

Low Cost, High Reach

Mobile interventions for foundational learning during school closures

Deepak Kumar¹ Naveen Sunder² Ruchi Chaudhary³
Siddhesh Mhatre³ Rukmini Banerji³ Wilima Wadhwa^{3,4}

This version: August 26, 2026

Abstract

India's schools were closed for 82 weeks during COVID-19; when they reopened in early 2022, children returned far behind on foundational skills. We ask whether the cheapest channel available, a text message to any mobile phone, can help them catch up, and how much a brief human follow-up adds. In a cluster-randomized trial in 1,162 villages across 19 Indian states, run in the first weeks after schools reopened, about 27,000 children in Grades 3–5 who could not yet subtract received one numeracy activity a day by SMS for four weeks, matched to each child's level and written for a caregiver to do with the child. In a second arm, a field worker from the implementing NGO, already known to the family, also called once a week to ask whether the activities were being done, but did not teach. Messages alone raise arithmetic by 0.14 standard deviations and the share of children who can subtract from 14 to 19 percent; with the calls, the gains rise to 0.24 standard deviations and 23 percent. There is also a spillover to reading, which the program did not address, of 0.07–0.15 standard deviations, and the gains are present in households with only a basic phone. Where one-way texting has mostly failed, level-matched activities from a trusted organization succeed, and a brief call with no instruction adds a further 0.10 standard deviations. The messages were virtually free to send, and the calls cost well under a dollar per child.

Keywords: Foundational learning; Mobile phones; SMS; Parental engagement; Teaching at the Right Level; COVID-19; India

JEL Codes: I21, I24, I25, O15

¹Department of Economics, Delhi School of Economics, University of Delhi.

²Department of Economics, Bentley University, Waltham, MA, USA. nsunder@bentley.edu.

³Pratham Education Foundation, New Delhi.

⁴All errors are our own.

1 Introduction

India’s schools were closed, fully or in part, for 82 weeks between March 2020 and early 2022, the second-longest closure in the world (UNESCO, 2022). When children returned, foundational learning had dropped considerably. In rural India the share of Grade 3 children who could do a two-digit subtraction fell from 28 to 26 percent between 2018 and 2022, and the share who could read a Grade 2 text fell from 27 to 21 percent. The decline was steeper for older children: among Grade 5 children the share who could read a Grade 2 text fell from 51 to 43 percent, and the share who could do a three-digit by one-digit division fell from 28 to 26 percent (ASER Centre, 2023). The pattern is not specific to India. Across 42 studies in 15 countries, pandemic learning deficits average 0.14 standard deviations (SD), are larger in mathematics than in reading, and are largest in middle-income countries (Engzell et al., 2021; Betthäuser et al., 2023; Moscoviz and Evans, 2022); they do not fade on their own (Lichand et al., 2022; Lichand and Doria, 2024; Singh et al., 2024). A school system in this position faces a practical question. What is the cheapest tool that still moves foundational skills, and can it reach homes that have no internet, no smartphone, and parents who may not read well themselves?

Text messaging is the natural place to look. India had about 1.1 billion mobile subscriptions in 2022, close to half of them rural (Telecom Regulatory Authority of India, 2022), and a text message arrives on the cheapest handset. The evidence on texting is discouraging, however. In Botswana, practice problems sent by SMS alone had no measurable effect (0.02 SD, SE 0.05), while the same messages paired with a twenty-minute tutoring call raised learning by 0.12 SD (Angrist et al., 2022). In five later trials in India, Kenya, Nepal, the Philippines and Uganda, phone tutorials raised test scores by 0.30–0.35 SD while SMS arms gave mixed results (Angrist et al., 2023). A review of thirty studies concludes that one-way texting is rarely enough on its own and that what works is a person on the other end of the line (Aker and Angrist, 2026). That person is costly, and not every call helps: live tutoring calls from teachers in Sierra Leone left test scores unchanged at \$40 per child (Crawford et al., 2023), and in Kenya brief accountability calls did nothing while short tutoring calls lowered the scores of children who later returned to school (Schueler and Rodriguez-Segura, 2023). Little is known about what lies between the two: messages written so that a caregiver can act on them, combined with human contact cut back to a few minutes a week. That is the combination we test.

This paper asks two questions. Can a text-message program whose content is matched to each child’s starting level raise foundational numeracy at national scale, delivered to basic phones by an organization that families already know? And how much does the lightest possible human follow-up—a brief weekly call that checks in but does not teach—add on top?

We answer them with a cluster-randomized trial run by Pratham, India’s largest education

NGO, in 1,162 villages across 170 districts and 19 states in March and April 2022, as schools were reopening. Pratham staff tested every child in Grades 3–5 in these villages one-on-one with the ASER arithmetic tool and enrolled the 30,955 who could not yet do two-digit subtraction. Villages were assigned, within blocks, to one of three arms. In the SMS-only arm, households received one short numeracy activity by text message six days a week for four weeks, in the state language, drawn from one of two sets matched to the child’s level: one for children who could not yet read two-digit numbers and one for those who could. The activities are things a caregiver can do with a child at home—count seeds held in two fists, read a number chart, walk twelve steps east and thirteen back and add them up. In the SMS+calls arm, households received the same messages and, once a week, a phone call from the Pratham staff member responsible for the village. The caller asked three questions—did the message arrive, did the child try the activity, who helped—and was told not to teach. Control villages received nothing. Four weeks later the same staff re-tested 26,751 children.

The messages alone raise arithmetic scores by 0.14 SD (SE 0.02). The share of children who can subtract rises from 14 percent in control villages to 19 percent, a relative gain of 35 percent, and the share who can read two-digit numbers rises by 6 percentage points. Adding the weekly call raises the effect to 0.24 SD and the subtraction share to 23 percent. The 0.10 SD gap between the two arms is precisely estimated ($p < 0.001$) and appears on every arithmetic outcome. There is a spillover to reading, which the program did not address: reading scores rise by 0.07 SD in the SMS-only arm and by 0.15 SD with calls. Effects are present in households with only a basic phone (0.12 and 0.20 SD) and are 0.06–0.07 SD larger where the household owns a smartphone. We find no differences by gender, grade or school type that can be distinguished from zero; the confidence intervals rule out gender gaps in the effect larger than about 0.07 SD. Against the distribution of effects from 96 learning RCTs in low- and middle-income countries, whose median is 0.10 SD and whose 75th percentile is 0.23 (Evans and Yuan, 2022), the SMS-only effect is above the median and the SMS+calls effect sits at the upper quartile—from a program with no teacher, no tutor, no device and no internet. In raw terms the effects are modest and in line with the closest studies: about a tenth of an ASER level with messages and a fifth with calls, against 0.15 levels for messages with tutoring calls in Botswana. Standard deviation units flatter them because the outcome is a five-level scale measured on children who all started at its lower levels (Section 5.1). Delivery cost almost nothing: the messages were virtually free to send, and the four weekly calls cost well under a dollar per child (Section 6).

What the call adds is not more content. Receipt of the messages is the same in both arms (93 and 95 percent, a difference we cannot distinguish from zero). What changes is what happens after the message arrives: households in the call arm are 10 percentage points more likely to report that someone helped the child with the activities (93 against 83 percent), and the

spillover to reading doubles. The data cannot separate the channels through which a brief call does this—a reminder that makes the task salient, accountability to a person the family knows, or the referral to a community volunteer that callers could offer—and the engagement measures are self-reports collected by the staff who made the calls. What the experiment does establish is that, holding content fixed, a non-instructional contact of a few minutes a week is worth 0.10 SD, about two-thirds of what the messages themselves deliver.

The paper makes three contributions. First, it shows that text messages alone can raise learning, at a scale no earlier SMS-only evaluation approaches, in a class of programs where one-way messaging has mostly failed. Our design differs from the programs that failed in three ways that the experiment cannot separate: the content is an activity for a caregiver to do with the child rather than a problem for the child to solve or a fact for the parent to know; it is matched to the child’s level; and it comes from an organization that had worked in these villages before. Two pieces of evidence point to targeting as the ingredient that mattered most. Pratham had sent untargeted activities to communities like these throughout 2020–21, and targeting by level was the one change it made for this trial. And the closest experiments point the same way: in the United States, text messages to parents that were differentiated by the child’s skill level outperformed identical messages that were not (Doss et al., 2019), and in Botswana effects were larger, though imprecisely so, where content was targeted (Angrist et al., 2022). We treat targeting as the most plausible explanation rather than as a finding; identifying it would require an untargeted arm. What this study adds to the low-tech learning literature (Aker et al., 2012; Aker and Ksoll, 2019, 2020; Escueta et al., 2020; Rodriguez-Segura, 2022) is a positive and precisely estimated SMS-only effect from 1,162 villages, together with a description of what such a program has to look like.

