Delivering Remote Learning Using a Low-tech Solution: Evidence from a Randomized Controlled Trial in Bangladesh

Liang Choon Wang

Monash University

Michael Vlassopoulos

University of Southampton and IZA Institute of Labor Economics

Asad Islam

Monash University, Centre for Development Economics and Sustainability (CDES) and J-PAL

Hashibul Hassan

Jagannath University, Dhaka

Abstract

The Covid-19 pandemic caused prolonged school closures worldwide, disproportionately affecting children in resource-poor settings. To address this, we implemented an educational intervention delivering audio lessons via mobile phones, employing Interactive Voice Response (IVR) technology. Our randomized intervention involved 1,763 primary school children across 90 villages in Bangladesh during the 2021 Covid-19 school closures. The intervention improved literacy and numeracy test scores by 0.60 Standard Deviations (SD), particularly benefitting academically weaker students and those with less-educated caregivers. Our results highlight the potential of this scalable and low-cost intervention to address learning deficits among marginalized students in similar contexts.

Keywords: School closures, Remote education, Interactive Voice Response (IVR), Covid-19, Randomized Controlled Trial, Bangladesh.

Authors' names appear in reverse alphabetical order.

We thank the editor and two anonymous reviewers for their insightful comments. We thank Buly Cardak, Gaurav Datt, Paul Glewwe, Claudio Labanca, Albert Park, Abu Siddique, Agne Suziedelyte, Rebecca Thornton, David Evans, seminar and conference participants at Asian Development Bank (ADB), Curtin University, Monash University, Bangladesh Institute of Development Studies, BRAC Institute of Governance and Development (BIGD), Lahore School of Economics, and the CDES Sustainable Development Conference 2023. This project is registered at the AEA RCT registry: AEARCTR-0007931; ethical clearance for this project was received from Monash University, Australia (Project no. 27474). This project is funded by an EdTech Hub Covid-19 research grant. Code replicating the tables and figures in this article can be found in the Harvard Dataverse, https://doi.org/10.7910/DVN/TCZ1QR. The replication package contains a copy of the programs (Stata do-file) used to create the final results and a readme file with further information. This paper was edited by Anna Dreber Almenberg.

1 Introduction

The Covid-19 pandemic affected the lives of billions of people around the world in numerous ways, with disruptions in education being a key domain. More than 1.5 billion students of all ages across 180 countries have been impacted by the closures of educational institutions (UNESCO 2021). Mounting evidence indicates that school closures have led to large learning losses worldwide, especially for children from disadvantaged backgrounds and those living in low and middle-income countries (Agostinelli et al. 2022; Engzell, Frey, and Verhagen 2021; Moscoviz and Evans 2022; Patrinos 2022; Singh, Romero, and Muralidharan 2022). These children are more vulnerable to disruptions in formal, in-person education, as they have limited access to distance learning resources and may lack adequate parental support for their learning (UNICEF 2020; Rahman and Sharma 2021). Therefore, there are growing concerns that school closures will exacerbate pre-existing education inequalities. These concerns highlight the need for low-cost and effective remote learning solutions that can be mobilized when schools are forced to close due to public health emergencies or when other causes, such as natural disasters, wars, strikes, and political unrests, trigger educational disruptions.

Bangladesh provides a good setting for studying remote learning interventions. The country experienced one of the world's longest periods of school closures during the Covid-19 pandemic, with around 37 million children having had their learning disrupted (UNESCO 2021). A rapid survey conducted by the World Bank found that only around 40% of students had access to remote learning in the first few months of the Covid-19 pandemic (Biswas et al. 2020). Even after a year of school closures, more than 40% of students still did not have access to remote learning (Rahman et al. 2021). Only a small percentage (5%) of children aged 5-15 years had access to a computer, and the active internet usage rate (28.8%) was also low due to the lack of compatible devices and high data costs (DataReportal 2021; Rahman et al. 2021). More than half of school-aged children did not have access to TV, and even those who did

often could not benefit from the TV lessons that were available during school closures. This situation appears particularly concerning for students in rural and disadvantaged areas (Hassan et al. 2021; Beam et al. 2021).

This paper reports evidence from a feature phone-based remote learning intervention aimed at addressing the learning needs of children during the Covid-19 pandemic in a resource-constrained context. Our educational program was delivered during school closures in Bangladesh. The education program in question involved the delivery of pre-recorded audio lessons using Interactive Voice Response (IVR) technology.[1] A key advantage of this technology is that it provides flexibility regarding learning levels and the timing of learning: lessons of different proficiency levels can be stored in a telecom server, and learners can choose lessons at their competence level, and when to access them without having to follow a pre-arranged schedule. More than 96% of households in rural Bangladesh have access to a mobile phone, but less than one-third of them own smart phones (Hassan et al. 2021), prompting us to offer a resource that could be accessed using a simple feature phone. In designing the audio lessons, we employed the distance learning method of Interactive Audio Instruction (IAI), in which learning content is delivered through pre-recorded audio broadcasts and learners engage actively through questions and exercises (Bosch 1997). This method was originally conceived to deliver lessons through radio, and has been shown to be effective in improving learning outcomes in conventional classrooms (Anzalone and Bosch 2005; Ho and Thukral 2009).

The educational program covered two main areas – literacy and numeracy – and was divided into 60 audio lessons. The lessons were delivered over a 15-week period, with each lesson lasting between 16 and 18 minutes. The audio lessons were structured as pre-recorded

[1] Interactive Voice Response (IVR) is an automated phone system that allows humans to access information via a touch tone system or a voice response system.

conversations among four characters, two teachers and two students, following the IAI methodology. The caregivers could select and access any lesson at any time for their child during the program period, as we did not impose that any specific sequence be followed.

Beyond numeracy and literacy, we were also interested in investigating whether children's leadership skills could be improved through this phone-based remote learning method. Generally, leadership is considered a complex and multidimensional advanced competency rather than a fixed, genetic personality trait (Karagianni and Jude Montgomery 2018). It is perceived as a dynamic skill that can be developed through appropriate interventions (Sisk 1993), particularly in childhood, when skills and personality traits tend to be more malleable (Murphy and Johnson 2011; Billsberry, Vega, and Molineux 2019).[2] Despite its potential importance, there has been limited research evaluating interventions aimed at enhancing leadership skills among children through randomized trials (Gutman and Schoon 2013). To address this gap, we offered an additional leadership skill module in 15 audio lessons, adapted from the Lead Africa program (LEAD 2021), to a separate treatment group of students.[3] As the contents of the leadership module covered a broad range of leadership skills, such as communication, planning, patience, empathy, sympathy, compassion, and perseverance, and encouraged extended interaction between caregivers and children at home, it has the potential to offer children a wide range of noncognitive skills that students usually acquire in school.

Furthermore, as schools were closed and social interactions with other children were limited, in addition to numeracy, literacy and leadership skills, we also examined whether the intervention had any effects on children's noncognitive skills and behavioral difficulties. For

[2] Some studies have shown that leadership development activities in schools are associated with measures of children's leadership trait (Ayman-Nolley and Ayman 2005; Billsberry, Vega, and Molineux 2019; Salmond and Fleshman 2010).
[3] The developer of this leadership module had previously delivered its leadership course to children in Liberia and Morocco, using a combination of in-person delivery and low-cost mobile technology. However, the effectiveness of this approach had not been previously evaluated.

example, the lockdown and associated limited interactions with other children could exacerbate children's behavioral problems, such as tantrums, nervousness, lack of control over emotions, and hyperactivity. The various modules that were offered could help children improve their noncognitive skills and decrease their behavioral difficulties.

We implemented this education program in a three-arm clustered Randomized Controlled Trial (RCT) targeting over 1700 primary school children across 90 villages (30 villages in each arm). In the *Standard* group, we offered participating children the literacy and numeracy module; in the *Extended* group, we offered the leadership module in addition to the literacy and numeracy module; and in the Control group, no intervention was offered. One challenge faced by remote learning educational programs is that participants might not engage with the material due to financial or time constraints or a lack of interest. To assess this, we monitored the take-up rate and usage of the learning material throughout the intervention. The data show that participants engaged substantially with the lessons, as more than 70% of participating children completed at least two-thirds of the lessons.

We find that the phone-based educational program led to substantial improvements in the learning outcomes of children, as measured by assessment tests on literacy and numeracy that we administered in the endline. Relative to children in the Control group, treated children in the Standard and Extended groups experienced a 0.60 standard deviation (SD) and 0.63 SD improvement, respectively, on the total test score. This is roughly equivalent to a 30% increase in their total test scores. Importantly, we find that the intervention was especially beneficial for academically weaker students and students with less-educated caregivers. As students with less-educated caregivers and weaker academic preparation are more likely to follow the recommended sequence of lessons that progressively built up the difficulties of learning materials, differences in how caregivers accessed the lessons potentially explain the heterogenous treatment effects. These findings suggest that the intervention combined with a

recommended lesson sequence is particularly effective for more vulnerable groups of students and could contribute toward reducing educational inequalities.

We do not find evidence that the intervention improved the leadership, communication, and planning skills of children, which was the aim of the leadership module. Furthermore, we do not find significant impacts on a range of noncognitive skills, such as impulsivity, grit, growth mindset, and empathy. With regard to behavioral issues, children in the Standard group experienced improvements in several dimensions—showing less emotional symptoms, conduct problems, hyperactivity, peer-related problems, and a more prosocial attitude—however, these treatment effects are not robust to multiple hypotheses testing corrections. These findings suggest that fostering leadership and noncognitive skills might require larger investments than what this type of short remote learning programs can provide.

Overall, our findings suggest that the phone-based educational program can be an effective and scalable distance learning tool to improve the learning outcomes of students, particularly in contexts with no access to formal, in-person education and limited access to alternative content through digital technological devices. An important aspect of the intervention is its wide accessibility, as audio lessons can be readily accessed via basic feature phones, which are widely available even among underprivileged families. A second feature is scalability. The total cost of this 15-week intervention amounted to USD 27.5 per student, of which USD 13.2 were variable costs and USD 14.3 were fixed costs. Scaling up the program is likely to reduce the per-student cost, further emphasizing its potential as a cost-effective solution to educational challenges in similar settings.

