SCHOOL'S OUT: EXPERIMENTAL EVIDENCE ON LIMITING LEARNING
LOSS USING "LOW-TECH" IN A PANDEMIC

Noam Angrist
Peter Bergman
Moitshepi Matsheng

School's Out: Experimental Evidence on Limiting Learning Loss Using "Low-Tech" in a
Pandemic
Noam Angrist, Peter Bergman, and Moitshepi Matsheng
NBER Working Paper No. 28205
December 2020, Revised January 2021
JEL No. I10,I20,I24,I25

ABSTRACT

Schools closed extensively during the COVID-19 pandemic and occur in other settings, such as
teacher strikes and natural disasters. This paper provides some of the first experimental evidence
on strategies to minimize learning loss when schools close. We run a randomized trial of low-
technology interventions – SMS messages and phone calls – with parents to support their child.
The combined treatment cost-effectively improves learning by 0.12 standard deviations. We
develop remote assessment innovations, which show robust learning outcomes. Our findings have
immediate policy relevance and long-run implications for the role of technology and parents as
partial educational substitutes when schooling is disrupted.

Noam Angrist
120 Walton St
Oxford OX2 6GG
United Kingdom
noam.angrist@bsg.ox.ac.uk

Peter Bergman
Department of Education Policy
and Social Analysis
Columbia University
525 W. 120th Street
New York, NY 10027
and NBER
bergman@tc.columbia.edu

Moitshepi Matsheng
Young 1ove
Gaborone, Botswana
and Botswana National Youth Council
mmatsheng@young1ove.org

I. Introduction

The COVID-19 pandemic paralyzed education systems worldwide; at one point, school closures forced over 1.6 billion learners out of classrooms (UNESCO 2020). While smaller in scale, widespread school closures are not unique to COVID-19: teacher strikes, summer breaks, earthquakes, viruses such as influenza and Ebola, and weather-related events cause schools to close. Closures 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 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.

Remote instruction also compelled several innovations in high-frequency, low-cost remote assessment. We adapted the widely used Annual Status of Education Report (ASER) into a phone-based learning assessment.[1] We incorporated time limits and a requirement that

---

[1] The ASER test is in active use in over 14 different countries and has been used consistently in the education literature (Banerjee et al. 2007; Banerjee et al. 2010; Banerjee et al. 2017; Duflo et al. 2020).

children explain their work to accurately identify their numeracy levels. To measure the reliability of our assessment, we randomly assigned problems measuring the same proficiency to students, a reliability test used in the psychometric literature (Crocker and Algina 1986). We further disentangled cognitive skills gains from effort effects, which have been shown to affect test scores (Gneezy et al. 2019). When learning outcomes are measured remotely in the household, effort might be particularly important. We test this hypothesis with a real-effort task. We also measure whether 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. The familiarity hypothesis is also partially tested with randomized problems of the same proficiency. Lastly, we demonstrate the value of high-frequency, remote assessment by using a midline assessment to target content to learning levels for a cross-randomized subgroup of students.

We find large, statistically significant learning differences between treatment and control groups. The combined phone and SMS intervention increases learning by 0.121 standard deviations (p=0.008). The SMS intervention alone has no effect on learning. For households who participated in all sessions, instrumental variables analysis shows learning gains are 0.167 standard deviations (p=0.007). The results for the phone plus SMS intervention translate to solving place values as well as fractions. Since fractions are a high-order learning construct not directly taught during the intervention this reveals skill complementarities, where, for example, learning division also enables learning fractions. Lastly, our results are robust to a series of sensitivity tests, do not appear to be driven by effort, and we find evidence that targeted instruction is more effective than non-targeted instruction.

These results demonstrate that certain types of instruction through "low-tech" mobile phones can provide an effective and scalable method to deliver educational instruction outside of the traditional schooling system and to personalize instruction. The phone and SMS intervention is highly cost-effective with 0.63 to 0.89 standard deviation learning gains per $100 USD. These results also reveal that some level of direct instruction, which can be done cheaply and virtually via phone, can be necessary and that automated SMS messages alone do not produce lasting learning gains. Moreover, parents exhibit strong demand for the intervention, with over 99 percent of households expressing interest in continuing the program

after the first four weeks.[2] Parental engagement with the interventions is high: 92 percent of parents report that their child attempted to solve the problems sent, with slightly higher engagement in the SMS plus phone group of 95 percent. Parents report 8.4 and 15.2 percent greater self-efficacy in supporting their child's learning because 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 parents' return to work post 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.