Second, the paper prices the human touch. Earlier work bundles the call with instruction: in Botswana, Sierra Leone and the five-country trials the caller taught, so the comparison is between tutoring and no tutoring (Angrist et al., 2022; Crawford et al., 2023; Angrist et al., 2023); the interactive programs that worked best in Bangladesh were teaching programs delivered by phone (Hassan et al., 2024; Wang et al., 2024), as is the online tutoring that worked in Italy (Carlana and La Ferrara, 2025). Our call arm holds the content fixed and removes the teaching, leaving a contact of a few minutes with a known person. The resulting 0.10 SD is close to the gap between SMS-only and SMS-plus-tutoring in Botswana (0.10 SD), which suggests that much of what a tutoring call adds to good content may not be the tutoring. This speaks to the literature on parents as a channel for learning, in which low-cost remote programs—text nudges, information about attendance and grades, parenting tips—have a bias-corrected mean effect of 0.05 SD on test scores across 82 trials (Bergman and Chohanajin, 2026); the underlying evidence runs from information given to parents in Malawi and Chile (Dizon-Ross,

2019; [Berlinski et al., 2025](#)) to a largely American literature on texting and teacher-to-parent communication ([Bergman, 2021](#); [York et al., 2019](#); [Kraft and Rogers, 2015](#)). Our caregivers are rural, many with little schooling, and the content asks them to do something rather than to know something; the effects are three to five times the meta-analytic mean, and the reading spillover from arithmetic-only content points to a general rise in household engagement rather than to transfer of content, in line with evidence that engaging mothers in children’s learning raises outcomes in rural India ([Banerji et al., 2017](#)). What this adds is an estimate of what a minimal human contact is worth on top of content, which no earlier design isolates.

Third, the program is one a government could run, and the study shows it running at close to government scale. The sample spans 19 states and 11 languages; delivery ran through 1,146 of Pratham’s own field staff under one protocol; half of the households had never received a Pratham message before; and the study window is the reopening period, when the question for a school system is what complements the classroom rather than what replaces it ([Aker and Angrist, 2026](#)). Effects are similar across school sector, gender and grade, and present without a smartphone, which bears on the concern that remote-learning programs reach mainly the better-off ([Beam et al., 2025](#); [Agostinelli et al., 2022](#)). Whether results from small trials survive implementation at scale is a standing question in this literature ([Muralidharan and Niehaus, 2017](#); [Bold et al., 2018](#); [Vivalt, 2020](#)); this trial was run at scale from the start, through the organization that would implement it, and its ingredients—a one-on-one level assessment, two sets of messages, a weekly check-in—map directly onto the assessment capacity that India’s national foundational-learning mission is now building ([Ministry of Education, Government of India, 2021](#)). In that sense the paper also adds to the evidence on how education systems can be made resilient to disruption ([Angrist et al., 2023](#)). It extends the Teaching at the Right Level approach, whose in-person and computer-assisted versions are among the best-documented ways to raise foundational skills ([Banerjee et al., 2007, 2017a,b](#); [Duflo et al., 2011](#); [Muralidharan et al., 2019](#); [Duflo et al., 2024](#); [Angrist et al., 2025](#)), to the thinnest delivery channel yet tested.

Three things should be said at the outset, and we return to them at the end of Section 5. As in most evaluations run with an implementing organization, the field staff who delivered the program also conducted the endline assessment, and the organization’s prior presence in these villages is part of what we evaluate. We observe outcomes at the close of the four-week program only. And the design does not separate level-matching from the other features of the messages.

Section 2 describes the setting and the program, Section 3 the design and data, Section 4 the empirical strategy, Section 5 the results, Section 6 costs, and Section 7 concludes.

2 Setting and Program

2.1 Learning after the closure

India’s foundational learning problem predates the pandemic. Enrollment in rural India has exceeded 95 percent for more than a decade, but in 2018 only 28 percent of Grade 3 children could do a two-digit subtraction and only 28 percent of Grade 5 children could do a simple division (ASER Centre, 2023). Much of the gap reflects a curriculum that moves faster than children do, compounded by large classes and high teacher absence (Chaudhury et al., 2006; Banerjee et al., 2017a). The government’s NIPUN Bharat mission, launched in 2021, made foundational literacy and numeracy by Grade 3 a national goal (Ministry of Education, Government of India, 2021).

The closure made the problem worse and changed where learning had to happen. Government primary schools closed in March 2020 and, in most states, did not reopen for regular classes until late 2021 or early 2022 (UNESCO, 2022). Households were asked to substitute for schools with little to work with. By 2021 two-thirds of enrolled rural children had a smartphone at home, up from 37 percent in 2018, but a quarter of those children had no access to it, and the share of children receiving any learning support at home fell from three-quarters to two-thirds as schools began to reopen (ASER Centre, 2021). Families with means turned to private tuition, whose take-up rose from under 30 to about 40 percent (ASER Centre, 2021); others had nothing. When the 2022 survey came, reading and arithmetic levels had fallen back to where they had been a decade earlier (ASER Centre, 2023). The pattern—losses concentrated among children whose parents could least compensate—is the one models of the closure predict (Agostinelli et al., 2022) and the one panel data from Tamil Nadu document (Singh et al., 2024).

What the school system could not offer through its own infrastructure was a digital channel. In 2021–22 only 24 percent of government schools had an internet connection (Ministry of Education, Government of India, 2022). What nearly every household did have was a mobile phone. India had about 1.1 billion wireless subscriptions in March 2022, close to half of them rural (Telecom Regulatory Authority of India, 2022), and a text message requires neither a smartphone nor data. That is the channel this program uses.

Pratham, which ran the program, is India’s largest non-governmental education organization. It developed the ASER assessment and the Teaching at the Right Level approach, in which children are grouped and taught by what they can do rather than by grade (Banerjee et al., 2007, 2017b). During the closure its field staff stayed in touch with families by phone. At the peak, daily or alternate-day activities went out by SMS and WhatsApp to about 481,000 contacts in 13,273 villages across 21 states. Those messages were the same for every child in

Grades 1–8, and their purpose was to keep families engaged rather than to teach; no attempt was made to measure learning. The trial reported here was Pratham’s first test of whether messages could teach, and the one change it made was to match them to the child’s level.

2.2 The program

Target group: The program was aimed at children in Grades 3–5 who, on the ASER arithmetic tool, could not yet do a two-digit subtraction with borrowing. That is roughly two-thirds of rural children in these grades and the group for whom foundational remediation matters most.

Messages: Each enrolled household received one activity by SMS a day, six days a week, from 30 March to 23 April 2022. Because the program began on a Wednesday, the first week ran for four days and the remaining three for six, so a child was sent 22 messages in all. Pratham’s content team in Delhi wrote two sets of messages in Hindi, one for children who could not yet recognize two-digit numbers (the “beginner” and “one-digit” levels of the ASER tool) and one for children who could but could not subtract. State teams translated them into the state language; messages went out in Hindi, Assamese, Bengali, Gujarati, Kannada, Marathi, Odia, Punjabi, Tamil, Telugu and Urdu. The activities were adapted from the worksheets Pratham uses in its in-person children’s clubs and had been piloted in small samples during the closure. Each asks a caregiver to do something with the child using what is at hand: seeds, leaves, steps, a hand-drawn number chart. Table 1 gives two messages from each set. Difficulty rises over the week and is higher in the second set. The message is meant to be read aloud to the child, so the caregiver needs to read but the child does not, and nothing requires a smartphone or an internet connection.

Calls: In the second treatment arm the same messages were accompanied by one phone call a week from the Pratham field worker responsible for the village. Families knew these workers, who had run Pratham’s school and community programs in the village before the pandemic, which is also why Pratham had their phone numbers. The callers were given an orientation and a short list of questions rather than a script. They greeted the parent and asked whether the previous day’s message had arrived, whether the child had tried the activity, and who had helped. They were instructed not to help with the arithmetic. If a parent said they could not support the child, the caller could point them to a volunteer in the community who could. Call days were fixed by the child’s level and were the same in every state—beginners on Tuesdays, one-digit recognizers on Thursdays, two-digit recognizers on Saturdays—so that each call asked about a message of the same difficulty sent the day before. Callers recorded the three answers on a weekly form. Calls were meant to be brief; we did not record their length.

