus, the interventions appear to promote engagement with the numeracy content provided by SMS while not crowding out engagement in educational activities overall. Moreover, both interventions shift parents' perceptions of their child's learning levels which might allow families to target learning activities to their child's level more effectively. At the same time that parent perceptions shift as a result of the intervention, still around 60 percent of parents cannot correctly identify their child's level. This motivated a follow-on intervention at the four-week interval, which provided more structure to targeting instruction to the child's level via SMS messages. Using data from wave 1, we sent tailored messages to students, for example, sending multiplication problems to students who know subtraction but not multiplication. We cross-randomize this intervention with phone calls and SMS with results to follow in the near future.

Figure 10: Proportion of Students whose Caretakers Correctly Predicted their Level



Notes: 90 percent confidence interval bars are shown on SMS Only and Phone & SMS and show treatment effects relative to the control group.

6.5 Demand

Demand for any combination of low-tech interventions was high: 98 percent of households wanted to continue the program after four weeks. The most sought after intervention combination was a combination of phone and SMS (70 percent), followed by SMS-only (18 percent), either phone or SMS (9 percent), phone only (3 percent) and none (1 percent). Appendix Figure 2 shows the distribution by treatment group. While the interventions do not affect demand on the extensive margin (desire for any service), likely since it was already nearly 100 percent, the intervention affects the intensive margin (type of demand). Parents in the phone or SMS group increase demand only for the intervention they received, with an increase in demand for phone and SMS services from 70 percent to 86 percent. Parents in the SMS-only group increase demand only for the intervention they received, with an increase in demand in SMS messages only from 18 percent to 25 percent. This finding suggests that receiving an intervention, even if not the preferred intervention at the outset (as is the case for SMS only), can increase subsequent demand. One reason this might be true is that families in the treated groups observe the benefits of their intervention, such as improved learning, and thus demand more of the intervention.

7 Cost-effectiveness

Both low-tech interventions are relatively low cost.[7] We estimate an upper bound on costs which includes total programmatic costs, personnel time, as well as fixed costs to collect phone numbers, set up new infrastructure, conduct training, and collect routine monitoring data.[8] 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 by the four-week juncture was about $3,200 USD. For phone calls, the marginal cost above the bulk text message was $17,800. This equates to $2.13 per child reached in the SMS group and $14 dollars per child reached in the phone and SMS group. Given average treatment effects of 0.16 and 0.29 standard deviations, this translates to $13.3 USD per standard deviation gain in learning for the SMS-only group and $48.28 USD per standard deviation gain in learning for the SMS and phone group.

These estimates are cost-effective relative to the literature. We make comparisons using a $50 benchmark which yields a 1 standard deviation gain for our phone and SMS treatment. As a comparison, conditional cash transfers in Malawi yielded an extra 0.01 standard deviation per $50; an extra contract teacher and streaming by ability yielded 0.47 standard deviation gain per $50 in Kenya; and remedial tutoring in India yielded an effect of 0.65 standard deviation gain per $50.[9] These comparisons suggest both low-tech interventions tested are cost-effective relative to other popular and effective interventions in the education literature.

8 Policy Implications

Our results have both immediate and long-term policy relevance. As of July 20th 2020, over 1 billion children continue to be affected by school disruption (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 short-term learning during school closures, with reductions in innumeracy of 34 percent and 52 percent reductions via SMS-only and Phone & SMS, respectively, and a full standard deviation gain in learning for between $13 to $50 per child.

[7] 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.

[8] 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 closed there are minimal opportunity costs in terms of time. In the next wave, as workplaces reopen, we will factor in these costs.

[9] We rescale these estimates which were originally expressed in terms of standard deviation gains per $100 from the Jameel Poverty Action Lab (J-PAL) website to provide a direct comparison to the cost of the phone and SMS intervention.

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 (Center for Global Development 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. If phone calls are included, on what dimension should calls be targeted? Given their added cost and complexity calls are unlikely to be able to reach everyone. Our results suggest that targeting low-performing students yields the highest returns, with both texts and phone calls closing a substantial gap between low and high-performing students. Direct phone calls at scale might consist of weekly teacher phone calls to the bottom 5 to 10 percent of their class.

Of note, the implementation and deployment 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 interaction points in low- or middle-income country contexts - which often focus on providing information on the child's performance - might be substantially enhanced with simple, easy-to-engage 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, and during weather-related shocks. 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 two low-tech SMS text message and phone call interventions have large and cost-effective effects on household engagement in education and learning during school closures. We find up to 52 percent reductions in innumeracy from phone calls and SMS texts, and 34 percent reductions in innumeracy using SMS texts only. These gains translate into 0.16 to 0.29 standard deviation gains. In terms of cost-effectiveness, we estimate a full standard deviation of learning can be gained for between $13 to $50 per child. We also find higher parental engagement in educational activities with their children and more accurate parent perceptions of their child's learning level. These results have substantial implications for global policy during current global school closures, as well as lessons for future school closures beyond the crisis.