Our study contributes to a recent literature focused on exploring innovative solutions to address educational disruptions, such as hiring paid instructors or volunteer tutors or SMS campaigns to help students with their learning over the phone (Angrist, Bergman, and Matsheng 2022; Crawfurd et al. 2023; Hassan et al. 2021; Lichand and Christen 2021; Schueler and Rodriguez-

Segura 2023).[4] One challenge facing these programs is that it can be difficult to scale them up, especially in a low-income country context, as paid instructors are costly and volunteers are difficult to retain (Islam et al. 2022). Importantly, our program distinguishes itself by offering flexible study hours, a feature absents in other phone-based educational initiatives. The flexible delivery method of our program accommodates the resource constraints rural households typically face, such as having only one phone in each household, and issues like unstable mobile networks and unreliable electricity supply.

More broadly, our study also contributes to a growing body of literature examining the role of technology in enhancing educational productivity, particularly in low-income countries. Previous work investigated technology's role both as a substitute and a facilitator of standard classroom teaching (Beg et al. 2022; Bianchi, Lu, and Song 2022; Cardim, Molina-Millán, and Vicente 2023; Johnston and Ksoll 2022; Muralidharan, Singh, and Ganimian 2019). The main contribution of the current paper is to provide evidence of the effectiveness of a remote learning intervention delivered during school closures. This approach could be a viable alternative to more human resource-intensive programs, as it is scalable and offers some flexibility regarding learning levels and learning delivery schedules to learners and their caregivers.

2 Intervention and Research Design

2.1 Background

When all schools were closed due to the Covid-19 pandemic in March 2020, the government of Bangladesh responded to the ensuing educational crisis by providing multimodal distance

[4] Research on feature-phone-based educational interventions is not a new area of interest. In the last decade, excluding the Covid-19 pandemic period, various studies have demonstrated the effectiveness of this platform. However, these studies have mostly been restricted to teacher–caregiver rather than teacher-student engagement (Bergman and Chan 2021; Berlinski et al. 2016; Hurwitz et al. 2015; Kraft and Dougherty 2013; Mayer et al. 2015), or high school graduates (Bird et al. 2021; Castleman and Meyer 2020), or adult learners (Aker and Ksoll 2019; Ksoll et al. 2015), rather than primary graders.

learning. By the first week of April 2020, the Ministry of Primary and Mass Education (MoPME) and the Ministry of Education (MoE) started remote learning through asynchronous classes broadcast via national television and online platforms (Rahman and Sharma 2021). Online resources were already developed and available on various sites but were expanded during the school closures. Radio broadcasting was introduced later in the year.

Despite the quick delivery of multimodal distance learning, a significant portion of students faced barriers to accessing government remote learning classes mainly due to a lack of TVs and radios. A rapid survey conducted by the World Bank found that only around 40% of students had access to remote learning in the first few months of the Covid-19 pandemic (Biswas et al. 2020). Even after a year, a significant portion of children remained outside of distance learning coverage; 44% and 36% of rural households and urban-slum households did not have access, respectively (Rahman et al. 2021). This lack of access led to poor learning outcomes, as only 18% of primary graders and 38% of secondary graders were actively learning through assignments as of August 2021 (Rahman et al. 2021).[5]

As the number of Covid-19 cases decreased in the second half of 2021, the government ordered all schools to reopen, allowing students to attend classes for one to two days per week starting on September 12, 2021. This partial reopening took place after 10 weeks of our intervention. However, schools were closed again on January 21, 2022, during the peak of the Omicron wave, and then fully reopened in mid-March 2022.

2.2 The intervention

The aim of the intervention was to deliver interactive audio content via IVR to improve the learning of primary-school students. We next explain the main features of the intervention.

[5] The MoPME and MoE gave various assignments via distance education programs and online platforms to engage students in learning and to assess their progress.

2.2.1 Interactive Voice Response (IVR)

Interactive Voice Response (IVR) is an automated phone system technology that allows incoming or receiving callers to access information by traversing or navigating a pre-designed flow. Navigation to different points of the flow can be done by either voice commands or keypad selections made by the caller. Once the caller makes a call to or receives a call from an IVR-enabled number, there is no need for human intervention. A pre-recorded message can guide them to the desired landing node with their preferred information. Though both radio and IVR platforms only support audio lessons, IVR has two important advantages over the radio: listeners can select lessons, i.e., there is no fixed broadcasting sequence, and can engage with the content at times that are convenient for them.

2.2.2 Interactive Audio Instruction (IAI)

The original design of Interactive Audio Instruction (IAI) was created back in 1970 as Interactive Radio Instructions (IRI) to teach mathematics via radio in Nicaragua (Bosch 1997). IAI is an instructional approach that turns a one-way technology into a tool for active learning as it requires learners to stop and react to questions and exercises through verbal response, to engage in group work, and physical and intellectual activities while the program is on the air (Bosch, Rhodes, and Kariuki 2002). Facilitators play an important role in supervising the progression of the lessons (Ho and Thukral 2009).

Just before the intervention, the field staff of the Global Development and Research Initiative (GDRI), our local partner, visited the sample households to provide a guidebook and briefing on the IAI method. In particular, they explained how the interaction would take place, what would be the role of caregivers, and what the caregivers would need during the lessons. Caregivers were the facilitators in our intervention. For example, they were invited to draw a few figures, show some letters, or write numbers to engage with their learners during the IAI

sessions played over the IVR. In Section A-1 in Appendix A we provide further examples of such activities.

The caregivers were present to ensure that the children followed the audio instructions and completed the learning activities. Thus, they were facilitators as well as supervisors or monitors. As the caregivers were always present during lessons, it is not possible for us to disentangle the effect of parental presence or supervision from the effect of the activities themselves. The guidebook serves an important role in our intervention as it prescribes a specific sequence of audio lessons for the caregivers to follow each week. Although all audio lessons were made available on day one, the caregivers were advised to follow the particular sequence outlined in the guidebook. This approach ensured that the learners started with the relatively easier materials before progressing to the more difficult ones.

2.2.3 Content of intervention

The 15-week intervention included three elementary educational modules – literacy, numeracy, and leadership – divided into 75 audio lessons, with each lesson lasting between 16 and 18 minutes (Section A-2 in Appendix A provides more details). Caregivers accessed these pre-recorded audio lessons via Interactive Voice Response (IVR) by dialing a toll-free number. Although the guidebook outlined a specific sequence of lessons to follow, caregivers could choose and access any lesson at any time for the child participants, without having to adhere to a specific curriculum order. Figure B1 and B2 in Appendix B provide an illustration of the IVR journey experienced by the participating caregiver–child pairs.

The audio lessons featured pre-recorded conversations among four characters: two teachers and two students. During the conversations, students were asked to do some activities according to the teachers' instructions, such as clapping, standing up, counting, and making plans. The teachers used regular pauses and cues, as well as playing music and songs during

the recorded lessons to assist the children in completing similar tasks with the help of their caregivers. The caregivers were mothers in 78% of cases, while in the remaining 22% of cases, any other adult members of the household. Before we started the intervention, we assessed whether the mothers could follow basic instructions by asking two simple questions (one in Bengali and one in English). 98.5% of the mothers answered the Bengali question correctly, while 73.3% of mothers answered the English question correctly. On average, mothers had completed 7 years of schooling, while fathers had completed 6 years. Since the lessons were primarily conducted in Bengali with occasional use of English, and almost all the parents have at least some years of primary schooling, they possessed the required literacy skills to ensure that the children engaged in the learning activities.

We designed the modules for this program with the support of two international organizations and a group of local curriculum experts. Overall, these modules were developed to supplement the national curriculum and support learning in household settings.

2.3 Treatments

We randomized each of the 90 participating villages into one of three groups (see Figure 1 for a summary of the research design):

- In the *Standard* treatment group, we offered the literacy and numeracy modules (60 lessons) to households in 30 villages.
- In the *Extended* treatment group, we offered the leadership module in addition to the literacy and numeracy modules to households in 30 villages. The leadership module focuses on development of leadership, listening and communication, and planning (see Section A-2.3 in Appendix).
- In the *Control* group (30 villages), no intervention was offered. These households did not have access to any of the modules offered to the two treatment groups.

Five lessons (two literacy, two numeracy and one leadership) were offered each week. After each lesson, a quiz was played to the listener, and the answers were recorded through the IVR system. To encourage listeners to complete the lessons, 30 listeners were randomly selected each week from the pool of listeners who answered quizzes correctly during that week to receive USD 3 (USD 8 in local purchasing power) as a prize via mobile financial services (MFS). Listeners who did not win the prize did not receive specific feedback on their quiz performance.

As previously mentioned, caregivers had the flexibility to choose and access any lesson for their child at any time during the program period. We did not mandate a fixed sequence in our curriculum, i.e., if learners found any lesson easy, they could skip the lesson and proceed to the next one. As students may learn better when they are provided with educational content that matches their level (Banerjee et al. 2007), we offered a menu of choices to allow participating children to tailor their learning experience according to their competencies. However, there is also evidence that lower-educated caregivers may face challenges in assessing their children's abilities (e.g., Dizon-Ross 2019). To prevent the intervention from potentially exacerbating existing educational inequalities, we recommended that caregivers follow the specific sequence outlined in the guidebook. Consequently, even though the caregivers had the option to skip lessons, among those who completed at least some lessons, nearly two thirds either completed all lessons or followed the sequence for the lessons completed.

3 Data and Empirical Method

3.1 Sample

In partnership with the Global Development and Research Initiative (GDRI), a local non-governmental organization (NGO) in Bangladesh, we conducted our study with a sample of

1,763 primary school-aged children and their caregivers in 90 villages in two southwestern districts (Khulna and Satkhira) (see Figure 2). GDRI had previously worked with a larger sample of children (more than 7,500) across 223 villages in these areas before the Covid-19 pandemic.[6] Consequently, we had access to household contact information and the pre-pandemic learning levels of these children.