Table 1: Example SMS activities, by level group and day (English translation)

	Beginner and one-digit group	Two-digit group
Week 2, Day 1 (Monday)	Count from 1 to 20 with the child. First you say “1,” then the child says “2,” and continue counting by taking turns. Now count backwards as well.	Make a number chart from 1 to 100. Read the chart with the help of an elder in the house: from top to bottom, bottom to top, right to left, and by picking numbers at random.
Week 2, Day 6 (Saturday)	Ask the child to take grains or seeds in both fists. Ask the child to say how many seeds there are in total across both fists. Then ask the child to add them up.	Walk 12 steps towards the east. Then turn around and walk 13 steps. Say the total number of steps you walked. Now solve this in your notebook as well and share the answer with everyone.

Notes: English translations of messages from the second week of the program. The beginner and one-digit group comprises children who could not recognize two-digit numbers at baseline (ASER arithmetic levels 0 and 1); the two-digit group comprises children who could recognize two-digit numbers but could not do a two-digit subtraction (level 2). Messages were written in Hindi by Pratham’s central content team and translated by state teams into the state language. One message was sent per day, six days a week, from 30 March to 23 April 2022; the first week ran for four days and the remaining three for six. Within each week the activities become harder from Monday to Saturday.

Control: Households in control villages were tested at baseline and endline and received no messages or calls. Pratham also suspended its other activities for children in Grades 3–5, such as library groups, in every study village regardless of arm, so that the messages and calls were the only inputs that differed across arms. Schools were reopening during the study, so children in all three arms had whatever instruction their schools and homes were providing; the treatment effects we report are the value added over that.

3 Experimental Design and Data

3.1 Sample, selection and randomization

The evaluation villages are a subset of the communities in which Pratham had an established presence. Two criteria were applied: the village had to be one where Pratham already worked, so that it had contact numbers and could reach families, and one where its team had done no direct instructional work with children in the 2020–21 or 2021–22 school years, so that the program was the only Pratham input children received during the study. The resulting sample is 1,162 villages in 170 districts and 19 states.

Villages were assigned to three arms by the ASER Centre, Pratham’s research unit, using a computer-generated sequence with two rules: each arm should receive a third of the villages, and the villages within each administrative block should be split equally across arms. Because the number of villages in a block is not always a multiple of three the arms are not exactly

equal: 380 villages in the control arm, 393 in the SMS-only arm and 389 in the SMS+calls arm. No village was assigned by hand. The study was not pre-registered; the questions it was designed to answer—whether level-matched messages raise arithmetic levels over a short period, whether weekly calls add to that, and who helps children at home—were fixed before the baseline, but no analysis plan was written.⁵ We therefore treat the main effects and the contrast between arms as confirmatory and the subgroup analyses as exploratory.

Each village was assigned to one Pratham field worker, and each worker handled villages of one arm only.⁶ Between 25 and 29 March 2022 these workers visited households door to door and tested every child in Grades 3–5 who would remain in those grades for the duration of the exercise, using the ASER arithmetic and reading tools. They were told only which villages they were responsible for, not the arm, so the baseline was collected blind to assignment. A household was asked at the door whether it had a phone on which it could receive messages; children in households without one were not tested or enrolled, since they could not receive the program, and no count of them was kept. Of 45,538 children tested, 30,955 scored below the subtraction level and form the study sample (Table 2). Assignments were released to the workers on 29 March, after testing had finished, and the messages began the following day. The same workers re-tested the children between 25 and 28 April, two days after the last message. The endline was therefore conducted by staff who knew the arm of their village; we return to this at the end of Section 5.

3.2 Outcomes

Learning is measured with the ASER tools, the one-on-one household assessments Pratham developed and that are used in India’s annual rural education survey and in economics research on Indian schooling (Singh et al., 2024; Kumar et al., 2024). The arithmetic tool places a child at the highest of five levels: (0) cannot recognize single-digit numbers; (1) recognizes single-digit numbers; (2) recognizes two-digit numbers; (3) can do a two-digit subtraction with borrowing; (4) can do a three-digit by one-digit division. Reading is measured with a modified ASER language tool whose reading component has five parallel levels: (0) beginner; (1) letters;

⁵The evaluation was designed and fielded within weeks, inside Pratham’s own response to the learning situation its teams were seeing as schools reopened: study villages were finalized on 22 March, the baseline began on 25 March, and the first messages went out on 30 March 2022. Data collection was confined to what that operational window allowed, and we note in the text the places where a purpose-built research study would have measured more.

⁶Pratham’s operational unit is a community, which is a village or a habitation within one; randomization was carried out over the 1,162 units listed in Table 2. Of these, 1,148 contain at least one child in the study sample and 1,121 contain at least one child tested at endline, so the estimation sample has 1,121 clusters. Because some units fall in the same revenue village, the number of distinct villages is about thirty to forty lower at each stage. Clustering standard errors by unit or by revenue village gives the same results, and we cluster by unit throughout.

Table 2: Sample: children at each stage, by arm

	Villages	Tested at baseline	Study sample	Tested at endline	Follow-up rate
Control	380	14,908	10,346	9,009	0.87
SMS only	393	16,009	10,855	9,202	0.85
SMS + calls	389	14,621	9,754	8,468	0.87
Total	1,162	45,538	30,955	26,679	0.86

Notes: *Tested at baseline:* all children in Grades 3–5 in participating households assessed with the ASER tools between 25 and 29 March 2022. *Study sample:* children whose baseline arithmetic level was below two-digit subtraction (ASER levels 0–2); these children were enrolled in the program. Households without a working phone were not enrolled. *Tested at endline:* study-sample children re-assessed four weeks after the program began. *Follow-up rate:* endline tested divided by study sample. Villages were assigned to arms within administrative blocks. Of the 1,162 units randomized, 1,148 contribute a child to the study sample and 1,121 a child tested at endline.

(2) words; (3) a short paragraph; (4) a short story. The tool also opens with a listening-comprehension question, which we do not use. Two content-distinct versions of each tool were prepared; a child received one at baseline and the other at endline, which limits practice on specific items. Responses were recorded on Pratham’s digital data platform.

Our primary outcome is the arithmetic level expressed as a z -score, standardized to have mean zero and standard deviation one in the control group at endline. Because the levels are ordered categories, we also report, for each threshold, an indicator for whether the child has reached it, which has a direct reading: the share of children who can subtract. Reading is a secondary outcome. The program sent no reading content, so any effect on reading is a spillover, and we treat it as evidence on how the program works rather than as a target it was designed to hit.

3.3 Take-up, and what was and was not measured

There is no administrative record of message delivery or of calls completed. What we have are two questions asked at the endline visit, of the parent where possible and otherwise of the child: “Have you received SMS from Pratham in the last month?” and, if so, whether anyone in the household helped the child with the activities. In the SMS-only arm 93 percent of households report receiving the messages and, among those, 83 percent report that someone helped the child. In the SMS+calls arm the figures are 95 and 93 percent. The difference in receipt is not distinguishable from zero; the difference in help is (Section 5.5). These are self-reports collected by the worker who delivered the program, and we read them as indicative of engagement rather than as measures of it. In the SMS+calls arm the same three questions were also put on each weekly call and recorded on a form, so weekly engagement data exist for that arm; whether a call was completed was not itself recorded, and we therefore do not use the number of calls a household received as a measure of dose.

Contamination across villages—messages or activities reaching control households—would bias the treatment effects toward zero. Each worker handled one arm, messages went to registered numbers only, and children were not attending school regularly during the program, so we think such contamination is small, but we have no direct test.

3.4 Balance and attrition

Table 3 compares the three arms at baseline. Arithmetic and reading levels, their distributions across ASER levels, gender, grade and school type are balanced; a joint test of equality across arms does not reject ($p = 0.37$). Smartphone ownership is two percentage points higher in the SMS+calls arm (54 against 52 percent, $p = 0.05$), which is what one expects across fifteen comparisons; we control for it throughout and the estimates do not move when we do not.

Of the 30,955 children in the study sample, 26,751 (86 percent) were tested at endline. Attrition is 13 percent in the control arm, 15 percent in the SMS-only arm and 13 percent in the SMS+calls arm; neither treatment indicator predicts attrition, with or without baseline controls, and the two arms do not differ from each other ($p > 0.85$; Appendix Table A1). With attrition this balanced, bounds on the treatment effect under selective attrition would be tight; we have not computed them.