Our work contributes to several literatures. The low-tech interventions we test 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.[3] We also contribute novel learning data via phone-based assessment.

This paper also relates 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, 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

---

[2] 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.
[3] The role of technology as a complement or substitute for the traditional schooling system is reviewed in Bettinger et al. (2020).

2019). We contribute to this literature by testing a particularly low-cost, and scalable approach to target instruction using phone-based assessments and 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). 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. Moreover, school closures occur in settings beyond the current pandemic, including teacher strikes, summer holidays, public health crises, 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). The results in this paper provide evidence that remote instruction by phone and SMS messages has the potential to operate as a partial educational substitute, limiting children's learning loss when schooling is disrupted using a low-cost and scalable model.

II. Background

A. 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 quality schooling (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). 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.

B. 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. 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 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.[4] Over 99 percent of parents reported demand for continued remote learning services even if schools reopened, likely due to uncertainty around whether schools would remain open, reduced school hours, and disrupted learning.

III. 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. These numbers were collected for students in grades 3 to 5. After phone collection and verification, facilitators called all numbers to confirm interest from parents in receiving remote learning support via phone.

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.

---

[4] 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.

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 S1 shows an example weekly message of practice problems focused on place value.

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. Figure S2 includes a subset of a sample phone call script.

A subset of phone numbers also received 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 used data collected at week four to have near real-time data to target instruction. At approximately week twelve, we collected additional endline survey data and conducted learning assessments which enabled evaluation of the targeted instruction component of the intervention.

IV. Experimental Design

We collected 7,550 phone numbers in primary schools throughout the country the week before the lockdown was instated. 4,550 households were reachable, interested, and gave consent to participate in the trial. For this cohort of 4,550 participants, we include a heat map in

Figure S3 of the location of the children's schools to demonstrate the distribution of participants across the country. Nine out of all ten regions in the country were represented, including the most remote and low-literacy areas.

We randomized the 4,550 phone numbers 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 further cross-randomized 2,250 numbers for a midline assessment, and approximately 1,600 phone numbers receive targeted instruction customized to their learning level using the data collected at midline. Randomization was stratified on whether at least one child in the household had previously participated in prior school-based educational programming, a proxy for having recently made substantial learning gains. Each phone number belongs to a caregiver and household.

Figure S4 provides a timeline of each step from initial phone number collection, piloting and training, program implementation and waves of data collection. Figure S5 provides an overview of the experimental design.

A. 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 own self-efficacy and their child's learning. A portion of the survey was conducted with the parent and learning outcomes were collected by directly assessing the child over the phone.

The learning assessment was adapted from the ASER test, which has been used in over 14 different countries and in the education literature (Banerjee et al. 2017). We adapt this assessment into a phone-based assessment. The ASER test consists of multiple numeracy items, including 2-digit addition, subtraction, multiplication, and division problems. Figure S6 shows a sample assessment. In order to maximize 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. 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. While imperfect, these measures provide a level of verification to maximize the likelihood the test captures child learning. We discuss practical steps to implement learning measurement via phone in Angrist et al. (2020). We also conduct several checks to validate measures, described further below.

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. We include 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 include information on whether the caregiver has returned to work. Finally, demographic questions record the child's age, grade, and gender.

We also conducted a midline assessment used to cross-randomize targeted instruction, described above, and asked about demand for remote learning services if schools were to reopen.

B. Balance and Attrition

Supplement Table S1 shows endline survey response rates. We successfully followed up with 64.9 percent of households. Nearly all of these 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 that, among respondents, analysis on endline outcomes is unbiased across study groups.

Supplement Table S2 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 provide evidence that randomization was successfully implemented.

V. 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 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 identifies causal effects of the intervention.

VI. Results

For our three primary outcomes—average level, place value and fractions—Figure 1 (see also Table S3) shows large, statistically significant learning differences between treatment and control groups after four months. 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 0.114 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 the SMS-only intervention across all learning proficiencies.