From the list of contacts provided by GDRI, our first step was to randomly select 90 villages. We then selected 3,000 households with mobile phone numbers. We were able to reach and complete a baseline survey for 2,400 children from 2,387 households in May 2021 (see Figure 3 for the project timeline). Other households did not respond, had inactive or invalid phone numbers, or were not interested in participating in this intervention. We randomly selected about 16–22 children from each of these 90 villages.[7] Our final sample comprised 1,763 children from 1,755 households at the baseline. At the endline, we reached 1,687 households for the endline survey and assessments. Survey attrition rates are not different across the treatment arms (for further details, see Table 1, and Table B1 and Table B2).

In Table B3 (Appendix B), we provide a comparison of household characteristics of our sample to that of a rural Bangladeshi sample of households with children in primary school drawn from the 2018-19 Bangladesh Integrated Household Survey (or BIHS) (IFPRI 2020). Parents in our sample appear to be slightly younger and more educated relative to the BIHS sample, but are fairly similar in other characteristics, such as income, occupation, household size, and access to TV and phone.

[6] These children were participants in a completed project named "Investing in our Future" conducted by GDRI. In that project, households in participating villages with children in the 30-60 months age group were randomly recruited to be part of an early childhood program.
[7] There are four villages with fewer than 16 children: two villages had 10 children each, one village had 11 children, and one village had 13 children. We capped the sample size at 22 children per village to match the budget allocated for this study.

3.2 Data collection

Baseline assessment tests were conducted a year before this intervention as a part of GDRI's activities in the area. Due to the Covid-19-related school closures and mobility restrictions during the baseline period, we did not reassess the children immediately before the intervention commenced. Throughout the intervention, we also collected IVR-flow usage data, i.e., overall duration of lesson-play, access time and date, frequency, etc. from the server. We also requested households to record the lessons they completed on a printed sheet to keep track of lesson completion. 75 households did not return the printed sheets and 9 households returned blanked sheets. After the 15-week intervention, we surveyed the households and children again.

Both the baseline and endline surveys contain information related to demographics, income, employment status, household asset composition, livelihood, caregiver involvement in education, the children's educational situation, and the households' private educational investment. At the endline, a team of two members — one assessor and one enumerator — visited each household. The assessor conducted the literacy and numeracy assessments with the children and elicited their noncognitive skills, while the enumerator also conducted the caregiver survey with the mothers, which also covered questions that measure the children's leadership skills and behavioral difficulties (see Appendix A, Section A-3 for details).

3.3 Outcomes

Our analysis focuses on the following groups of prespecified outcomes.[8]

[8] One minor deviation from the pre-analysis plan relates to the assessment of children's learning outcomes. In the pre-analysis plan, we had specified the inclusion of a general knowledge component, as this is part of the national curriculum. However, as we finalized the material, we decided not to cover general knowledge questions in the assessment, as general knowledge was not covered in the intervention lesson plans. Consequently, the general knowledge component is not included in our measure of children's learning outcomes. In addition, our pre-analysis plan had initially proposed that we would examine whether parenting style could also be influenced by the intervention. However, we decided not to report and discuss this in the paper, as our primary focus is on children's outcomes, and we have already included more direct measures of parental input in our analysis.

Learning outcomes. Children's learning outcomes were measured using an assessment test that comprises 15 questions on literacy (English and Bangla) and numeracy. All questions were sourced from the national curriculum of Bangladesh. The test totals 80 points. We developed three sets of questions as our study involved students from different grades. The answers were set deliberately in binary form to avoid assessment bias. The questions of the assessment test are listed in Table A4 (Appendix A).

Leadership, communication, and planning skills. We employed the "Scales for Rating the Behavioral Characteristics of Superior Students" developed by Renzulli (2002). This scale has 14 subscales. We selected only the leadership, communication, and planning subscales as our modules focus on these dimensions. Items of these three characteristics are listed in Table A5 (Appendix A).

Noncognitive skills. We measured four types of noncognitive skills of the children. First, we measured their self-control by using the Impulsivity Scale for Children (ISC), an 8-item survey that assesses domain-specific impulsivity, defined as the "inability to regulate behavior, attention, and emotions in the service of valued goals" (Tsukayama, Duckworth, and Kim 2013). Second, we measured the grit of the participants using an 8-item grit scale developed by Duckworth and Quinn (2009). Grit is defined as "perseverance and passion for long-term goals" (Rimfeld et al. 2016) and this scale encompasses both perseverance of effort and consistency of interests. Grit and conscientiousness, one of the Big Five personality traits, are largely the same trait (Rimfeld et al. 2016). The conscientiousness scale includes competence, order, dutifulness, achieving striving, self-discipline and deliberation (John and Srivastava 1999), differentiating it slightly from the grit scale. As studies have shown that grit remains a significant predictor of life outcomes even after controlling for Big Five personality traits (Duckworth 2016; Duckworth and Eskreis-Winkler 2013; Duckworth et al. 2007; Eskreis-Winkler et al. 2014), we focus on grit instead of conscientiousness. Third, we assessed the

extent to which participating children view intelligence as a fixed behavioral trait rather than a feature that can be improved with effort using the 3-item growth mindset scale developed by Dweck, Chiu, and Hong (1995). Finally, we measured the impact of the intervention on the prosocial attitude of the children using the Empathy Questionnaire for Children and Adolescents (EmQue-CA) developed by Overgaauw et al. (2017). Items of these scales are listed in Table A6 (Appendix A).

While these measures have been validated for their English version, they lacked a version in Bangla. To address this, we engaged local psychologists to perform translations. Moreover, these measures have been shown to have associations with various behaviors, traits, and outcomes. Table A8 (Appendix A) provides an overview of both the validation status and the behavioral correlates associated with each measure.

Behavioral difficulties. We used the 25-item parent-report Strengths and Difficulties Questionnaire (SDQ) (1997), which has been validated for both their English and Bangla versions. The main motivation for using this scale was to evaluate whether school closures had any impact on students' fatigue, abnormal behavior, or conduct issues. Furthermore, we wanted to investigate whether participation in our program helped mitigate such problems by providing additional activities for students to engage in at home. Items of this scale are listed in Table A7 (Appendix A).

Homeschooling time (student). Students' time investment in homeschooling.

Homeschooling time (caregiver): Caregiver involvement in children's educational activities.

All outcome variables have been standardized following Kling, Liebman, and Katz (2007). First, we normalized the raw values by subtracting the mean values of the Control group sample and then dividing by the standard deviation of the raw values in the Control group sample.

3.4 Sample descriptive statistics

The average age of the children participating in this study was 7.4 years and the age range was 5.1 to 9.9 years. Around 51.3% of the children were girls. The mean years of schooling of their father and mother were 5.9 and 7.1 years, respectively. These children primarily came from households with a low socio-economic status, with an average monthly income of BDT 11,003 (USD 130.8). Slightly more than half of the sample reported to have access to private tuition. In terms of access to distance learning modalities, these households were also quite disadvantaged. In May 2021, only 46.5% of them had access to TV, and less than 1% had access to computers or radio. However, all households had access to mobile phones, of which 35.6% were smartphones. Nevertheless, these smartphones were rarely used for educational purposes due to the high data costs and low internet speeds, particularly in rural areas. Table 1 presents summary statistics and balance tests for various characteristics of our endline sample, showing that these characteristics are balanced across the treatment and control groups.

3.5 Usage

Providing access to distance education does not guarantee learning as students may not use the resources for many reasons, such as difficulties with navigating the system, technical challenges with accessing the lessons, and time constraints of parents. It is therefore important to consider the level of student engagement with the audio lessons offered during our intervention.

Overall, participants' engagement was high. On average, each student completed 46 recorded lessons in the Standard group and 58 recorded lessons in the Extended group. Around 70% of participating children completed more than two thirds of lessons in each of the modules (Figure B3). Furthermore, there is a downward trend in the number of listeners and total hours of audio lessons accessed as the program advances (see Figure B4 and Figure B5). One potential reason

is that some of the children might have lost interest after a few lessons or decided to skip some lessons for not finding them interesting or needed as they progressed. The other reason may be that schools were partially reopened in the middle of week 11 of our intervention. In particular, the government ordered all schools to reopen for one to two days per week starting from September 12, 2021. The government made this announcement roughly one week in advance. Figure B5 shows that the total hours of lessons accessed started to fall after reaching its peak in week 8 of our intervention. The drop in the total hours of audio lessons accessed after week 10 largely follows the ongoing downward trend. Despite the partial reopening of all schools after week 10, the total hours accessed in the remaining weeks stabilized around at least half of those in week 8. Given that the reopening of schools did not entirely replace students access to the audio lessons, the intervention appears to work as a supplementary resource alongside formal schooling. All in all, the number of lessons completed is moderately high, indicating that the intervention was well-received by the target group of households.

3.6 Empirical specification

To assess the overall effects of the treatments on the various outcomes, we estimate the following Ordinary Least Squares (OLS) regression specification:

$$Y_i = \alpha + \beta_1 T_{1i} + \beta_2 T_{2i} + \theta' X_i + \epsilon_i, \qquad \text{(Equation 1)}$$

where Y_i is an outcome of a child from household i measured at the endline. The treatment indicator T_{1i} takes the value of one if the child is in the Standard group, and zero otherwise. The treatment indicator T_{2i} takes the value of one if the child is in the Extended group, and zero otherwise. The coefficients of interest are β_1 and β_2, which capture the causal effect of a treatment on an outcome. We also include a vector of individual and household-specific characteristics in the regression specification. X_i includes the child's age, gender, access to

private tuition, parental education in years, family income, religion, access to TV, access to smartphone, homestead size, number of members in the household, and the relevant outcome measured at baseline. Finally, the error term ϵ_i captures all other unobserved influences.

We cluster the standard errors at the village level. We also separately report the Family Wise Error Rate (FWER) adjusted p-values corrected for multiple hypotheses testing using the free step-down resampling approach to account for the large set of outcomes that we considered in this study (Westfall and Young 1993). Furthermore, to account for uncertainty in the estimated treatment effects that arise naturally from the random assignment of participants into the treatments, we also report p-values using randomization-based inference (RI) (Young 2019).