Table 3: Balance at baseline

	Control		SMS only		SMS + calls		p -value
	Mean	SD	Mean	SD	Mean	SD	(equality)
<i>Panel A. Learning levels</i>							
Arithmetic level (0–4)	1.36	0.63	1.36	0.62	1.35	0.63	0.374
Beginner (level 0)	0.08	0.27	0.08	0.28	0.08	0.27	0.704
One-digit (level 1)	0.48	0.50	0.48	0.50	0.49	0.50	0.704
Two-digit (level 2)	0.44	0.50	0.44	0.49	0.43	0.50	0.894
Reading level (0–4)	1.35	1.09	1.37	1.11	1.33	1.09	0.421
Beginner	0.23	0.42	0.22	0.42	0.23	0.42	0.636
Letter	0.40	0.49	0.40	0.49	0.41	0.49	0.537
Word	0.21	0.41	0.20	0.40	0.20	0.40	0.556
Paragraph	0.11	0.32	0.12	0.33	0.11	0.31	0.215
Story	0.05	0.21	0.05	0.22	0.05	0.21	0.823
<i>Panel B. Child and household characteristics</i>							
Household owns a smartphone	0.52	0.50	0.52	0.50	0.54	0.50	0.048
Female	0.50	0.50	0.51	0.50	0.50	0.50	0.180
Grade 3	0.40	0.49	0.42	0.49	0.40	0.49	0.428
Grade 4	0.34	0.47	0.33	0.47	0.34	0.47	0.510
Grade 5	0.26	0.44	0.25	0.43	0.26	0.44	0.694
Government school	0.86	0.34	0.87	0.34	0.86	0.35	0.441
Private school	0.13	0.34	0.12	0.32	0.13	0.33	0.554
Children	10,346		10,855		9,754		

Notes: Means and standard deviations of baseline characteristics for the study sample (children below the subtraction level at baseline), by arm. Arithmetic and reading levels are the ASER ordinal levels (0–4); the indented rows give the share of children at each level. *Household owns a smartphone* records the type of handset the field worker noted at enrollment as the one on which the messages would be sent, coded one for a smartphone and zero for a basic phone; either can receive an SMS. The last column reports the p -value from a test that the means are equal across the three arms, with standard errors clustered by community. The joint test across all characteristics gives $p = 0.374$.

4 Empirical Strategy

We estimate intent-to-treat effects by ordinary least squares:

$$y_{ivd} = \alpha_1 \text{SMS}_v + \alpha_2 \text{SMS+Calls}_v + \beta y_{ivd,0} + \mathbf{X}'_{iv} \boldsymbol{\gamma} + \mu_d + \varepsilon_{ivd}, \quad (1)$$

where y_{ivd} is the endline outcome of child i in village v and district d , SMS_v and SMS+Calls_v indicate the arm to which the village was assigned, $y_{ivd,0}$ is the baseline value of the outcome, $\mathbf{X}_{iv}$ contains gender, grade, school type and household smartphone ownership, and μ_d are district fixed effects.⁷ Conditioning on the baseline outcome is the efficient choice when the outcome is noisy and its autocorrelation is moderate (McKenzie, 2012). Standard errors are clustered by community, the unit of assignment (Bruhn and McKenzie, 2009); with more than a thousand clusters the usual asymptotic approximation is reliable, and clustering instead by revenue village leaves the estimates and their standard errors unchanged. The coefficients α_1 and α_2 are the effects of being offered each program, and we test $\alpha_1 = \alpha_2$ in every table.

The intent-to-treat effect is the policy-relevant quantity, since a government or NGO can offer the program but cannot make families use it. Because take-up was high it is also close to the effect on those who received the messages: dividing the intent-to-treat estimates by the 93 and 95 percent receipt rates gives effects on recipients of about 0.15 SD in the SMS-only arm and 0.25 SD in the SMS+calls arm.

Heterogeneity is examined by adding to equation (1) interactions of the two arm indicators with a child or household characteristic. We look at baseline arithmetic level, gender, school type, smartphone ownership and grade. These analyses were not pre-specified. Across the four characteristics other than baseline level there are 40 interaction coefficients, and for these we report sharpened false discovery rate q -values (Benjamini et al., 2006; Anderson, 2008) alongside conventional p -values; the q -value is the smallest expected share of false discoveries at which the interaction would be called significant.

5 Results

5.1 Arithmetic

Table 4 reports the main estimates. The messages alone raise the arithmetic z -score by 0.144 SD (SE 0.024). Adding the weekly call raises it by 0.239 SD (SE 0.024). The two arms differ by 0.095 SD, and the difference is precisely estimated ($p < 0.001$). Columns (2)–(4) translate

⁷Randomization was stratified by administrative block, which is nested within district. District fixed effects therefore absorb part rather than all of the stratification; this affects precision, not unbiasedness, and the estimates are unchanged without fixed effects or controls.

these into levels. At endline, 14 percent of control children can do a two-digit subtraction; the messages raise this by 4.9 percentage points and messages with calls by 8.6 points, relative gains of 35 and 61 percent. The share who can recognize two-digit numbers, 51 percent in the control arm, rises by 5.6 and 8.2 points. Almost every child can recognize single-digit numbers by endline (95 percent), so there is little room to move on that margin, but both arms move it. On every outcome the call arm does better than the message arm, and on three of the four the difference is significant at the 1 percent level.

Table 4: Effects on arithmetic

	Arithmetic <i>z</i> -score (1)	Recognizes one-digit numbers (2)	Recognizes two-digit numbers (3)	Can subtract (4)
SMS only	0.144*** (0.024)	0.014*** (0.004)	0.056*** (0.010)	0.049*** (0.010)
SMS + calls	0.239*** (0.024)	0.022*** (0.004)	0.082*** (0.009)	0.086*** (0.012)
Control mean at endline	0.000	0.950	0.510	0.140
<i>p</i> -value, SMS only = SMS + calls	0.000	0.061	0.006	0.001
Baseline outcome and controls	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes
Observations	26,751	26,751	26,751	26,751

Notes: OLS estimates of equation (1). The sample is children in Grades 3–5 who scored below the subtraction level at baseline and were tested at endline. Column (1): ASER arithmetic level standardized on the control group at endline. Columns (2)–(4): indicators for having reached ASER level 1, 2 and 3 at endline. Controls are the baseline value of the outcome, gender, grade, school type and household smartphone ownership; district fixed effects (170 districts). Standard errors clustered by community (1,121 clusters) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Two benchmarks put these numbers in perspective, and both call for care. Across 96 randomized trials with learning outcomes in low- and middle-income countries the median effect is 0.10 SD and the 75th percentile 0.23 SD; among the largest quarter of studies by sample size the median is 0.10 (Evans and Yuan, 2022). On that scale the SMS-only effect is above the median and the SMS+calls effect at the upper quartile, from a sample larger than almost all of the studies in the distribution, so the usual small-sample inflation is not the explanation. Two features of this study are. The first is measurement. The outcome is a five-level ordinal scale, and the sample is restricted to children who started at its three lowest levels, so the control-group standard deviation at endline is small—about 0.8 levels, which is what one obtains by comparing the *z*-score effects in column (1) with the threshold effects in columns (2)–(4)—and a modest share of children crossing one threshold becomes a sizeable effect in standard deviation units. Standardized effects in this literature depend as much on the test and the sample as on the learning (Kraft, 2020; Evans et al., 2026). In raw units, summing the threshold effects in columns (2)–(4) gives an average gain of about 0.12 levels with messages and 0.19 levels with calls: roughly one child in eight, and one in five, moves up a level who would not have otherwise.

The second is alignment and timing. The thresholds measure the skills the activities practised, and the endline came at the close of the program, before any fading; effects measured this way are typically larger than effects on broad tests or at a later follow-up (Kraft, 2020; Angrist et al., 2024).

Read in raw units, the effects are in line with the closest studies rather than above them. In Botswana, SMS with a tutoring call raised the same ASER scale by 0.15 levels and SMS alone by nothing measurable (Angrist et al., 2022); our messages alone produce slightly less than the Botswana tutoring effect and our messages with calls slightly more. In the five-country trials, phone tutorials produced 0.30–0.35 SD on a range of assessments (Angrist et al., 2023); in standard deviation units our call arm is below that, with no tutoring. The one earlier Indian study of a phone-based program during the closure, in three districts of Odisha, found gains of about 5 percentage points on number recognition and arithmetic using a difference-in-differences design (Ojha and Yadav, 2023), the same order as our 5–9 point gains on the two-digit and subtraction thresholds. What is unusual about this study is not the size of the effect per child but that it was produced by the cheapest arm in this literature, at a scale at which effects usually shrink.