The results suggest that combined phone and SMS "low-tech" interventions can generate substantial learning gains. These results also reveal that some level of direct instruction, which can be done cheaply and virtually via phone, can be necessary and that automated SMS messages alone do not produce lasting learning gains. 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. A review by Evans and Yuan (2020) also finds 0.10 median effect sizes across 130 RCTs in low- and middle-income countries. Our findings show effect sizes that are around or above the median effect size with a relatively cheap and scalable intervention.[5]

We run a series of validity checks for our remote assessments and treatment effects. First, we randomized problems that test the same proficiency, a version of a reliability test used in the psychometric literature (Crocker and Algina 1986). We randomize 5 problems for each proficiency including for addition, subtraction, multiplication, division, and fractions. Table 1 shows results. We find that each random problem across all proficiencies is not statistically significantly different relative to a base random problem.[6] 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 which 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 2, 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. Columns (3) through (4) contrast the lack of significant effect on effort by exploring effects across different skills, showing reductions on innumer-

---

[5] Of note, the learning gains observed might be driven by either learning gains, minimizing learning loss, or a combination of both.
[6] Relatedly, we find no difference in treatment effects by the random question received for each proficiency. Results available on request.

acy and improvements on division. Column (2) shows the effect on average level as a reference. 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 as on the endline assessment. 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, which does not change our 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 can translate to this skill.

Lastly, we demonstrate the validity and benefits of high-frequency, low-cost remote assessment by testing whether it can be used to target instruction effectively. In Table S3, we show that the effect on average level is similar between targeted and nontargeted instruction, with effects of 0.076 standard deviations (p=0.097) for targeted content compared to 0.070 standard deviation effect (p=0.130) non-targeted content. There is suggestive evidence that targeted instruction increases impacts on learning place values and fractions relative to non-targeted instruction. Targeted instruction improves understanding of place values by 0.098 standard deviations (p=0.026) compared to 0.026 standard deviations (p=0.572) for non-targeted instruction. Targeted instruction also benefits learning higher-order competencies—understanding fractions—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. The p-values for the test of the difference between targeted and non-targeted for these three outcomes—average level, place values and fractions—are 0.896, 0.098 and 0.160, respectively. These results reveal the potential of using a particularly low-cost and scalable approach to target instruction using phone-based assessments.

VII. Mechanisms

A. Engagement and Demand

We explore parental demand and educational engagement mechanisms. Parents exhibit strong demand for the intervention, with over 99 percent of households expressing interest in continuing the program after four weeks. Supplement Table S4 also explores how demand changes as a result of the intervention, with households demanding more of the intervention they received.

Parental engagement in both interventions 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-by-week engagement, defined as spending any time on the phone with the instructor. In Figure 2, we see that weekly engagement is high overall, starting at around 85 percent and at 60 percent by the final week.[7] In addition, we find that the *type* of engagement changes over time, 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. This also suggests the potential for larger treatment-on-the-treated (TOT) effects than earlier reported intention-to-treat effects. We estimate TOT effects by coding a continuous treatment variable for the number of sessions attended and instrument this endogenous variable with treatment assignment. Table S5 in the supplement reports effects of 0.167 (p=0.007) standard deviations for those who participate in all intervention sessions.

B. Parent Beliefs, Self-Efficacy, and Potential Labor Market Crowd Out

The effects on parents elucidate how remote instruction improved learning outcomes. 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 learning might cause

---

[7] 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).

parents to update their beliefs and attenuate 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 3 we see that in the SMS group, students learn, but at best 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. We also find that parents have slightly more accurate beliefs as a result of the phone and SMS intervention. This reveals that more intensive involvement in a child's learning can be important for belief updating.

Table 3 also shows significant increases in parents' self-efficacy and perceptions as a result of both interventions. 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. We also find parents' confidence that their child made progress on their learning increases by 6.6 (p=0.002) to 10.5 (p<0.001) percentage points.

Parents' engagement in their children's learning might displace other activities, such as returning to work when lockdowns were lifted. We find no evidence of such crowd out effects. Rather, in column (5) in Table 3 we find a reduction in parents who remain out of any type of work of 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. We do not focus on explaining these effects; instead, our goal is to test concerns that additional 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. Altogether, these results show that remote instruction can change parental investments, which play an important role in their child's learning.