4 Results

4.1 Learning outcomes

Figure 4 summarizes the treatment effects of the intervention on the two main learning outcomes (literacy and numeracy), along with their combined score. These treatment effects are obtained from estimating Equation 1 with OLS regressions. In panel A of Figure 4, which considers the full sample, we see that treated children in the Standard and Extended groups experienced a 0.60 SD and 0.63 SD improvement in total score relative to children in the control group, respectively. In the literacy component, treatment effects were 0.55 SD ($p<0.001$) and 0.59 SD ($p<0.001$) for the Standard and Extended treatment arms, respectively. In the numeracy component, treatment effects were 0.53 SD ($p<0.001$) and 0.54 SD ($p<0.001$) for the Standard and Extended treatment arms, respectively (see Table 2 for details).

Our standardized effect sizes are substantial. It is informative to consider alternative metrics to better gauge the magnitude of the effects. In terms of raw test scores, the treatment groups achieved about 30% higher total scores than the control group, which translates to about 2.5 additional correct answers (out of 15), when the average number of correct answers in the

control group was 8 (see Table B4). We can also provide an internal benchmark for the effect sizes measured in SDs by comparing them to learning gaps observed between different demographic groups in our sample. Among children in the control group, we estimate large differences in two dimensions: between children who received private tuition and those who did not (a gap of 0.29SD), and between Hindu and Muslim children (with a gap favoring Hindus of 0.47SD). In these terms, the intervention can be thought of as generating nearly double the effect observed between the first group (private tuition) and is comparable to the gap between the second group (religion). Furthermore, in Figure B6, Panel A presents the distribution of test scores, and panel B presents a percentile-to-percentile comparison of the distributions of treated and control children. This figure indicates that, for instance, the 30^{th} percentile of the treatment group distribution corresponds to the 60^{th} percentile of the control group distribution.

Since the questions in our assessment tests are designed to be correctly answered by students meeting the learning standards for their grade, we can classify students into those who met the numeracy and literacy standards (answering all questions correctly) and those who did not. Using this classification, we observe that 9% of the students in the control group reached the literacy standard, and 21% reached the numeracy standard. When pooling observations across the two treatment groups, the respective proportions are significantly higher at 17% for literacy and 36% for numeracy, implying that the intervention more than doubled the literacy and numeracy levels, starting from a markedly low standard. Table B5 (Appendix B) presents a formal treatment effect analysis based on the number of correct answers.

These results indicate that the provision of audio lessons in a context where no access to formal education is available can result in significant improvements in students' learning outcomes, both in literacy and numeracy.

We also examine the relationship between the number of lessons accessed and our main outcomes of interest, test scores on numeracy and literacy. We categorize the number of lessons

accessed into five groups. The first group is the control group. The second group includes those who completed one to 19 lessons (roughly 25% of those who completed any lesson), the third group includes those who completed 20 to 29 lessons (roughly 25% of those who completed any lesson), and the fourth group includes those who completed all lessons (roughly 45% of those who completed any lesson). The final group includes those who reported that they accessed some lessons.[9] These results are presented in Figure B7.[10] Having accessed and completed some lessons are both associated with a significant increase in literacy and numeracy test scores. However, we do not find significant differences across dosage levels. For example, those who completed some lessons performed similarly to those who completed all lessons. It is important to note that these results should not be interpreted as causal, given that the number of lessons accessed or completed is likely to be endogenous.

Panels B and C of Figure 4 show the treatment effects by gender. The intervention seems to have benefitted equally boys' and girls' test scores. This is confirmed through regression analysis in which treatment is interacted with gender (see Table 3).

4.2 Leadership

The Extended treatment group was offered an additional module that focused on leadership, qualities of a leader, active listening, communication and presenting, and planning. At the endline, we collected measures of these skills using scales for Rating the Behavioral Characteristics of Superior Students (Renzulli et al. 2002). Treatment effects on these measures are presented in Figure 5, Panel A. We do not find any evidence that children in the Extended

[9] Before the intervention began, participating households provided us with the phone numbers they intended to use to access the audio lessons. In addition, we requested households to record their lessons on a printed sheet, which served as a backup method to keep track of lesson completion. This last group of individuals indicated that they had accessed the audio lessons on the printed sheet, but they did not indicate which lessons were completed. Their exact number of lessons accessed or completed also could not be verified with the IVR server data due to the use of non-matching phone numbers.

[10] We exclude six individuals that reported to have never completed any lessons, 84 individuals who did not provide any information on the printed sheet (and we are unable to verify their access with the IVR server data due to the use of non-matching phone numbers), and 12 individuals whose parents did not complete the endline survey (to be consistent with results reported in previous tables).

treatment improved on these measures relative to the control group—treatment effects are small and statistically insignificant. This suggests that improving leadership skills might be difficult to achieve through this distance learning medium for children of this age.

4.3 Noncognitive skills

While learning outcomes are the core targets of this educational intervention, the development of noncognitive skills, such as critical thinking, problem-solving skills, social skills, persistence, creativity, and self-control, is an integral part of any educational program. Therefore, in our intervention, through various examples and discussions, we aimed to cultivate these noncognitive skills of treated children. For example, listening carefully to others, being patient, and setting a goal and then following it up, are key steps that children must undergo during each lesson. These steps provide children with opportunities to develop their impulse control and grit. Similarly, in the leadership module, children were also taught the importance of empathy, compassion, humility, patience, and perseverance. During the endline, we measured the level of noncognitive skills of the children. We chose commonly used noncognitive skill measurement scales, including impulsivity, grit, growth mindset, and empathy instruments.

Treatment effects on noncognitive skills are presented in Figure 5, Panel B. We find that our intervention was effective in reducing the impulsive behavior of children only in the Standard treatment group, with the overall impulsivity component decreasing by 0.27 SD ($p<0.05$). However, this effect is not robust when considering FWER adjusted p-values (Table 2). We also do not find any significant treatment effects on the other three measures: grit, growth mindset, and level of empathy.

4.4 Behavioral difficulties

Our intervention started at a time when the children were not attending school, meaning that they had not been following a formal educational routine for about 15 months. Because of the lack of school attachment, one might be concerned that children could develop behavioral problems, e.g., tantrums, nervousness, lack of control over emotions, and hyperactivity. These problems might be reduced by participating in our intervention. To check this possibility, we use the 25-item parent-report Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997). This popular scale covers five domains of children's behaviors, i.e., emotional symptoms, conduct problems, hyperactivity, peer problem and prosociality.

In all the five domains, we find statistically significant treatment effects for children in the Standard treatment (Figure 5, Panel C). Children in this group showed less emotional symptoms, conduct problems, hyperactivity, peer problems and a more prosocial attitude compared to the children in the control group. Treatment effects in various domains of SDQ range from -0.17 SD to -0.24 SD ($p<0.05$). However, these effects are not robust when we consider FWER adjusted p-values (Table 2). In the Extended group, treatment effects in various domains of SDQ range from -0.02 SD to -0.13 SD but are not statistically significant whether we consider p-values based on robust standard errors clustered at the village level, FWER adjusted p-values, or RI p-values. One potential explanation for the smaller effects in the Extended group is that the greater number of activities provided might unintentionally disrupt the flow of the learning process and consequently reduce any potential associated benefits.

4.5 Heterogeneous effects

In Table 3, we present some regression-based tests of heterogeneity of the intervention on test scores along three dimensions: baseline test scores, household income, and parents' education. For each dimension, we create a binary variable for being above the median and interact it with

the treatment. We find large treatment effects across all subgroups with a tendency for the treatment effects to be larger for groups below the median. However, the two dimensions of heterogeneity that are statistically significant are baseline test scores and parents' education concerning numeracy. For literacy, the differences in the treatment effects between the high and low groups are smaller and not statistically significant.

We further explore these heterogeneous treatment effects in Figure 6. The top subfigure presents the mean test score of the two treatment groups and the control group by quartile of baseline test scores. The differences in test scores are most pronounced for students who scored the lowest in the baseline assessment. For students in the 1st and 2nd quartiles (weakest performers), the gaps between the treatment and control groups are as large as 40%, whereas, in the other two quartiles, the gaps are less than 30%. All these gaps are statistically significant at the 5% level. However, when we conduct pairwise comparisons of treatment effects across quartiles, we find that they are not statistically distinguishable (see columns 1 and 4 of Table B6). These results indicate that the program helped students across the spectrum of prior academic standing to improve their learning.

The middle subfigure of Figure 6 shows the total scores by family income. Again, the treatment effects are most pronounced for children from households in lower-income quartiles. The gains in test scores are more than 35% in the 1st and 2nd quartiles, whereas they are around 20% in the top two quartiles. However, similar to the baseline test scores, treatment effects across income quartiles are not statistically distinguishable (see columns 2 and 5 of Table B6).

Finally, the bottom subfigure of Figure 6 shows the heterogeneity in the treatment effects based on parental education. In this case, we find that treatment effects are significantly higher for students with less-educated parents (lowest quartile) compared to those in the first quartile, in both treatment groups (columns 3 and 6 of Table B6). The differential effects by parental education are consistent with findings in other recent distance learning interventions. For

example, Hassan et al. (2021) find stronger treatment effects on the literacy and numeracy tests of children with less-educated parents than children with more-educated parents in Bangladesh in the one month after a 13-week telementoring educational intervention ended during the Covid-19 pandemic. Similarly, Carlana and La Ferrara (2021) also find stronger treatment effects on academic performance of middle school students with less-educated mothers in Italy after five weeks of online tutoring that lasted roughly 3 hours per week during the Covid-19 pandemic. Overall, these results suggest that our intervention was especially beneficial for students with low baseline test-scores and less-educated caregivers.