5.2 Spillover to reading

The program sent arithmetic activities only, so Table 5 is a test of spillover. Reading z -scores rise by 0.073 SD (SE 0.017) in the SMS-only arm and by 0.145 SD (SE 0.018) with calls; the gap between arms is again significant ($p < 0.001$). The gains show up at every reading level: the share of children who can read letters, words, a paragraph and a story each rise by 2–3 percentage points with messages alone and by 2–5 points with calls. Spillovers of this kind are not unusual when parents are drawn into a child’s learning—programs for preschool parents in Costa Rica and the Philippines improved several domains at once (Hernández-Agramonte et al., 2024; Angrist et al., 2024)—but in those programs the content covered several domains. Here it covered one. We read the spillover as evidence that the program changes how much households engage with the child’s learning in general, and we return to it in Section 5.5.

5.3 Effects by baseline level

Because the messages were matched to the child’s level, a natural question is whether they worked at every level. Table 6 estimates the effects separately for children who started as beginners, as one-digit recognizers and as two-digit recognizers, on the thresholds above each group’s starting point. Effects are positive in all three groups. Among one-digit recognizers, the largest group, the share who can recognize two-digit numbers at endline rises from 18 percent by 10.6 points with messages and by 15.9 points with calls; among two-digit recognizers, the share

Table 5: Spillover to reading

	Reading z-score (1)	Reads letters (2)	Reads words (3)	Reads a paragraph (4)	Reads a story (5)
SMS only	0.073*** (0.017)	0.027*** (0.008)	0.027*** (0.009)	0.017* (0.009)	0.017** (0.007)
SMS + calls	0.145*** (0.018)	0.053*** (0.009)	0.052*** (0.009)	0.044*** (0.011)	0.023*** (0.006)
Control mean at endline	0.000	0.820	0.440	0.220	0.070
p -value, SMS only = SMS + calls	0.000	0.004	0.009	0.010	0.382
Baseline outcome and controls	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	26,751	26,751	26,751	26,751	26,751

Notes: OLS estimates of equation (1). Sample, controls, fixed effects and clustering as in Table 4. Column (1): ASER reading level standardized on the control group at endline. Columns (2)–(5): indicators for having reached ASER reading level 1, 2, 3 and 4 at endline. The program sent no reading content. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

who can subtract rises from 29 percent by 10.4 and 18.0 points. Among beginners the messages raise the share recognizing one-digit numbers by 10.8 points and the calls by 18.2 points, but the group is small (2,270 children) and the two arms cannot be told apart ($p = 0.13$). In the two larger groups the call arm does significantly better, and the call premium is largest among two-digit recognizers, for whom the next step is the subtraction procedure itself. Each group can, by construction, only be observed crossing the thresholds above it, so the table should be read as showing that the program moved children at every starting level rather than as ranking the levels.

Table 6: Effects by baseline arithmetic level

Outcome: recognizes / can	Baseline: beginner			Baseline: one-digit		Baseline: two-digit
	One-digit (1)	Two-digit (2)	Subtract (3)	Two-digit (4)	Subtract (5)	Subtract (6)
SMS only	0.108** (0.044)	0.004 (0.016)	−0.005 (0.004)	0.106*** (0.016)	0.009* (0.005)	0.104*** (0.021)
SMS + calls	0.182*** (0.044)	0.031** (0.015)	0.004 (0.005)	0.159*** (0.015)	0.021*** (0.006)	0.180*** (0.024)
Control mean at endline	0.44	0.05	0.01	0.18	0.03	0.29
p : SMS only = SMS + calls	0.130	0.120	0.177	0.001	0.036	0.002
Observations	2,270	2,270	2,270	13,223	13,223	11,258

Notes: OLS estimates of equation (1) on subsamples defined by the child’s baseline ASER arithmetic level: beginner (level 0), one-digit recognition (level 1) and two-digit recognition (level 2). The three groups together make up the estimation sample. Outcomes are indicators for having reached the stated level at endline. Controls, district fixed effects and village-clustered standard errors as in Table 4. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.4 Effects by household and child characteristics

Appendix Tables A2–A5 interact the two arm indicators with gender, school type, smartphone ownership and grade. Three findings stand out. First, girls and boys gain alike. The interaction of the SMS+calls arm with a female indicator on the arithmetic z -score is -0.030 (SE 0.020), so the data rule out a gap larger than about 0.07 SD in favour of boys or 0.01 SD in favour of girls; the SMS-only interaction is -0.002 (SE 0.021). Second, children in private schools, 13 percent of the sample, gain about as much as children in government schools on the z -score; the interactions are imprecise (SE 0.047–0.049) and the only ones that survive the false discovery rate adjustment are negative effects on the lowest thresholds, which private-school children had largely reached already. The grade interactions show the same pattern: older children gain less on the lower thresholds and more on subtraction, which is where they started.

Third, the program works without a smartphone, and it works somewhat better with one. For households with a basic phone only, the SMS-only and SMS+calls effects on the z -score are 0.115 SD and 0.200 SD. Households with a smartphone gain a further 0.057 SD (SE 0.030) and 0.073 SD (SE 0.032), and on the subtraction outcome a further 4–5 percentage points; the subtraction interactions survive the false discovery rate adjustment ($q = 0.02$), the z -score interactions narrowly do not ($q = 0.12$ and 0.07). A long message is easier to read on a larger screen, and smartphone ownership also marks a better-off household, and we cannot say which of these is at work. The pattern differs from that in Bangladesh, where a phone-based information program raised learning only among wealthier households (Beam et al., 2025); here the poorer half gains most of what the richer half gains. It is consistent with evidence that demand for SMS-based practice is highest in households with the fewest alternatives (Figueiredo Walter et al., 2023).

5.5 What explains the call premium

The call arm delivered the same messages on the same schedule. Table 7 compares the two treatment arms on the two engagement questions asked at endline. Receipt of the messages is 93 percent in the SMS-only arm and 1.7 percentage points higher with calls (SE 1.6), a difference that could be zero. Among households that received the messages, the share reporting that someone helped the child with the activities is 83 percent in the SMS-only arm and 9.7 points higher with calls (SE 2.4). The call did not change what arrived; it changed what happened after it arrived.

Three channels could produce this, and the design does not separate them. The first is salience: a weekly call is a reminder, and reminders alone have been shown to move parents’ behaviour as much as new information does (Lichand et al., 2024). The second is accountability: the

Table 7: Receipt of messages and help at home: SMS + calls relative to SMS only

	Household received the messages (1)	Someone helped the child with the activities (2)
SMS + calls (relative to SMS only)	0.017 (0.016)	0.097*** (0.024)
Mean, SMS-only arm	0.93	0.83
Controls and district fixed effects	Yes	Yes
Observations	17,670	16,545

Notes: OLS estimates on the two treatment arms only. Column (1): indicator that the household reported receiving SMS from Pratham in the past month, asked at the endline visit of the parent where possible and otherwise of the child; all children in the two treatment arms. Column (2): indicator that any household member helped the child with the activities; children in households that reported receiving the messages. Both measures are self-reports collected by the field worker who delivered the program. Controls (gender, grade, school type, smartphone ownership) and district fixed effects as in Table 4. Standard errors clustered by community in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

caller was a person the family knew, who would ask next week whether the activity had been done. Teacher-to-parent messages in the United States raise effort through a similar channel (Kraft and Rogers, 2015). The third is access to help: a caller who learned that a parent could not support the child could point the family to a community volunteer, so some households in the call arm may have obtained instruction that households in the message arm did not. The reading spillover, which doubles with the calls, fits all three, since each raises the household’s engagement with the child’s learning in general rather than with arithmetic in particular. What the data rule out is that the calls worked by delivering more content, and what they establish is the size of the premium: 0.10 SD, about two-thirds of the message effect, from a contact that carried no instruction.