Our results reveal promise for low-tech interventions that are relatively cheap and easy to scale. Of note, while our results are promising, replication and 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 as well as Interactive Voice Response (IVR). Moreover, follow-on research can help disentangle the mechanism behind learning gains, including how much of the gains are driven by access to more learning content, effort, motivation, or accountability. Follow-on trials could be implemented in a rapid and adaptive approach, generating real-time data to optimize and improve interventions as they scale.

In this paper, we present a first wave of results after four weeks, which will be followed by a ten-week follow-up and a second intervention which uses data from the first wave of the trial to target SMS and phone call instruction to a child's learning level. We are also conducting a longer-term follow-up to determine persistence of learning gains, validate learning measures, and collect a series of non-cognitive outcomes such as well-being and parental self-efficacy. Finally, we will explore the implications of these low-tech interventions as substitutes or complements of the traditional schooling system as schools open.

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. 2020. "Practical lessons for phone-based assessments of learning." *BMJ Global Health* 5, no. 7: e003030.

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." NBER Working Paper Series.

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.

Center for Global Development. 2020. "Schools Out, Now What?" CGD Blog.

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.

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

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.

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.

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

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.

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

Horowitz, Joel L., and Charles F. Manski. 2000. "Nonparametric analysis of randomized experiments with missing covariate and outcome data." *Journal of the American statistical Association* 95, no. 449: 77-84.

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. 2019. “Interpreting Effect Sizes of Education Interventions. EdWorkingPaper No. 19-10.” Annenberg Institute for School Reform at Brown University.

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. "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review* 109, no. 4 (2019): 1426-60.

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

Pansiri, O. Nkobi and Gabatshwane T. Tsayang. 2017. “A Situational Analysis of Basic Literacy and Numeracy levels at early grade levels in Botswana.” Background Report.

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 No. 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.

Odhiambo, Moses. 2019. “State to spend Sh8.4bn more on school tablets.” The Star.

Trucano, Michael. 2013. “A new wave of educational efforts across Africa exploring the use of ICTs.” World Bank Blog.

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

Southern African Consortium for Monitoring Education Quality (SACMEQ). 2013. Retrieved from http://www.sacmeq.org/

UNESCO Institute for Statistics (UIS). 2014. Retrieved from http://uis.unesco.org/en/country/bw

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

Appendix

Appendix Table 1: Learning Outcomes

	(1) Student Level	(2) Beginner Level	(3) Addition Level	(4) Subtraction Level	(5) Multiplication Level	(6) Division Level	(7) Answered Place Value Question Correctly
Phone & SMS	0.42***	-0.15***	0.00	0.07***	0.04	0.04	0.08***
	(0.10)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
SMS Only	0.23**	-0.10***	0.04	0.03	-0.01	0.04	0.01
	(0.10)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Observations	1,330	1,330	1,330	1,330	1,330	1,330	1,339
Control Mean	1.73	0.29	0.21	0.12	0.25	0.13	0.74
Phone & SMS v. SMS Only P-Value	0.04	0.04	0.15	0.08	0.09	0.98	0.01

Notes: 1. *** p<0.01, ** p<0.05, * p<0.1. 2. Robust standard errors clustered at household level.

Appendix Table 2: Heterogeneous Treatment Effects

	(1)	(2)	(3)	(4)
	Student Level	Beginner Level	Student Level	Beginner Level
Prev in TaRL	0.57***	-0.21***		
	(0.17)	(0.04)		
Phone & SMS X Prev in TaRL	-0.25	0.19***		
	(0.23)	(0.06)		
SMS Only X Prev in TaRL	-0.14	0.14**		
	(0.25)	(0.06)		
Phone & SMS	0.50***	-0.20***	0.56***	-0.21***
	(0.11)	(0.03)	(0.14)	(0.04)
SMS Only	0.29**	-0.13***	0.31**	-0.16***
	(0.11)	(0.04)	(0.14)	(0.04)
Female			0.34**	-0.11**
			(0.14)	(0.05)
Phone & SMS X Female			-0.24	0.11**
			(0.19)	(0.06)
SMS Only X Female			-0.15	0.11*
			(0.19)	(0.06)
Observations	1,329	1,329	1,330	1,330
Conditional Control Mean	1.60	0.34	1.55	0.35
Condition	Not Prev in TaRL	Not Prev in TaRL	Not Female	Not Female

Notes: 1. *** p<0.01, ** p<0.05, * p<0.1. 2. Robust standard errors clustered at household level.