4.6 Robustness

Social desirability bias

As some of our outcome variables were collected via surveys, social desirability bias may arise, which can lead to over or underestimation of the treatment effects on these outcomes. To address this concern, during the endline, we elicited the social desirability bias of parents using the short-form Crowne-Marlowe module (Crowne and Marlowe 1960; Reynolds 1982) and of the children using the Children's Social Desirability (CSD-S) scale (Miller et al. 2015) (Section A-4). In Table B7 and B8 of Appendix B, we report an analysis that suggests that the significant effects of the intervention are mainly driven by participants with lower social desirability bias, thereby alleviating concerns about social desirability bias driving our results.

Assessment reliability

Our assessment tests are based on the curriculum designed by the National Curriculum and Textbook Board (NCTB) of Bangladesh. This curriculum has been already rigorously tested and modified over the years based on the changing aptitude of the learners. We picked 15 questions from the latest version of the textbooks for primary grades. To verify the coherence

of these selected items, we report some statistical tests (see Section A-5) that indicate that our assessment tests were stable, reliable, and coherent.

5 Potential Channels and Mechanisms

5.1 Did the intervention impact homeschooling time investment?

We next examine the impact of the intervention on time spent homeschooling, which we split into two components: students' study time and caregivers' homeschooling time. These variables are defined in Appendix A, Sections A-3.6 and A-3.7.

We do not find a statistically significant treatment effect on the students' study time (See Table 4). It is important to note that our sample children were in primary school during the intervention. Self-study is not very common for children of these ages, which probably explains why we do not find a significant impact of the intervention on students' study time.

However, we find that caregivers who participated in the intervention devoted more time to their children's education-related tasks compared to caregivers in the Control group — 10.4 minutes per day in the Standard group and 3.1 minutes per day in the Extended group. The treatment effect in the Standard group was 0.20 SD ($p<0.05$). When we consider FWER adjusted p-values (Table 4), this effect is generally more robust measured in the Likert scale, than measured in minutes. In the case of the Extended group, although the treatment effect is positive, it is not statistically significant. These results suggest that the phone-based education program may have encouraged caregivers to engage more in their children's education, extending their involvement beyond the direct program time. This sustained engagement of caregivers may have enhanced the effectiveness of children's learning activities.

5.2 Can targeted learning explain the treatment effects?

While caregivers were advised to follow the sequence of lessons outlined in the guidebook, they had the option to select only audio lessons they believed suitable for their children based

on their level of prior academic preparation. When caregivers opted for this targeted learning approach, it often resulted in certain lessons being skipped. Such targeted learning is likely to be less common among caregivers with lower levels of education, as they might find it more challenging to assess their children's competencies (Dizon-Ross 2019). Instead, these caregivers might just follow the recommended sequence. Similarly, students with lower baseline test scores might target their learning by following the sequence, as it gradually increased the complexity of the learning materials. As we observed stronger treatment effects among students with less-educated parents, those with lower baseline test scores, and to some extent, students from lower-income families (see Table 3), differences in how they adhered to the sequence of lessons by parental education, baseline achievement, and family income, could potentially explain why our intervention is particularly effective for these groups.

Table 5 explores the relationships between different types of lesson completion and parental education type, baseline achievement type, and family income type. Note that we did not specify the following exploratory analysis in our registered pre-analysis plan. In columns 1 and 2, we examine whether parents with lower levels of education, lower baseline test score of children, and lower household income are associated with students strictly following the recommended sequence without skipping any lesson. In columns 3 and 4, we examine whether these characteristics are associated with the total number of lessons completed. In columns 5 and 6, we examine whether these characteristics are associated with the percentage of completed lessons ever repeated. Column 1 indicates that less-educated parents were 7 percentage points more likely to follow the recommended sequence of literacy lessons than more-educated parents. Column 2 indicates that children with lower baseline numeracy scores were 8 percentage points more likely to follow the recommended sequence of numeracy lessons than children with higher baseline numeracy scores. These findings may explain why children with less-educated parents and those with lower baseline numeracy scores experienced larger

treatment effects. Columns 3 and 5 also show that children with lower baseline literacy scores tend to complete fewer literacy lessons and are less likely to revisit those they have completed. In contrast, children with higher baseline literacy scores tend to complete more literacy lessons and repeat them. These patterns may explain the observed similarity in treatment effects between children with lower and higher baseline literacy scores. Lastly, there is no statistically significant association between family income type and any of the three lesson completion types.

Overall, this exploratory analysis provides suggestive evidence that different types of targeted learning approaches were employed, and the heterogenous treatment effects by parental education and baseline test scores are consistent with the approaches taken.

6 Discussion

The intervention proved highly effective in improving the learning outcomes of children in the treatment groups, especially those from a low-socioeconomic background. The effect sizes fall within the range of other educational interventions implemented during the Covid-19 pandemic. On one hand, they are somewhat larger than the effect of an eight-week after-school online math tutoring program provided to secondary school students (N=356; 16 hours) in highly disadvantaged neighborhoods in Spain six months after Covid-19 school closures ended (Gortazar, Hupkau, and Roldán 2022), a five-week volunteer-based online tutoring (N=1,059; 17 hours) in Italy (Carlana and La Ferrara 2021), and a 12-week pilot program of online tutoring (N=560; 4 hours) by college-volunteers in the US (Kraft et al. 2022). The effect sizes are also larger than those reported for an eight-week direct phone call based tutoring program (N=4,550; 3 hours) in Botswana (Angrist, Bergman, and Matsheng 2022) and a 16-week live phone tutorial intervention tied to radio instruction (N=4,399; 4 hours) in Sierra Leone, which increased educational activity but not test scores (Crawfurd et al. 2023). On the other hand, they are slightly smaller than the effect of a 13-week phone-based mentoring program provided

to primary school age children (N=838; 6.5 hours) in rural Bangladesh by volunteers during Covid-19 school closures (Hassan et al. 2021). Besides differences in sample size, type and intensity of engagement, and timing of the intervention across these studies, other differences that might explain the variation in effects sizes include the mode of assessment (phone-based or in-person), the age of targeted children, the type of tutor, and the extent of parental involvement.

The large effects on learning outcomes found in this study are likely due to the 18-month-long school closures in Bangladesh, which prevented children in the control group from having access to any educational services. This implies that the learning levels of these children during this period were likely low, making large improvements easier to achieve.[11] One might expect that the large differences between the treatment and control groups would diminish over time after children in the control group gained access to formal education. For instance, in a similar setting, it has been reported that one year after a phone-based tutoring intervention ended, with children in the control group returning to school, the intervention's effect decreased by roughly 20% to 55% (Hassan et al. 2021).

The intervention was largely ineffective in changing behavioral difficulties and noncognitive skills. The null results here are consistent with those of two online goal-setting interventions examined in Dobronyi, Oreopoulos, and Petronijevic (2019), which involved approximately 1,400 first-year undergraduate students. These goal-setting interventions, which lasted for approximately two hours, did not significantly affect the treated students' likelihood of persisting into their second year of undergraduate studies. In contrast, Alan, Boneva, and Ertac (2019) evaluated two in-person grit interventions and found an increase in participants' grit

[11] In a similar vein, Eble et al. (2021) also attribute the large effect sizes of their educational intervention to the low learning levels in the low-income and remote rural setting in The Gambia that they studied.

score by roughly 0.3 SD. These two studies involved 2,600 and 1,500 fourth-grade students, respectively, and the interventions were delivered in after-school classes for at least two hours per week over the course of 12 weeks. One noteworthy feature of these grit interventions is that the participating teachers were also encouraged to adopt a teaching philosophy that emphasizes the role of effort in everyday classroom practices. Thus, it is possible that the ineffectiveness of our intervention in altering behaviors and noncognitive skills can be attributed to its remote delivery and lack of integration into the teaching philosophy.

Based on the exchange rate of 80 BDT to 1 USD (at the time of the intervention), the 15-week intervention cost USD 27.5 per student, of which USD 13.2 was attributed to variable costs and the remaining USD 14.3 to fixed costs. Fixed costs include expenses related to IVR platform development, module content development, and program administration. Variable costs mainly include voice and SMS charges. It is important to note that these cost figures encompass both provider and recipient costs. Table B13 in the Appendix B provides a detailed breakdown of these various cost components. Our intervention is among the most cost-effective ones implemented during the Covid-19 school closures. Using the approach in Kremer, Brannen, and Glennerster (2013) to compare cost-effectiveness, our treatment effect estimates indicate that this intervention could achieve improvements of 2.18 SD and 2.29 SD per USD 100 of spending in the Standard and Extended treatment groups, respectively. In comparison Hassan et al. (2021), achieved a slightly higher effect size of 3.1 SD per USD 100, while Angrist, Bergman, and Matsheng (2022) achieved a slightly lower effect size of 0.89 SD per USD 100. This places our intervention at a middle point in terms of effectiveness when compared to these two studies. If the duration of the intervention were to be lengthened, the total cost would increase proportionately to our 15-week intervention. As only a total of 1,182 students across two districts received this intervention, the fixed cost per student was high relative to the variable cost. Scaling up the intervention to include more students would likely

lower the per-student fixed cost. Furthermore, there is further scope to lower the variable cost if the intervention were to be scaled up, as it is likely that a lower phone call rate can be negotiated with telecommunication companies.

7 Conclusion

The household environment plays a critical role in education, but most education policies primarily focus on school-based interventions, as it is believed that it is more feasible to improve schools than to intervene at the household level at scale (Muralidharan and Singh 2021). However, the school closures induced by the Covid-19 pandemic have sharply shifted the focus from the school to the household environment. Due to the weak information communication technology's ecosystem in most low-income developing countries, widely accessible basic feature phones have become popular in educating the mass of students during the Covid-19 pandemic (Hassan et al. 2021). Existing studies using basic feature phones are limited to SMS reminders or brief calls to parents to follow up on their children's homework (Angrist, Bergman, and Matsheng 2022; Lichand and Christen 2021; Muralidharan and Singh 2021).