That number sits usefully against earlier work. In Botswana the difference between SMS with a tutoring call and SMS alone was 0.10 SD (Angrist et al., 2022); the calls there taught, ours did not, and the premium is the same. In Sierra Leone and Kenya calls from teachers who did not know the families, with or without tutoring content, produced nothing (Crawford et al., 2023; Schueler and Rodriguez-Segura, 2023). A review of this literature concludes that the phone is a platform rather than a pedagogy, and that parents who are brought into the activity amplify its effect (Aker and Angrist, 2026). Our reading of the evidence is that a call adds value when it connects a usable activity to a person the household will answer to; the instruction may be the smaller part. One more possibility deserves a sentence. Households that expect a program to do the work sometimes pull back their own effort (Das et al., 2013); a weekly question about who helped makes that harder, which may be part of why messages with calls outperform messages alone. The engagement measures are self-reports to the same worker who made the calls, so

the 10-point difference in help should be taken as indicative; the learning effects do not depend on it.

A few caveats apply to these estimates. The program was delivered and evaluated through Pratham’s own field staff, as most evaluations run with an implementing organization are, and this has two implications. The workers who tested children at endline had also delivered the program in their villages, so the endline was not blind. The assessment is a sequence of pass-or-fail tasks with alternate forms at baseline and endline, administered under the same protocol in every arm, which limits the room for this to matter, but it cannot be ruled out. And the caller was a person the family knew, so the call premium is an estimate for a known caller; how much of it would carry over to a caller without that relationship is a question for a different design. Two further limits are the usual ones for a short program. We observe outcomes at the end of four weeks, and gains from parent-focused programs can fade (Angrist et al., 2024); whether these persist, or are built on when children return to class, requires a follow-up we do not have. And because every message was matched to the child’s level, the design cannot separate level-matching from the activity format or the sender; the case for targeting in Section 1 rests on comparison with other programs rather than on an arm of this one. Take-up and help at home are self-reports, and we read them as indicative.

6 What the Program Costs

Pratham did not cost the program, and two facts stand in for a cost study. The first is that the messages were virtually free to send: bulk SMS in India cost a fraction of a rupee per message at the time, so the delivery cost of the SMS-only arm is, to a close approximation, zero. The second is that the calls are the only delivery cost of the second arm, and a back-of-envelope calculation puts them at roughly Rs 35–75, or US\$0.45–0.95, per child over the four weeks.⁸ Even if the salary and time assumptions are off by a factor of five, the calls arm costs under US\$3 per child. For comparison, the Botswana program cost US\$5 per child for SMS and US\$19 with tutoring calls, the five-country phone tutorials about US\$11, the Sierra Leone tutoring calls US\$40, and a two-month information campaign in Bangladesh US\$2.77 per household (Angrist et al., 2022, 2023; Crawford et al., 2023; Beam et al., 2025); cost accounting differs across these studies and the comparison is rough. Put in the units the cost-effectiveness literature uses, US\$100 spent on delivering the calls arm buys learning gains two orders of magnitude larger than the 0.89 SD per US\$100 that placed the Botswana program among the most cost-effective interventions on

⁸Assumptions: 22 messages per child, sent at effectively zero marginal cost; four calls per child of five to eight minutes each, including dialling and unanswered attempts; a field worker’s cost to Pratham taken from Pratham’s salary range for these staff in April 2022; and airtime of about Rs 1 per call. At an exchange rate of Rs 78 per US dollar these give the figures in the text.

record (Angrist et al., 2022, 2025)—a comparison that flatters this program precisely because delivery is not where its cost lies.

These are delivery costs, and delivery is not where the money goes. The program required a one-on-one assessment of every child to place them at a level, which in turn required a household visit by a trained worker; in-person assessments of this kind in India have been costed at several dollars per child (Crawford et al., 2023). It also required content written and translated into eleven languages, phone numbers collected in advance, and the supervision of 1,146 field workers. None of these are in the figures above, and the first of them is the price of targeting. A government that wanted to copy the program would find the messages nearly free and the calls cheap; what it would have to pay for is knowing each child’s level. That is the same investment the in-person Teaching at the Right Level programs require (Banerjee et al., 2017a), and the reviews that rank targeted instruction among the most cost-effective things a school system can do already count it (Angrist et al., 2025). Against that benchmark, a text message is the cheapest way yet tested to act on the information once it exists.

7 Conclusion

Schools in India reopened in early 2022 to children who were, on average, further behind on foundational skills than any cohort in a decade. This paper asked what the cheapest available tool could do for them. The answer, from 1,162 villages in 19 states, is that a month of daily text messages, each asking a caregiver to do one small numeracy activity with the child at the child’s level, raises arithmetic by 0.14 SD, about a tenth of an ASER level, and moves one child in twenty across the subtraction threshold; that a brief weekly call from a known field worker, asking only whether the activity was done, raises the effect to 0.24 SD; and that both arms produce a spillover to reading, which the program never addressed. The messages reached basic phones, required no internet, and were delivered by the organization’s own staff under a single protocol.

The findings change the picture in a few places. In the literature on phones and learning, the cheap end of the ladder—the one-way text—has looked like a dead end; here it is not, and the difference from the programs that failed lies in what the text asks for, whom it is matched to and whom it comes from. In the literature on parents as a channel, remote nudges typically move test scores by a twentieth of a standard deviation; here caregivers with little schooling, asked to act rather than to know, produced several times that. In the literature on tutoring by phone, the call has been bundled with instruction; here a call with no instruction is worth two-thirds of the content it accompanied, which suggests that a good part of what a tutor adds is presence rather than teaching. And in the debate on whether results survive scale, this one

was produced at scale.

The findings also have limits that the next study should address. An untargeted arm would identify what level-matching is worth. A follow-up round would say whether a four-week gain lasts or is overtaken. An endline collected by assessors who did not deliver the program would tighten the measurement. And a call arm run by callers the family does not know would say how much of the premium travels. None of these is expensive relative to the program, and all would raise its value to policy.

For policy the arithmetic is simple. The messages are virtually free to send, the calls cost well under a dollar a child, and what the program really requires is what the Teaching at the Right Level approach has always required: a reliable reading of each child’s level. India’s national foundational-learning mission is building exactly that capacity, and Pratham’s field network already has it ([Ministry of Education, Government of India, 2021](#); [Banerjee et al., 2017b](#)). Where the level is known, a text message is the cheapest way yet tested to act on it, and a brief call from someone the family knows makes it work better. Neither replaces a classroom. But classrooms close, for strikes, disasters and epidemics as well as pandemics, and the losses from those closures last ([Jaume and Willén, 2019](#); [Andrabi et al., 2023](#)). Tools that keep learning moving through a disruption are what resilience means for an education system ([Angrist et al., 2023](#)). When classrooms are closed, or when children sit in them years behind, both are worth having.

References

- Agostinelli, F., Doepke, M., Sorrenti, G., and Zilibotti, F. (2022). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of Public Economics*, 206:104574.
- Aker, J. C. and Angrist, N. (2026). Leveraging mobile phones for learning. J-PAL Policy Insight, Abdul Latif Jameel Poverty Action Lab, <https://www.povertyactionlab.org/policy-insight/leveraging-mobile-phones-learning>.
- Aker, J. C. and Ksoll, C. (2019). Call me educated: Evidence from a mobile phone experiment in Niger. *Economics of Education Review*, 72:239–257.
- Aker, J. C. and Ksoll, C. (2020). Can ABC lead to sustained 123? the medium-term effects of a technology-enhanced adult education program. *Economic Development and Cultural Change*, 68(3):1081–1102.
- Aker, J. C., Ksoll, C., and Lybbert, T. J. (2012). Can mobile phones improve learning? evidence from a field experiment in Niger. *American Economic Journal: Applied Economics*, 4(4):94–120.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.
- Andrabi, T., Daniels, B., and Das, J. (2023). Human capital accumulation and disasters: Evidence from the Pakistan earthquake of 2005. *Journal of Human Resources*, 58(4):1057–1096.
- Angrist, N., Ainomugisha, M., Bathena, S. P., Bergman, P., Crossley, C., Cullen, C., Letsomo, T., Matsheng, M., Panti, R. M., Sabarwal, S., and Sullivan, T. (2023). Building resilient education systems: Evidence from large-scale randomized trials in five countries. Working Paper 31208, National Bureau of Economic Research. Revised June 2025.
- Angrist, N., Bergman, P., and Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour*, 6(7):941–950.
- Angrist, N., Evans, D. K., Filmer, D., Glennerster, R., Rogers, H., and Sabarwal, S. (2025). How to improve education outcomes most efficiently? a review of the evidence using a unified metric. *Journal of Development Economics*, 172:103382.
- Angrist, N., Kabay, S., Karlan, D., Lau, L., and Wong, K. (2024). Human capital at home: Evidence from a randomized evaluation in the Philippines. EdWorkingPaper 24-1044, Annenberg Institute at Brown University.