Appendix Table 3: Bounded Learning Outcomes

	All of Week 5 Data		Random Subset of Week 5 Data		Top Subset of Week 5 Data	
	(1)	(2)	(3)	(4)	(5)	(6)
	Student Level	Beginner Level	Student Level	Beginner Level	Student Level	Beginner Level
Phone & SMS	0.34***	-0.12***	0.36***	-0.14***	0.16*	-0.11***
	(0.09)	(0.02)	(0.09)	(0.03)	(0.10)	(0.03)
SMS Only	0.15	-0.07**	0.17*	-0.08***	-0.03	-0.05*
	(0.09)	(0.03)	(0.10)	(0.03)	(0.10)	(0.03)
Observations	1,541	1,541	1,418	1,418	1,418	1,418
Control Mean	1.81	0.26	1.78	0.27	1.99	0.24
Phone & SMS v. SMS Only P-Value	0.04	0.03	0.04	0.03	0.04	0.03

Notes: 1. *** p<0.01, ** p<0.05, * p<0.1. 2. Robust standard errors clustered at household level. 3. For columns (1) and (2), all 212 control students and all 15 SMS Only students evaluated in week 5 are included in the analysis. For columns (3) and (4), a random subset of 73 control students from week 5 and all 15 SMS Only students evaluated week 5 are included. For columns (5) and (6), the top 73 performing control students from week 5 and all 15 SMS Only students evaluated week 5 are included. The number 73 for columns 3-6 was chosen as it is the value that ensures balance between the three experiment arms.

Appendix Table 4: Attrition

	Week 4 Data Only	Week 4 Data Only	Week 4 + Week 5 Data	Week 4 + Subset of Week 5 Data
	(1)	(2)	(3)	(4)
	Survey Happened	Student Assessment Happened	Student Assessment Happened	Student Assessment Happened
SMS Only	0.11***	0.06**	-0.14***	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)
Phone & SMS	0.07***	0.09***	-0.12***	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Observations	2,757	2,757	2,757	2,757
Control Mean	0.64	0.43	0.64	0.52
Phone & SMS v. SMS Only P-Value	0.14	0.21	0.50	0.50

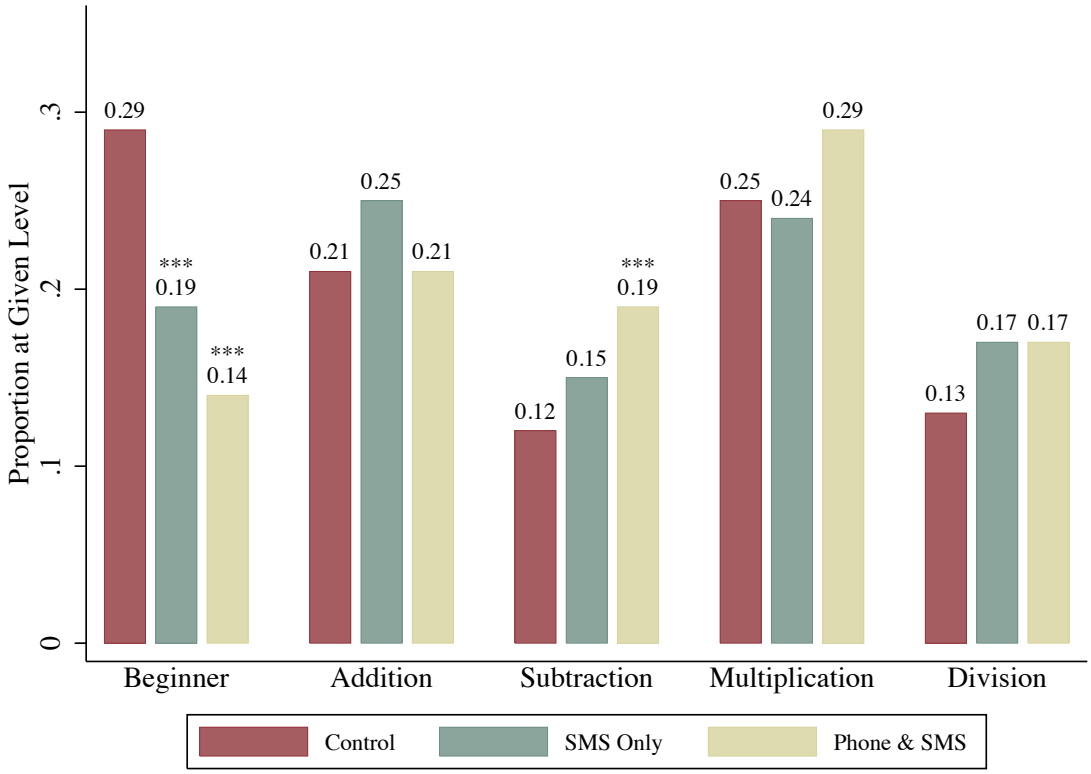
Notes: 1. *** p<0.01, ** p<0.05, * p<0.1. 2. Robust standard errors clustered at household level. 3. For columns (1) and (2), only students evaluated during week 4 are included. For columns (3), all students evaluated in week 4, all 212 control students evaluated in week 5 and all 15 SMS Only students evaluated in week 5 are included. For column (4), all students evaluated in week 4, all 15 SMS Only students evaluated in week 5, and a random subset of 73 control students evaluated in week 5 are included. The number 73 for column 4 was intentionally chosen to ensure balanced attrition.