VIII. Cost-effectiveness

Both low-tech interventions are relatively low cost. For the SMS-only treatment arm, the total cost was about $7,825 USD. For phone calls, the additional cost above the bulk text message was $28,775. This equates to $5 per child in the SMS group and $19 dollars per child 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 gains for the phone and SMS group per $100 USD. For those who 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 and remedial tutoring in India yielded around a standard deviation per $100 (Kremer, Brannen, and Glennerster 2013). Another relevant cost-effectiveness comparison is tutoring programs.[8] A recent review by Nickow, Oreopoulos, and Quan (2020) shows that tutoring programs have been consistently effective across 96 randomized trials. The phone call intervention in our trial compares closely to one of these tutoring programs which yielded 0.19 to 0.31 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.

IX. Conclusion

This paper provides some of the first experimental estimates on minimizing the fallout of the COVID-19 pandemic on learning. We show that remote instruction and remote assessment can promote learning. We find that low-tech phone calls plus SMS interventions have large and cost-effective effects on household engagement in education and learning, while SMS messages alone do not. For the former, we find 0.12 to 0.17 standard deviation gains and up to .89 standard deviation gains per $100.

We also show how high-frequency, remote assessments can be used to target instruction and find suggestive evidence that targeted interventions outperform non-targeted 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

---

[8] Carlana and La Ferrara (2020) evaluate remote tutoring with college students in Italy during covid-19 with results pending.

cognitive skills with real-effort tasks. We further find that gains persist in the phone and SMS treatment across multiple waves of assessment.

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.

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 COVID-19 pandemic, including teacher strikes, summer holidays, public health crises, during adverse weather events, natural disasters, and in refugee and conflict settings. In moments when schooling is disrupted, particularly for families with fewer resources at home, outside-school interventions are needed. Doing so at scale requires cheap, low-technology solutions that can reach as many families as possible. To this end, the results from this trial have long-run implications for the role of technology and parents to serve as partial educational substitutes during school closures and provide cost-effective remote instruction and assessment.

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." *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. 2020 "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. 1986. *Introduction to Classical and Modern Test Theory.* Fort Worth: Holt, Rinehart, and Winston.

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.

Evans, David K., and Fei Yuan. 2020. "How big are effect sizes in international education studies." Center for Global Development, Working Paper 545.

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

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.

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.

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

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.
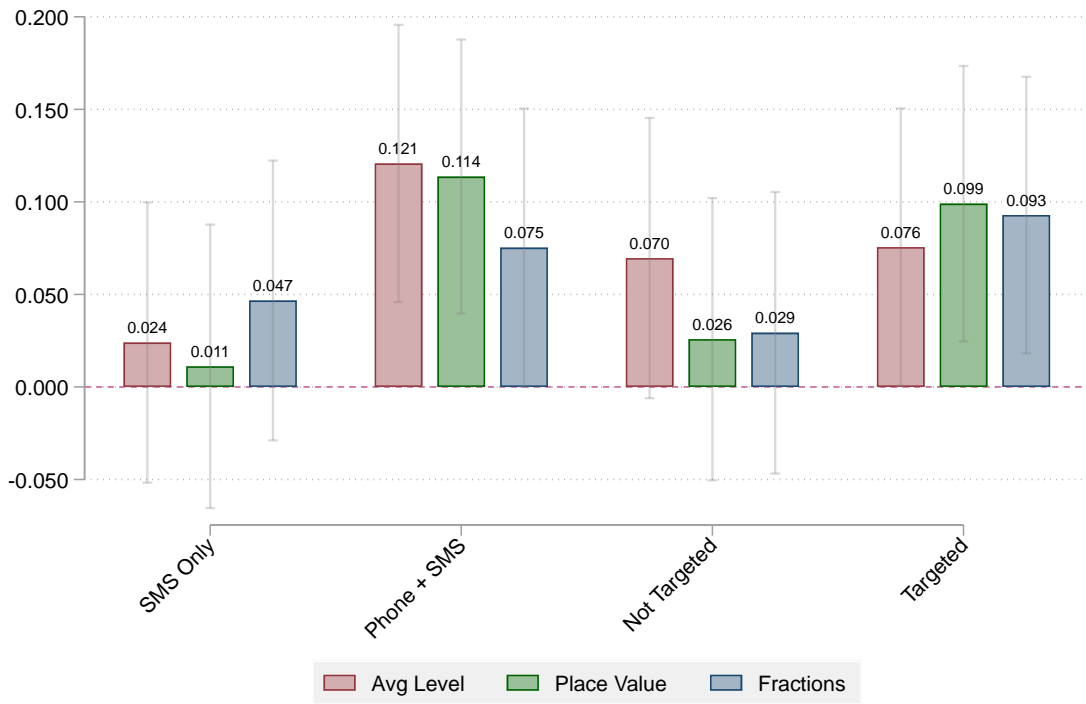
FIGURE 1 - TREATMENT EFFECTS ON LEARNING OUTCOMES