- ASER Centre (2021). Annual status of education report (rural) 2021.
- ASER Centre (2023). Annual status of education report (rural) 2022.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. (2017a). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4):73–102.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., and Walton, M. (2017b). Mainstreaming an effective intervention: Evidence from randomized evaluations of “Teaching at the Right Level” in India. *Science*, 357(6346):1146–1148.
- Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3):1235–1264.
- Banerji, R., Berry, J., and Shotland, M. (2017). The impact of maternal literacy and participation programs: Evidence from a randomized evaluation in India. *American Economic Journal: Applied Economics*, 9(4):303–337.
- Beam, E. A., Mukherjee, P., and Navarro-Sola, L. (2025). Lowering barriers to remote education: Experimental impacts on parental responses and learning. Working paper, University of Vermont, University of Wisconsin–Madison, and IIES Stockholm. October 2025.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.
- Bergman, P. (2021). Parent-child information frictions and human capital investment: Evidence from a field experiment. *Journal of Political Economy*, 129(1):286–322.
- Bergman, P. and Chowanajin, N. (2026). Characterizing the file drawer: Evidence from a meta-analysis of parent-interventions around the world. Working Paper 35299, National Bureau of Economic Research.
- Berlinski, S., Busso, M., Dinkelman, T., and Martínez A., C. (2025). Reducing parent–school information gaps and improving education outcomes: Evidence from high-frequency text messages. *Journal of Human Resources*, 60(4):1284–1322.
- Bethhäuser, B. A., Bach-Mortensen, A. M., and Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, 7(3):375–385.
- Bold, T., Kimenyi, M., Mwabu, G., Ng’ang’a, A., and Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, 168:1–20.

- Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.
- Carlana, M. and La Ferrara, E. (2025). Apart but connected: Online tutoring, cognitive outcomes, and soft skills. *American Economic Review*, 115(10):3487–3513.
- Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives*, 20(1):91–116.
- Crawford, L., Evans, D. K., Hares, S., and Sandefur, J. (2023). Live tutoring calls did not improve learning during the COVID-19 pandemic in Sierra Leone. *Journal of Development Economics*, 164:103114.
- Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, 5(2):29–57.
- Dizon-Ross, R. (2019). Parents’ beliefs about their children’s academic ability: Implications for educational investments. *American Economic Review*, 109(8):2728–2765.
- Doss, C., Fahle, E. M., Loeb, S., and York, B. N. (2019). More than just a nudge: Supporting kindergarten parents with differentiated and personalized text messages. *Journal of Human Resources*, 54(3):567–603.
- Duflo, A., Kiessel, J., and Lucas, A. M. (2024). Experimental evidence on four policies to increase learning at scale. *Economic Journal*, 134(661):1985–2008.
- Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–1774.
- Engzell, P., Frey, A., and Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17):e2022376118.
- Escueta, M., Nickow, A. J., Oreopoulos, P., and Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, 58(4):897–996.
- Evans, D. K., Rossiter, J., Hares, S., and Henny, C. (2026). The illusion of comparability among standardised effect sizes: Why education evaluations should report raw effects. Working Paper 748, Center for Global Development.

- Evans, D. K. and Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, 44(3):532–540.
- Figueiredo Walter, T., Gray-Lobe, G., and Kabay, S. (2023). Can brick phones bridge the digital learning divide? evidence from SMS-based math practice. EdWorkingPaper 23-791, Annenberg Institute at Brown University.
- Hassan, H., Islam, A., Siddique, A., and Wang, L. C. (2024). Telementoring and homeschooling during school closures: A randomised experiment in rural Bangladesh. *Economic Journal*, 134(662):2418–2438.
- Hernández-Agramonte, J. M., Namen, O., Näslund-Hadley, E., and Biehl, M. L. (2024). Supporting early childhood development remotely: Experimental evidence from SMS messages. *Journal of Development Economics*, 166:103201.
- Jaume, D. and Willén, A. (2019). The long-run effects of teacher strikes: Evidence from Argentina. *Journal of Labor Economics*, 37(4):1097–1139.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4):241–253.
- Kraft, M. A. and Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review*, 47:49–63.
- Kumar, D., Sunder, N., Sabates Aysa, R., and Wadhwa, W. (2024). Improving children’s foundational learning through community-school participation: Experimental evidence from rural India. *Labour Economics*, 91:102615.
- Lichand, G., Cunha, N., Madeira, R. A., and Bettinger, E. (2024). When the effects of informational interventions are driven by salience: Evidence from school parents in Brazil. Working paper, Stanford University.
- Lichand, G. and Doria, C. A. (2024). The lasting impacts of remote learning in the absence of remedial policies: Evidence from Brazil. *Proceedings of the National Academy of Sciences*, 121(22):e2316300121.
- Lichand, G., Doria, C. A., Leal-Neto, O., and Fernandes, J. P. C. (2022). The impacts of remote learning in secondary education during the pandemic in Brazil. *Nature Human Behaviour*, 6(8):1079–1086.
- McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2):210–221.

- Ministry of Education, Government of India (2021). NIPUN Bharat: National initiative for proficiency in reading with understanding and numeracy.
- Ministry of Education, Government of India (2022). Unified district information system for education plus (UDISE+) 2021–22: Flash statistics.
- Moscoviz, L. and Evans, D. K. (2022). Learning loss and student dropouts during the COVID-19 pandemic: A review of the evidence two years after schools shut down. Working Paper 609, Center for Global Development.
- Muralidharan, K. and Niehaus, P. (2017). Experimentation at scale. *Journal of Economic Perspectives*, 31(4):103–124.
- Muralidharan, K., Singh, A., and Ganimian, A. J. (2019). Disrupting education? experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4):1426–1460.
- Ojha, M. and Yadav, K. (2023). In good times and bad: Low-cost mobile teaching during a pandemic. *Economics of Education Review*, 96:102439.
- Rodriguez-Segura, D. (2022). EdTech in developing countries: A review of the evidence. *World Bank Research Observer*, 37(2):171–203.
- Schueler, B. E. and Rodriguez-Segura, D. (2023). A cautionary tale of tutoring hard-to-reach students in Kenya. *Journal of Research on Educational Effectiveness*, 16(3):442–472.
- Singh, A., Romero, M., and Muralidharan, K. (2024). COVID-19 learning loss and recovery: Panel data evidence from India. *Journal of Human Resources*. Published online, April 2024.
- Telecom Regulatory Authority of India (2022). Telecom subscription data as on 31st march 2022.
- UNESCO (2022). Global monitoring of school closures caused by the COVID-19 pandemic. UNESCO Institute for Statistics, <https://covid19.uis.unesco.org/global-monitoring-school-closures-covid19/>.
- Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association*, 18(6):3045–3089.
- Wang, L. C., Vlassopoulos, M., Islam, A., and Hassan, H. (2024). Delivering remote learning using a low-tech solution: Evidence from a randomized controlled trial in Bangladesh. *Journal of Political Economy Microeconomics*, 2(3):562–601.
- York, B. N., Loeb, S., and Doss, C. (2019). One step at a time: The effects of an early

literacy text-messaging program for parents of preschoolers. *Journal of Human Resources*, 54(3):537–566.