In this paper, we offer an extension of the existing applications of basic feature phones in education by using the IVR system to deliver lessons. We delivered these lessons via basic mobile phones because their penetration rate in rural Bangladesh is significantly higher than other one-way technologies such as radio and television. Our results indicate that this approach delivered substantial learning benefits to students, especially for those who are more disadvantaged. However, it is worth noting that impacts of the intervention on measures of leadership and other noncognitive skills were not found to be significant. Considering the importance of soft skill development among children and the challenges in measuring these skills in developing countries (Laajaj et al. 2019), it would be very valuable for future research

to delve into the mechanisms and dynamics behind the development of these skills through remote learning approaches.

Although the extent of learning disruptions caused by the Covid-19 pandemic was unprecedented, educational disruptions on a smaller scale is not uncommon in low- and middle-income countries. In many developing countries, climate change, natural and human-induced events (e.g., cyclones, floods, wars, and political unrest) often damage educational infrastructure and limit school operations. Therefore, policymakers may consider expanding education delivery in out-of-school settings using accessible distant learning methods, such as the one studied in this paper, to better support children's learning in these situations. Importantly, the relatively low cost of the intervention examined in this study, especially if it is provided to a large number of children, makes it a promising option for providing remedial educational support to poor and academically left-behind students in hard-to-reach areas, even outside of times of crisis. A fruitful avenue for future research would be to examine the effectiveness of this type of intervention in such settings.

References

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2022. "When the great equalizer shuts down: Schools, peers, and parents in pandemic times." *Journal of public economics* 206: 104574.

Aker, Jenny C, and Christopher Ksoll. 2019. "Call me educated: Evidence from a mobile phone experiment in Niger☆." *Economics of Education Review* 72: 239-257.

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit." *The Quarterly Journal of Economics* 134 (3): 1121-1162.

Angrist, Noam, Peter Bergman, and Moitshepi Matsheng. 2022. "Experimental evidence on learning using low-tech when school is out." *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01381-z. https://doi.org/10.1038/s41562-022-01381-z.

Anzalone, Stephen, and Andrea Bosch. 2005. "Improving educational quality with interactive radio instruction: a toolkit for policymakers and planners."

Ayman-Nolley, Saba, and Roya Ayman. 2005. "Children's implicit theory of leadership." *Implicit leadership theories: Essays and explorations*: 227-275.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying education: Evidence from two randomized experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235-1264.

Beam, Emily A., Priya Mukherjee, Laia Navarro-Sola, Junnatul Ferdosh, and Md. Afzal Hossain Sarwar. 2021. Take-Up, Use, and Effectiveness of Remote Technologies. In *Working Paper Series*. Washington, DC: The World Bank.

Beg, Sabrin, Waqas Halim, Adrienne M Lucas, and Umar Saif. 2022. "Engaging teachers with technology increased achievement, bypassing teachers did not." *American Economic Journal: Economic Policy* 14 (2): 61-90.

Bergman, Peter, and Eric W Chan. 2021. "Leveraging Parents through Low-Cost Technology The Impact of High-Frequency Information on Student Achievement." *Journal of Human Resources* 56 (1): 125-158.

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." *Unpublished Manuscript*.

Bianchi, Nicola, Yi Lu, and Hong Song. 2022. "The effect of computer-assisted learning on students' long-term development." *Journal of Development Economics* 158: 102919.

Billsberry, Jon, Claudia Escobar Vega, and John Molineux. 2019. "Think of the children: Leader development at the edge of tomorrow." *Journal of Management & Organization* 25 (3): 378-381.

Bird, Kelli A, Benjamin L Castleman, Jeffrey T Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger. 2021. "Nudging at scale: Experimental evidence from FAFSA completion campaigns." *Journal of Economic Behavior & Organization* 183: 105-128.

Biswas, Kumar, TM Asaduzzaman, David K Evans, Sebastian Fehrler, Deepika Ramachandran, and Shwetlena Sabarwal. 2020. "TV-Based Learning in Bangladesh."

Bosch, Andrea. 1997. "Interactive Radio Instruction: Twenty-three years of improving education quality." *Washington, DC: World Bank Group*.

Bosch, Andrea, Rebecca Rhodes, and Sera Kariuki. 2002. *Interactive Radio Instruction: an Update from the Field.* Knowledge Enterprise, Inc.

Cardim, Joana, Teresa Molina-Millán, and Pedro C Vicente. 2023. "Can technology improve the classroom experience in primary education? An African experiment on a worldwide program." *Journal of Development Economics* 164: 103145.
Carlana, Michela, and Eliana La Ferrara. 2021. *Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic.*
Castleman, Benjamin L, and Katharine E Meyer. 2020. "Can text message nudges improve academic outcomes in college? Evidence from a West Virginia initiative." *The Review of Higher Education* 43 (4): 1125-1165.
Crawfurd, Lee, David K. Evans, Susannah Hares, and Justin Sandefur. 2023. "Live tutoring calls did not improve learning during the COVID-19 pandemic in Sierra Leone." *Journal of Development Economics* 164: 103114. https://doi.org/https://doi.org/10.1016/j.jdeveco.2023.103114. https://www.sciencedirect.com/science/article/pii/S030438782300069X.
Crowne, Douglas P, and David Marlowe. 1960. "A new scale of social desirability independent of psychopathology." *Journal of consulting psychology* 24 (4): 349.
DataReportal. 2021. "Digital 2021: Bangladesh." Hootsuite & We Are Social. Accessed 18 November. https://datareportal.com/reports/digital-2021-bangladesh.
Dizon-Ross, Rebecca. 2019. "Parents' beliefs about their children's academic ability: Implications for educational investments." *American Economic Review* 109 (8): 2728-65.
Dobronyi, Christopher R, Philip Oreopoulos, and Uros Petronijevic. 2019. "Goal setting, academic reminders, and college success: A large-scale field experiment." *Journal of Research on Educational Effectiveness* 12 (1): 38-66.
Duckworth, Angela Lee. 2016. *Grit: The power of passion and perseverance*. Vol. 234. Scribner New York.
Duckworth, Angela Lee, and Lauren Eskreis-Winkler. 2013. "True grit." *Observer*.
Duckworth, Angela Lee, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: perseverance and passion for long-term goals." *Journal of personality and social psychology* 92 (6): 1087.
Duckworth, Angela Lee, and Patrick D Quinn. 2009. "Development and validation of the Short Grit Scale (GRIT–S)." *Journal of personality assessment* 91 (2): 166-174.
Dweck, Carol S, Chi-yue Chiu, and Ying-yi Hong. 1995. "Implicit theories and their role in judgments and reactions: A word from two perspectives." *Psychological inquiry* 6 (4): 267-285.
Eble, Alex, Chris Frost, Alpha Camara, Baboucarr Bouy, Momodou Bah, Maitri Sivaraman, Pei-Tseng Jenny Hsieh, Chitra Jayanty, Tony Brady, and Piotr Gawron. 2021. "How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in The Gambia." *Journal of Development Economics* 148: 102539.
Engzell, Per, Arun Frey, and Mark D Verhagen. 2021. "Learning loss due to school closures during the COVID-19 pandemic." *Proceedings of the National Academy of Sciences* 118 (17): e2022376118.
Eskreis-Winkler, Lauren, Elizabeth P Shulman, Scott A Beal, and Angela Lee Duckworth. 2014. "The grit effect: Predicting retention in the military, the workplace, school and marriage." *Frontiers in psychology* 5: 36.
Goodman, Robert. 1997. "The Strengths and Difficulties Questionnaire: a research note." *Journal of child psychology and psychiatry* 38 (5): 581-586.
Gortazar, Lucas, Claudia Hupkau, and Antonio Roldán. 2022. "Online tutoring works: Experimental evidence from a program with vulnerable children." *Working Paper*.

Gutman, Leslie Morrison, and Ingrid Schoon. 2013. "The impact of non-cognitive skills on outcomes for young people. A literature review."
Hassan, Hashibul, Asad Islam, Abu Siddique, and Liang Choon Wang. 2021. "Telementoring and homeschooling during school closures: A randomized experiment in rural Bangladesh."
Ho, Jennifer, and Hetal Thukral. 2009. "Tuned in to student success: Assessing the impact of interactive radio instruction for the hardest-to-reach." *Journal of Education for International Development* 4 (2): 34-51.
Hurwitz, Lisa B, Alexis R Lauricella, Ann Hanson, Anthony Raden, and Ellen Wartella. 2015. "Supporting Head Start parents: impact of a text message intervention on parent–child activity engagement." *Early Child Development and Care* 185 (9): 1373-1389.
IFPRI. 2020. Bangladesh Integrated Household Survey (BIHS) 2018-2019. edited by International Food Policy Research Institute: Harvard Dataverse.
Islam, Asad, Abdul Malek, Sakiba Tasneem, and Liang Choon Wang. 2022. "Can public recognition reward backfire? Field experimental evidence on the retention and performance of volunteers."
John, Oliver P, and Sanjay Srivastava. 1999. "The Big-Five trait taxonomy: History, measurement, and theoretical perspectives."
Johnston, Jamie, and Christopher Ksoll. 2022. "Effectiveness of interactive satellite-transmitted instruction: Experimental evidence from Ghanaian primary schools." *Economics of Education Review* 91: 102315.
Karagianni, Despoina, and Anthony Jude Montgomery. 2018. "Developing leadership skills among adolescents and young adults: a review of leadership programmes." *International Journal of Adolescence and Youth* 23 (1): 86-98.
Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75 (1): 83-119.
Kraft, Matthew A, and Shaun M Dougherty. 2013. "The effect of teacher–family communication on student engagement: Evidence from a randomized field experiment." *Journal of Research on Educational Effectiveness* 6 (3): 199-222.
Kraft, Matthew A, John A List, Jeffrey A Livingston, and Sally Sadoff. 2022. "Online tutoring by college volunteers: Experimental evidence from a pilot program." AEA Papers and Proceedings.
Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The challenge of education and learning in the developing world." *Science* 340 (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.
Laajaj, Rachid, Karen Macours, Daniel Alejandro Pinzon Hernandez, Omar Arias, Samuel D Gosling, Jeff Potter, Marta Rubio-Codina, and Renos Vakis. 2019. "Challenges to capture the big five personality traits in non-WEIRD populations." *Science advances* 5 (7): eaaw5226.
LEAD. 2021. LEAD Learning. 2021 (15 January).
Lichand, Guilherme, and Julien Christen. 2021. *Behavioral nudges prevent student dropouts in the pandemic.*
Mayer, Susan E, Ariel Kalil, Philip Oreopoulos, and Sebastian Gallegos. 2015. *Using behavioral insights to increase parental engagement: The parents and children together (PACT) intervention.* National Bureau of Economic Research.
Miller, Patricia H, Suzanne D Baxter, Julie A Royer, David B Hitchcock, Albert F Smith, Kathleen L Collins, Caroline H Guinn, Alyssa L Smith, Megan P Puryear, and Kate