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.

TABLE 1 – 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 2 – 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.
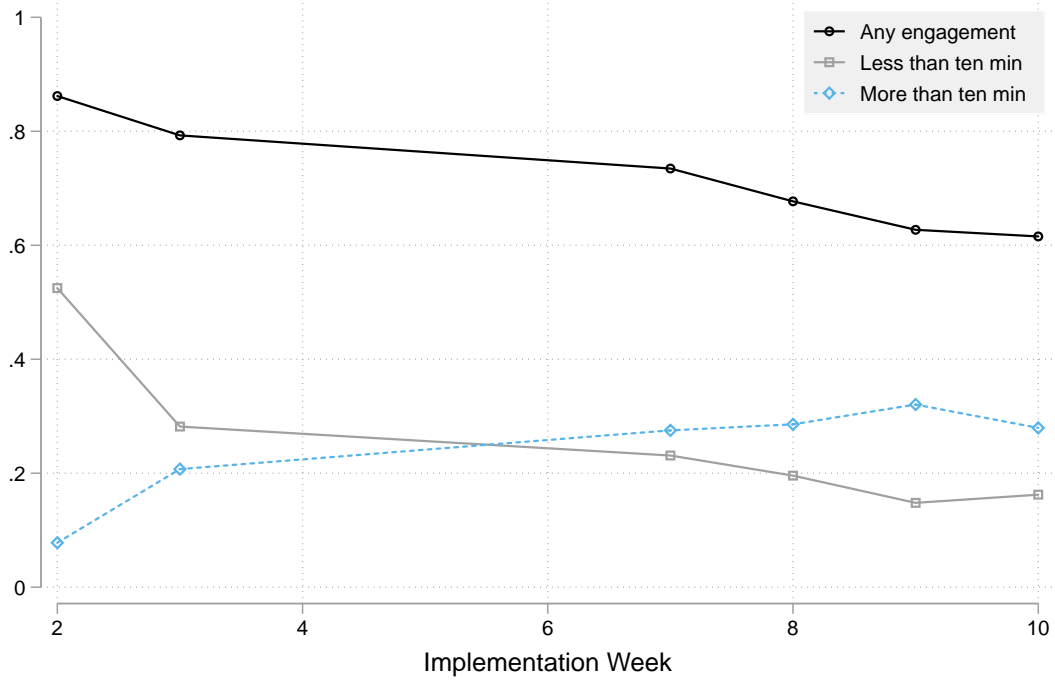
FIGURE 2 – WEEK ON WEEK ENGAGEMENT IN THE PHONE AND SMS TREATMENT

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).

TABLE 3 – PARENT MECHANISMS: BELIEFS, SELF-EFFICACY, AND POTENTIAL CROWD OUT

	(1)	(2)	(3)	(4)	(5)
	Parent Reported Child Level	Parent Correct about Child Level	Parent Self-Efficacy	Parent Perception that Child Learned	Parent Did not Return to Work
Panel A					
SMS Only	0.025	-0.012	0.049	0.066	-0.000
	(0.050)	(0.023)	(0.021)	(0.022)	(0.018)
	[0.621]	[0.594]	[0.023]	[0.002]	[0.994]
Phone + SMS	0.153	0.039	0.086	0.105	-0.029
	(0.050)	(0.023)	(0.021)	(0.022)	(0.017)
	[0.002]	[0.099]	[0.000]	[0.000]	[0.092]
Panel B					
Not Targeted	0.050	-0.001	0.050	0.071	-0.010
	(0.051)	(0.023)	(0.022)	(0.022)	(0.018)
	[0.323]	[0.957]	[0.020]	[0.001]	[0.565]
Targeted	0.125	0.028	0.084	0.099	-0.018
	(0.049)	(0.023)	(0.021)	(0.022)	(0.017)
	[0.012]	[0.239]	[0.000]	[0.000]	[0.296]
Control Mean	2.500	0.398	0.566	0.492	0.190
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	2957	2650	3127	3127	2990
p-val: SMS = Phone	0.009	0.029	0.071	0.075	0.088
p-val: Targeted = Not Targeted	0.128	0.217	0.115	0.194	0.640