A Additional Tables

Table A1: Attrition between baseline and endline

	Dependent variable: not tested at endline	
	(1)	(2)
SMS only	0.004 (0.021)	0.011 (0.023)
SMS + calls	0.000 (0.019)	0.001 (0.019)
Control mean	0.12	0.12
p -value, SMS only = SMS + calls	0.850	0.877
Baseline controls	No	Yes
District fixed effects	Yes	Yes
Observations	30,955	30,955

Notes: OLS estimates on the study sample. The dependent variable equals one if the child was in the study sample at baseline and was not tested at endline. Column (2) adds baseline arithmetic and reading levels, gender, grade, school type and household smartphone ownership. Standard errors clustered by community in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Effects on arithmetic, by gender

	Arithmetic z -score (1)	Recognizes one-digit (2)	Recognizes two-digit (3)	Can subtract (4)
SMS only	0.145*** (0.026)	0.016*** (0.005)	0.057*** (0.011)	0.045*** (0.012)
SMS + calls	0.254*** (0.027)	0.024*** (0.005)	0.084*** (0.011)	0.098*** (0.014)
Female	0.016 (0.012)	0.002 (0.004)	−0.005 (0.007)	0.016** (0.006)
SMS only $\times$ Female	−0.002 (0.021) [0.616]	−0.005 (0.005) [0.321]	−0.002 (0.010) [0.616]	0.007 (0.010) [0.413]
SMS + calls $\times$ Female	−0.030 (0.020) [0.184]	−0.003 (0.005) [0.441]	−0.003 (0.010) [0.555]	−0.023** (0.011) [0.090]
Controls and district fixed effects	Yes	Yes	Yes	Yes
Observations	26,751	26,751	26,751	26,751

Notes: OLS estimates of equation (1) with the arm indicators interacted with a female indicator. The arm coefficients give the effects for boys. Standard errors clustered by community in parentheses. Sharpened false discovery rate q -values (Anderson, 2008), computed over the 40 interaction coefficients in Tables A2–A5, in square brackets. Controls and fixed effects as in Table 4. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Effects on arithmetic, by school type

	Arithmetic z-score (1)	Recognizes one-digit (2)	Recognizes two-digit (3)	Can subtract (4)
SMS only	0.146*** (0.024)	0.016*** (0.004)	0.062*** (0.010)	0.044*** (0.010)
SMS + calls	0.232*** (0.024)	0.025*** (0.005)	0.085*** (0.009)	0.078*** (0.011)
Private school	0.130*** (0.031)	0.017*** (0.005)	0.092*** (0.013)	−0.001 (0.016)
SMS only × Private	−0.025 (0.047) [0.445]	−0.020*** (0.007) [0.036]	−0.050** (0.019) [0.046]	0.030 (0.023) [0.231]
SMS + calls × Private	0.035 (0.049) [0.413]	−0.019*** (0.007) [0.040]	−0.024 (0.021) [0.266]	0.054** (0.026) [0.103]
Controls and district fixed effects	Yes	Yes	Yes	Yes
Observations	26,751	26,751	26,751	26,751

Notes: OLS estimates of equation (1) with the arm indicators interacted with an indicator for attending a private school (86–87 percent of children attend government schools). The arm coefficients give the effects for government-school children. Standard errors clustered by community in parentheses; sharpened q -values in square brackets (see Table A2). Controls and fixed effects as in Table 4. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Effects on arithmetic, by household smartphone ownership

	Arithmetic z-score (1)	Recognizes one-digit (2)	Recognizes two-digit (3)	Can subtract (4)
SMS only	0.115*** (0.025)	0.018*** (0.007)	0.056*** (0.012)	0.027*** (0.010)
SMS + calls	0.200*** (0.024)	0.031*** (0.007)	0.078*** (0.011)	0.059*** (0.010)
Smartphone	0.004 (0.019)	0.013*** (0.005)	0.016** (0.008)	−0.018** (0.009)
SMS only × Smartphone	0.057* (0.030) [0.123]	−0.008 (0.006) [0.266]	−0.001 (0.012) [0.616]	0.041*** (0.013) [0.023]
SMS + calls × Smartphone	0.073** (0.032) [0.070]	−0.016** (0.007) [0.056]	0.007 (0.013) [0.441]	0.053*** (0.016) [0.023]
Controls and district fixed effects	Yes	Yes	Yes	Yes
Observations	26,751	26,751	26,751	26,751

Notes: OLS estimates of equation (1) with the arm indicators interacted with an indicator for household smartphone ownership at baseline (52–54 percent of households). The arm coefficients give the effects for households with a basic phone only. Standard errors clustered by community in parentheses; sharpened q -values in square brackets (see Table A2). Controls and fixed effects as in Table 4. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Effects on arithmetic, by grade

	Arithmetic z-score (1)	Recognizes one-digit (2)	Recognizes two-digit (3)	Can subtract (4)
SMS only	0.157*** (0.028)	0.020*** (0.007)	0.076*** (0.013)	0.036*** (0.012)
SMS + calls	0.244*** (0.030)	0.032*** (0.007)	0.096*** (0.012)	0.066*** (0.014)
Grade 4	0.083*** (0.018)	0.011** (0.005)	0.036*** (0.008)	0.026*** (0.009)
Grade 5	0.160*** (0.023)	0.021*** (0.005)	0.069*** (0.009)	0.041*** (0.012)
SMS only $\times$ Grade 4	-0.014 (0.024) [0.441]	-0.006 (0.007) [0.323]	-0.027** (0.013) [0.090]	0.019 (0.012) [0.172]
SMS only $\times$ Grade 5	-0.031 (0.030) [0.306]	-0.014* (0.007) [0.123]	-0.041*** (0.015) [0.040]	0.024 (0.016) [0.172]
SMS + calls $\times$ Grade 4	-0.016 (0.026) [0.441]	-0.013* (0.007) [0.130]	-0.020 (0.013) [0.172]	0.024* (0.013) [0.130]
SMS + calls $\times$ Grade 5	0.002 (0.032) [0.616]	-0.019*** (0.007) [0.046]	-0.026* (0.015) [0.130]	0.048*** (0.017) [0.039]
Controls and district fixed effects	Yes	Yes	Yes	Yes
Observations	26,751	26,751	26,751	26,751

Notes: OLS estimates of equation (1) with the arm indicators interacted with grade indicators (Grade 3 is the reference; 40–42 percent of the sample is in Grade 3, 33–34 percent in Grade 4 and 25–26 percent in Grade 5). The arm coefficients give the effects for Grade 3 children. Standard errors clustered by community in parentheses; sharpened q -values in square brackets (see Table A2). Controls and fixed effects as in Table 4. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: The closest experiments on phone-based learning support in low- and middle-income countries

Study	Setting and period	Channel	Content	Matched to child's level	Human contact	Sample	Effect on learning	Delivery cost per child
This paper	India, 19 states; March–April 2022, as schools reopened	SMS to any phone	Daily numeracy activity for caregiver and child	Yes (two sets)	None; or a brief weekly check-in call by a known NGO worker, no teaching	26,751 children, 1,162 villages	0.14 SD (SMS); 0.24 SD (SMS + calls); spillover to reading 0.07 / 0.15 SD	Virtually zero (SMS); about US\$0.45–0.95 (SMS + calls), assumed
Angrist et al. (2022)	Botswana; 2020 closure	SMS; SMS + phone	Practice problems; 20-minute tutoring call	Cross-randomized subgroup	NGO tutors	4,550 households	0.02 SD (SMS, n.s.); 0.12 SD (SMS + phone)	US\$5; US\$19
Angrist et al. (2023)	India, Kenya, Nepal, Philippines, Uganda; 2020–22	SMS + phone tutoring	Targeted tutoring	Yes	Government teachers or NGO staff	Five trials	0.30–0.35 SD (phone); SMS mixed	About US\$11
Crawford et al. (2023)	Sierra Leone; 2020 closure	Radio + tutoring calls	Instruction by teachers	No	Public and private school teachers	4,399 students	0.006 SD (null)	US\$2 (SMS); US\$40 (calls)
Schueler and Rodriguez-Segura (2023)	Kenya; 2020 closure	Phone calls	5-minute accountability check, or 15-minute tutoring	No	Teachers	105 schools	Null; tutoring negative for children who returned to school	—
Hassan et al. (2024)	Bangladesh; 2020 closure	Phone tele-mentoring	Instruction to mother and child	—	Volunteer mentors	200 villages	Test score 35 percent higher; persisted one year	US\$20
Wang et al. (2024)	Bangladesh; 2021 closure	Interactive voice response	Audio lessons	No	Automated	1,763 children, 90 villages	0.60 SD	—
Beam et al. (2025)	Bangladesh; 2021 closure	Phone messages	Information about a free learning platform	No	None	7,576 households	0.15 SD math (information arm), concentrated among wealthier households	US\$2.77 per household
Hernández-Agramonte et al. (2024)	Costa Rica; 2020 closure, preschool	SMS to parents	Activities, 15 weeks	No	None	1,072 parents	0.11–0.12 SD	About US\$0.8
Ojha and Yadav (2023)	Odisha, India; 2020–21 closure	Calls + SMS + voice messages	Activity-based learning	No	NGO callers	Three districts; difference-in-differences	+4.7 pp arithmetic; +5.5 pp language	—

Notes: Effects are intent-to-treat estimates on the primary learning outcome reported by each study, in standard deviations unless stated. Delivery costs are as reported by each study and are not on a common accounting basis; the figures for this paper are the assumption-based estimates of Section 6. “—” indicates not reported or not applicable.