K Vaadi. 2015. "Children's social desirability: Effects of test assessment mode." *Personality and individual differences* 83: 85-90.
Moscoviz, Laura, and David K Evans. 2022. Learning loss and student dropouts during the covid-19 pandemic: A review of the evidence two years after schools shut down. In *Center for Global Development, Working Paper*.
Muralidharan, Karthik, and Abhijeet Singh. 2021. "India's new National Education Policy: Evidence and challenges." *Science* 372 (6537): 36-38. https://doi.org/10.1126/science.abf6655. https://science.sciencemag.org/content/sci/372/6537/36.full.pdf.
Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109 (4): 1426-60. https://doi.org/10.1257/aer.20171112. https://www.aeaweb.org/articles?id=10.1257/aer.20171112.
Murphy, Susan Elaine, and Stefanie K Johnson. 2011. "The benefits of a long-lens approach to leader development: Understanding the seeds of leadership." *The Leadership Quarterly* 22 (3): 459-470.
Overgaauw, Sandy, Carolien Rieffe, Evelien Broekhof, Eveline A Crone, and Berna Güroğlu. 2017. "Assessing empathy across childhood and adolescence: Validation of the Empathy Questionnaire for Children and Adolescents (EmQue-CA)." *Frontiers in psychology* 8: 870.
Patrinos, Harry Anthony. 2022. "Learning loss and learning recovery." *DECISION*: 1-6.
Rahman, Hossain Zillur, Imran Matin, Atiya Rahman, Narayan C Das, Umama Zillur, Md. Shakil Ahmed, Syed M. Hashemi, Mohammad Abdul Wazed, Tanvir Ahmed Mozumder, Sabrina Miti Gain, Fatema Mohammad, Tahsina Naz Khan, Montajina Tasnim, Nusrat Jahan, Md. Saiful Islam, Avinno Faruk, and Namira Shameem. 2021. "PPRC-BIGD COVID-19 Livelihoods & Recovery Panel Survey." BIGD. Accessed 18/11. https://bigd.bracu.ac.bd/study/rapid-survey-on-immediate-economic-vulnerabilities-created-by-covid-19-and-the-coping-mechanisms-of-poor-and-marginal-people/.
Rahman, Tashmina, and Uttam Sharma. 2021. "A Simulation of COVID-19 School Closure Impact on Student Learning in Bangladesh." *World Bank*.
Renzulli, Joseph S, Linda H Smith, Alan J White, Carolyn M Callahan, Robert K Hartman, and Karen L Westberg. 2002. *Scales for rating the behavioral characteristics of superior students. Technical and administration manual*. ERIC.
Reynolds, William M. 1982. "Development of reliable and valid short forms of the marlowe-crowne social desirability scale." *Journal of Clinical Psychology* 38 (1): 119-125.
Rimfeld, Kaili, Yulia Kovas, Philip S Dale, and Robert Plomin. 2016. "True grit and genetics: Predicting academic achievement from personality." *Journal of personality and social psychology* 111 (5): 780.
Salmond, Kimberlee, and Paula Fleshman. 2010. "Gender and Women's Leadership: A Reference Handbook." edited by Karen O'Connor. Thousand Oaks

Thousand Oaks, California: SAGE Publications, Inc. https://doi.org/10.4135/9781412979344. https://sk.sagepub.com/reference/womensleadership.
Schueler, Beth E., and Daniel Rodriguez-Segura. 2023. "A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya." *Journal of Research on Educational Effectiveness* 16 (3): 442-472. https://doi.org/10.1080/19345747.2022.2131661. https://doi.org/10.1080/19345747.2022.2131661.

Singh, Abhijeet, Mauricio Romero, and Karthik Muralidharan. 2022. COVID-19 Learning Loss and Recovery: Panel Data Evidence from India. National Bureau of Economic Research.
Sisk, Dorothy A. 1993. "Leadership education for the gifted." In *International handbook of research and development of giftedness and talent*, edited by Kurta A. Heller, Franz J. Mönks and A. Harry Passow, 491–505. New York, NY: Pergamon.
Tsukayama, Eli, Angela Lee Duckworth, and Betty Kim. 2013. "Domain-specific impulsivity in school-age children." *Developmental Science* 16 (6): 879-893.
UNESCO. 2021. "One year into COVID: Prioritizing education recovery to avoid a generational catastrophe." Accessed 10 May. https://en.unesco.org/news/one-year-covid-prioritizing-education-recovery-avoid-generational-catastrophe.
UNICEF. 2020. Students in Bangladesh adjust to remote learning via national TV during COVID-19 lockdown. UNICEF.
Westfall, Peter H, and S Stanley Young. 1993. *Resampling-based multiple testing: Examples and methods for p-value adjustment*. Vol. 279. John Wiley & Sons.
Young, Alwyn. 2019. "Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results." *The Quarterly Journal of Economics* 134 (2): 557-598.

Tables

Table 1. Sample characteristics and balance at endline

Variable	(1) T1: Standard	(2) T2: Extended	(3) C: Control	Pairwise difference		
				(4) T1 vs. T2	(5) T1 vs. C	(6) T2 vs. C
Children's Age in years (as of 01/01/2021)	7.39 (0.04)	7.38 (0.03)	7.38 (0.04)	0.97	0.97	1.00
Gender (1 if Boy)	0.49 (0.02)	0.48 (0.02)	0.49 (0.02)	0.74	0.90	0.84
Baseline literacy score	17.03 (0.33)	16.68 (0.25)	16.87 (0.35)	0.17	0.54	0.47
Baseline numeracy score	14.93 (0.21)	14.70 (0.17)	14.85 (0.17)	0.21	0.65	0.42
Access to private tuition	0.51 (0.03)	0.59 (0.04)	0.57 (0.04)	0.01	0.06	0.53
Father's education (in years of schooling)	6.21 (0.27)	5.84 (0.22)	5.72 (0.24)	0.12	0.05	0.64
Mother's education (in years of schooling)	7.25 (0.24)	6.95 (0.18)	7.00 (0.19)	0.14	0.21	0.83
No. of household member	4.82 (0.13)	4.75 (0.07)	4.91 (0.08)	0.43	0.35	0.06
Family income (in BDT/ month)	10963.24 (297.76)	10852.40 (347.97)	11196.26 (438.98)	0.73	0.46	0.31
Access to TV (1 if yes)	0.46 (0.03)	0.46 (0.04)	0.48 (0.03)	0.87	0.44	0.36
Access to smartphone (1 if yes)	0.34 (0.03)	0.36 (0.02)	0.35 (0.03)	0.36	0.57	0.73
Homestead land (in decimal)	9.13 (0.61)	9.41 (0.81)	11.12 (1.12)	0.72	0.10	0.16
Religion (1 if Islam)	0.80 (0.05)	0.83 (0.04)	0.77 (0.05)	0.12	0.21	0.01
Observations [HHs]	567 [566]	562 [560]	561 [558]	1129 [1126]	1128 [1124]	1123 [1118]
Village	30	30	30	60	60	60

Note: Columns 1-3 report the mean values of the background characteristics of participants in the various treatment groups. Columns 4-6 report the p-values of the respective pairwise t-test. Robust standard errors clustered at the village level are in parentheses. Total number of distinct households is in the squared brackets. *** p<0.005, ** p<0.01, * p<0.05.

Table 2. Treatment effects on key outcomes

Variables	(1)	T1: Standard		T2: Extended	
		(2)	(3)	(4)	(5)
	Control Mean	Treatment effect (Raw)	Treatment effect (SD)	Treatment effect (Raw)	Treatment effect (SD)
Total score	42.75	13.29***	0.60***	13.97***	0.63***
	(1.63)	(1.74)	(0.08)	(1.64)	(0.07)
		{0.00}	{0.00}	{0.00}	{0.00}
		[0.00]	[0.00]	[0.00]	[0.00]
Literacy	25.29	8.10***	0.55***	8.69***	0.59***
	(1.12)	(1.19)	(0.08)	(1.15)	(0.08)
		{0.00}	{0.00}	{0.00}	{0.00}
		[0.00]	[0.00]	[0.00]	[0.00]
Numeracy	17.47	5.19***	0.53***	5.28***	0.54***
	(0.58)	(0.65)	(0.07)	(0.63)	(0.06)
		{0.00}	{0.00}	{0.00}	{0.00}
		[0.00]	[0.00]	[0.00]	[0.00]
Overall impulsivity	2.10	-0.21*	-0.27*	-0.11	-0.15
	(0.06)	(0.08)	(0.11)	(0.08)	(0.11)
		{0.23}	{0.23}	{0.88}	{0.88}
		[0.01]	[0.01]	[0.18]	[0.18]
Schoolwork impulsivity	2.22	-0.25**	-0.30**	-0.15	-0.17
	(0.06)	(0.08)	(0.10)	(0.09)	(0.10)
		{0.09}	{0.09}	{0.75}	{0.75}
		[0.00]	[0.00]	[0.10]	[0.10]
Interpersonal impulsivity	1.99	-0.17	-0.19	-0.08	-0.09
	(0.06)	(0.08)	(0.10)	(0.09)	(0.10)
		{0.57}	{0.57}	{0.98}	{0.98}
		[0.05]	[0.05]	[0.36]	[0.36]
Grit score	3.05	0.10	0.16	0.09	0.16
	(0.04)	(0.06)	(0.09)	(0.06)	(0.10)
		{0.74}	{0.74}	{0.79}	{0.79}
		[0.10]	[0.10]	[0.13]	[0.13]
Growth mindset	3.51	0.19	0.15	0.28*	0.22*
	(0.11)	(0.14)	(0.11)	(0.13)	(0.10)
		{0.88}	{0.88}	{0.45}	{0.45}
		[0.19]	[0.19]	[0.03]	[0.03]
Affective empathy (Contagion)	1.15	-0.02	-0.04	-0.01	-0.04
	(0.02)	(0.03)	(0.07)	(0.03)	(0.07)
		{0.98}	{0.98}	{0.99}	{0.99}
		[0.57]	[0.57]	[0.63]	[0.63]
Cognitive empathy (Understanding)	0.88	-0.03	-0.07	-0.03	-0.05
	(0.03)	(0.03)	(0.07)	(0.03)	(0.07)
		{0.98}	{0.98}	{0.98}	{0.98}
		[0.37]	[0.37]	[0.41]	[0.41]
Prosocial motivation (Support)	1.29	0.02	0.05	0.02	0.05
	(0.02)	(0.03)	(0.08)	(0.03)	(0.07)
		{0.98}	{0.98}	{0.98}	{0.98}
		[0.49]	[0.49]	[0.51]	[0.51]
Leadership	27.99	0.71	0.12	0.83	0.14
	(0.36)	(0.56)	(0.09)	(0.46)	(0.08)
		{0.89}	{0.89}	{0.64}	{0.64}
		[0.20]	[0.20]	[0.07]	[0.07]
Communication	14.82	0.41	0.10	0.11	0.03
	(0.31)	(0.44)	(0.11)	(0.41)	(0.10)
		{0.98}	{0.98}	{0.99}	{0.99}
		[0.37]	[0.37]	[0.80]	[0.80]