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. This table also 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 bracket.

Online Appendix

Supplement Figure S1: Intervention SMS Text Message Example
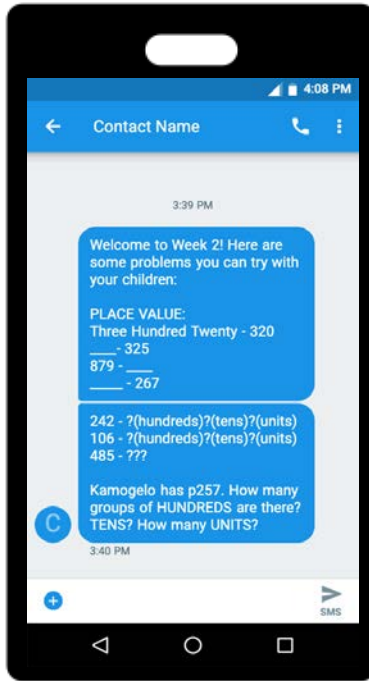
Supplement Figure S2: Sample phone call introduction

Supplement Figure S3: Distribution of Schools of Student Participants across Botswana
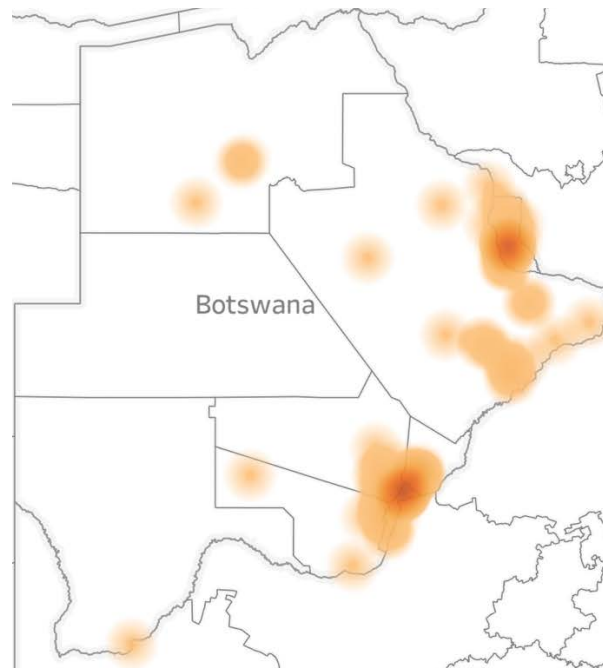
Notes: this density map of schools in Botswana shows the relative distribution of schools linked to students in our sample. Darker regions correspond to higher concentrations of schools for study participants. The sample in the study includes nearly all regions in Botswana (9 out of 10).

Supplement Figure S4: Intervention and Evaluation Timeline

Schools Closed

Schools Partially Open

Program Launch

Midline

Endline

Mid April

Late May

Late July
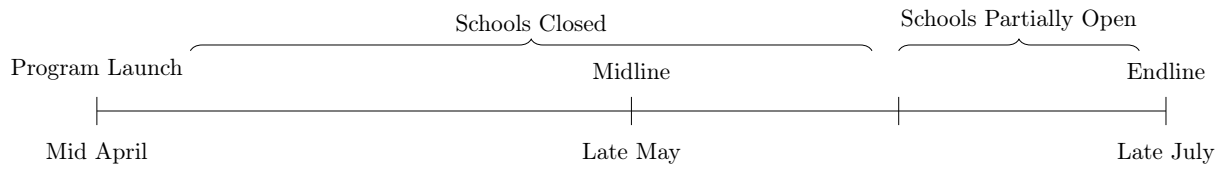
Notes: All dates refer to the year 2020.