Appendix Table 5: Learning Perceptions

	(1) Underestimated Student's Level	(2) Correctly Predicted Student's Level	(3) Overestimated Student's Level	(4) Reported not knowing Student's Level
Phone & SMS	0.02	0.11***	0.02	-0.14***
	(0.03)	(0.03)	(0.03)	(0.02)
SMS Only	0.07**	0.06**	-0.02	-0.11***
	(0.03)	(0.03)	(0.03)	(0.02)
Observations	1,393	1,393	1,393	1,651
Control Mean	0.17	0.31	0.32	0.17
Phone & SMS v. SMS Only P-Value	0.10	0.17	0.23	0.04

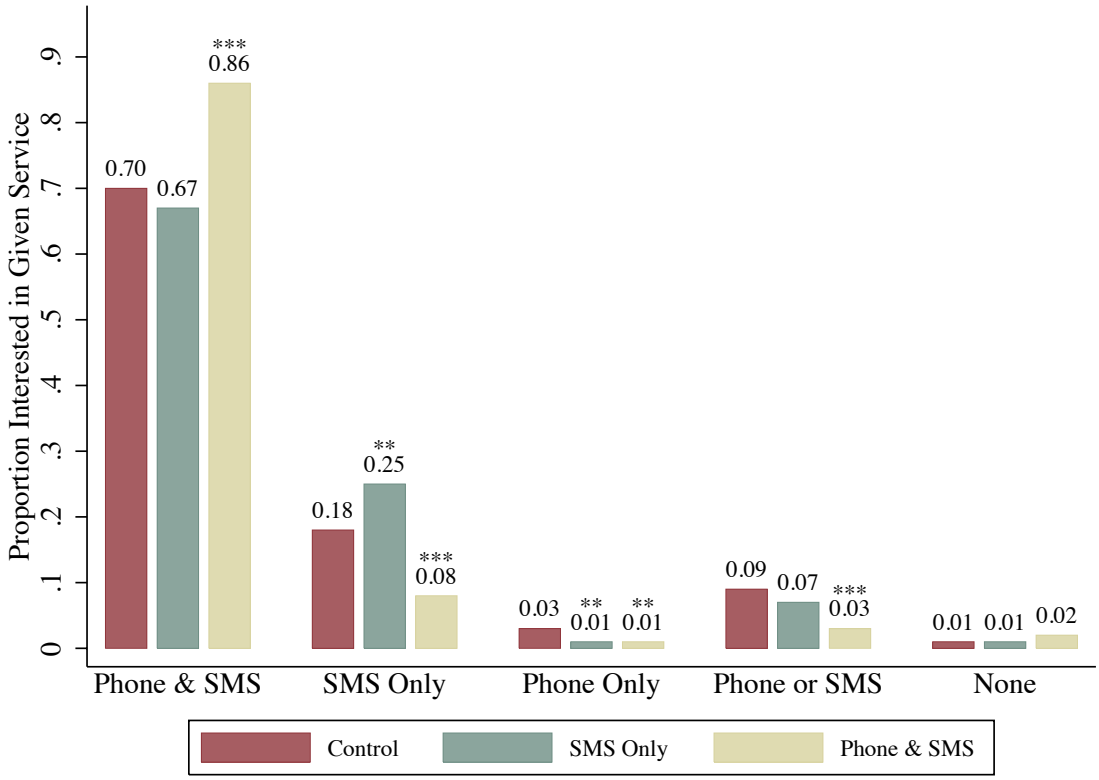
Notes: 1. *** p<0.01, ** p<0.05, * p<0.1. 2. Robust standard errors clustered at household level.

Appendix Figure 1: Distribution of Learning Levels



Notes: Stars denote statistical significance in relation to the control mean. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Figure 2: Demand for Low-tech Services



Notes: Stars denote statistical significance in relation to the control mean. *** p<0.01, ** p<0.05, * p<0.1.