Supplement Figure S5: Experimental Design

Numbers Collected — 7,550

Sensitization Data Collected — 6,375

Enrolled — 4,550

Control	SMS Only	Phone + SMS
1,516	1,516	1,518

Control	SMS Only	Phone + SMS	Control	SMS Only	Phone + SMS
758	758	758	758	758	760

Wave 1 Data Collection

Targeted SMS Messages

Wave 2 Data Collection
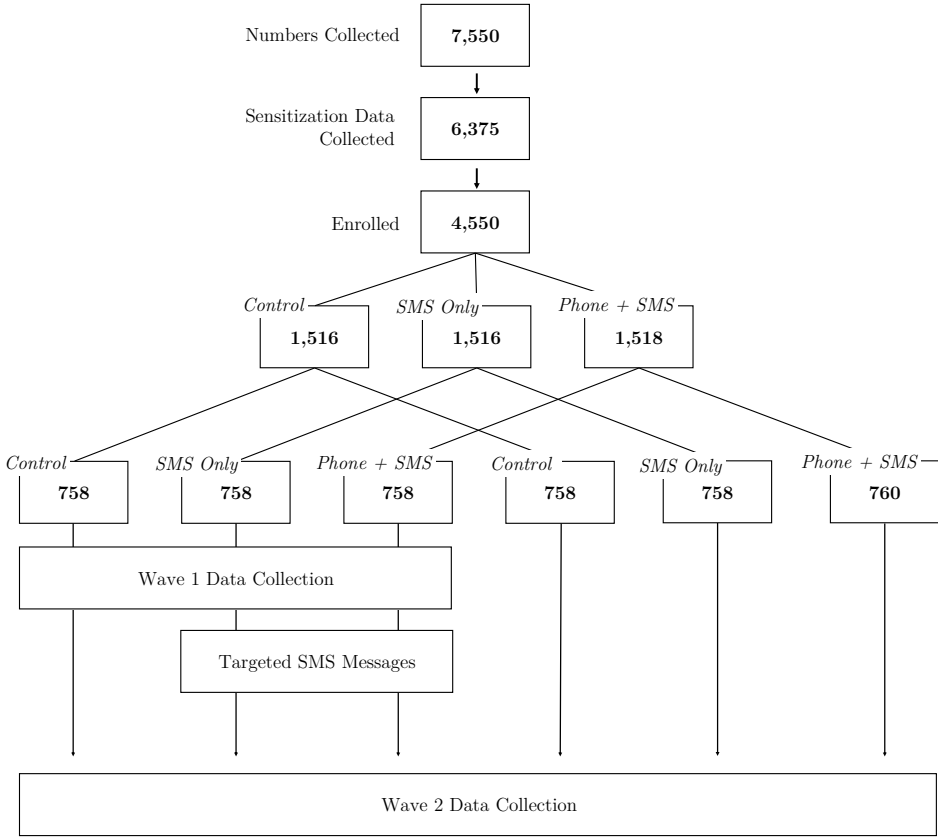
Notes: Counts represent the quantity of phone numbers. Each phone number corresponds to one household.

Supplement Figure S6: Sample of ASER test used in Botswana

Levelling Tool (Version 5)
Basic Operations

62	33	16		91	42	81
+ 18	+ 49	+ 47		− 52	− 38	− 43

26	38	12		6) 93	4) 53	3) 49
x 3	x 2	x 5				

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.

Supplement Table S1 – 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 = Not 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.

Supplement Table S2 – 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 = Not 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.

SUPPLEMENT TABLE S3 – TREATMENT EFFECTS ON LEARNING OUTCOMES

	(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 = Not 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.

	Engagement	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 half-way 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.

SUPPLEMENT TABLE S5 – TREATMENT ON THE TREATED EFFECTS

	(1) Avg Level	(2) Avg Level	(3) Avg Level
Phone + SMS	0.121 (0.046) [0.008]		
Phone + SMS - Per Session		0.028 (0.010) [0.007]	
Phone + SMS - All Sessions			0.167 (0.062) [0.007]
Observations	2815	1878	1878

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)-(3) than Column (1) since we exclude the SMS group in the regression. Standard errors are in parentheses and p-values are in square brackets