Variables	(1) Control Mean	T1: Standard (2) Treatment effect (Raw)	T1: Standard (3) Treatment effect (SD)	T2: Extended (4) Treatment effect (Raw)	T2: Extended (5) Treatment effect (SD)
Planning	50.04	0.75	0.05	0.45	0.03
	(1.05)	(1.63)	(0.12)	(1.39)	(0.10)
		{0.99}	{0.99}	{0.99}	{0.99}
		[0.66]	[0.66]	[0.73]	[0.73]
Emotional symptoms	2.73	-0.32*	-0.17*	-0.25	-0.13
	(0.11)	(0.14)	(0.07)	(0.14)	(0.07)
		{0.34}	{0.34}	{0.64}	{0.64}
		[0.02]	[0.02]	[0.07]	[0.07]
Conduct problem	2.67	-0.41**	-0.21**	-0.25	-0.13
	(0.12)	(0.14)	(0.07)	(0.15)	(0.08)
		{0.12}	{0.12}	{0.75}	{0.75}
		[0.01]	[0.01]	[0.10]	[0.10]
Hyperactivity	4.40	-0.42**	-0.20**	-0.04	-0.02
	(0.11)	(0.14)	(0.07)	(0.16)	(0.07)
		{0.10}	{0.10}	{0.99}	{0.99}
		[0.00]	[0.00]	[0.82]	[0.82]
Peer problem	2.78	-0.23*	-0.17*	-0.12	-0.09
	(0.07)	(0.09)	(0.07)	(0.10)	(0.07)
		{0.23}	{0.23}	{0.89}	{0.89}
		[0.01]	[0.01]	[0.21]	[0.21]
Prosocial	7.06	0.50**	0.24**	0.25	0.12
	(0.12)	(0.16)	(0.08)	(0.15)	(0.08)
		{0.06}	{0.06}	{0.76}	{0.76}
		[0.00]	[0.00]	[0.11]	[0.11]

Note: This table presents the treatment effects on raw and standardized outcome variables. Column 1 displays the mean of the control group. Columns 2 and 4 show the treatment effects on the raw outcome variables for the Standard and Extended groups, respectively. Columns 3 and 5 present the treatment effects on the standardized outcome variables for the Standard and Extended groups, respectively. Outcome variables are standardized [(y_i – mean of the control group)/standard deviation of control group]. Coefficients are estimated using OLS regressions. The total number of observations is 1690. Baseline controls include children's age, gender, baseline literacy score, baseline numeracy score, access to private tuition, parents' education in years, family income, religion, access to TV & smartphone, homestead size, and the number of members in the household. Children's grade fixed effects are included in all regressions. Robust standard errors, clustered at the village level, are in parentheses. Westfall-Young FWER adjusted p-values are in the curly brackets and Randomized Inference (RI) P-values are in the squared brackets. These p-values are calculated based on 5,000 replications. *** p<0.005, ** p<0.01, * p<0.05.

Table 3. Heterogenous treatment effects on learning and homeschooling

Variables	W: Gender			X: Baseline score			Y: Household income			Z: Parental education		
	(1) Boy	(2) Girl	(3) Inter-action	(4) Above median	(5) Below median	(6) Inter-action	(7) Above median	(8) Below median	(9) Inter-action	(10) Above median	(11) Below median	(12) Inter-action
Panel A: Standard treatment												
Total score	0.55***	0.65***	-0.09	0.48***	0.75***	-0.25*	0.57***	0.65***	-0.08	0.53***	0.75***	-0.21*
	(0.09)	(0.10)	(0.10)	(0.10)	(0.09)	(0.11)	(0.08)	(0.11)	(0.11)	(0.08)	(0.11)	(0.10)
Literacy	0.54***	0.57***	-0.02	0.48***	0.65***	-0.17	0.50***	0.64***	-0.13	0.51***	0.64***	-0.13
	(0.10)	(0.10)	(0.10)	(0.11)	(0.09)	(0.12)	(0.09)	(0.11)	(0.11)	(0.08)	(0.12)	(0.11)
Numeracy	0.43***	0.62***	-0.18	0.38***	0.73***	-0.32*	0.54***	0.51***	0.02	0.42***	0.72***	-0.29**
	(0.09)	(0.08)	(0.10)	(0.08)	(0.09)	(0.10)	(0.08)	(0.09)	(0.11)	(0.07)	(0.09)	(0.09)
Student's study time	0.10	0.11	-0.02	0.13	0.04	0.08	0.04	0.18	-0.14	0.11	0.13	0.00
	(0.11)	(0.12)	(0.14)	(0.11)	(0.10)	(0.12)	(0.10)	(0.12)	(0.13)	(0.11)	(0.14)	(0.15)
Caregiver's time in homeschooling	0.22*	0.19	0.04	0.24*	0.16	0.08	0.21*	0.17	0.02	0.35***	0.03	0.30*
	(0.11)	(0.11)	(0.13)	(0.11)	(0.10)	(0.11)	(0.11)	(0.11)	(0.12)	(0.10)	(0.14)	(0.14)
Observations	551	577	1128	612	516	1128	623	505	1128	667	461	1128
Panel B: Extended treatment												
Total score	0.61***	0.69***	-0.09	0.60***	0.69***	-0.10	0.52***	0.78***	-0.26*	0.56***	0.77***	-0.22*
	(0.08)	(0.09)	(0.10)	(0.09)	(0.10)	(0.10)	(0.07)	(0.10)	(0.10)	(0.08)	(0.10)	(0.10)
Literacy	0.60***	0.63***	-0.03	0.61***	0.61***	-0.00	0.46***	0.76***	-0.29**	0.53***	0.72***	-0.20
	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)	(0.11)	(0.08)	(0.11)	(0.10)	(0.09)	(0.10)	(0.11)
Numeracy	0.48***	0.62***	-0.15	0.44***	0.65***	-0.22*	0.49***	0.62***	-0.14	0.47***	0.66***	-0.19
	(0.08)	(0.08)	(0.10)	(0.08)	(0.09)	(0.09)	(0.07)	(0.09)	(0.10)	(0.07)	(0.09)	(0.10)
Student's study time	0.14	0.02	0.11	0.04	0.12	-0.07	0.01	0.16	-0.16	0.12	0.03	0.12
	(0.12)	(0.10)	(0.13)	(0.11)	(0.10)	(0.11)	(0.11)	(0.11)	(0.12)	(0.10)	(0.11)	(0.12)
Caregiver's time in homeschooling	0.03	0.12	-0.05	0.15	-0.00	0.13	0.07	0.08	-0.03	0.07	0.08	0.00
	(0.09)	(0.10)	(0.13)	(0.09)	(0.10)	(0.10)	(0.09)	(0.11)	(0.11)	(0.09)	(0.11)	(0.11)
Observations	543	580	1123	576	547	1123	597	526	1123	635	488	1123

Note: This table presents the heterogeneous treatment effects of the intervention on the learning outcomes and homeschooling variables. Coefficients are estimated using OLS regressions. The dependent variable for each regression is listed in the first row. The same list of control variables is used as before. Children's grade fixed effects are included in all regressions. Boy = dummy variable for boy participant; above-median = dummy (1 if the corresponding value is above the median); interaction = interaction term between treatment and gender or above median variable. Robust standard errors clustered at the village level are in parentheses. *** p<0.005, ** p<0.01, * p<0.05.

Table 4. Treatment effect on homeschooling outcomes

Variables	(1)	T1: Standard		T2: Extended	
		(2)	(3)	(4)	(5)
	Control Mean	Treatment effect (Raw)	Treatment effect (SD)	Treatment effect (Raw)	Treatment effect (SD)
Student's study time	87.03	5.42	0.10	3.72	0.07
	(3.70)	(4.83)	(0.09)	(4.66)	(0.09)
		{0.58}	{0.58}	{0.61}	{0.61}
		[0.27]	[0.27]	[0.43]	[0.43]
Extent of study	2.65	0.18*	0.20*	0.21**	0.24**
	(0.04)	(0.07)	(0.08)	(0.07)	(0.08)
		{0.07}	{0.07}	{0.02}	{0.02}
		[0.01]	[0.01]	[0.00]	[0.00]
Caregiver's time in homeschooling	75.48	9.72*	0.20*	3.38	0.07
	(3.41)	(4.33)	(0.09)	(3.93)	(0.08)
		{0.12}	{0.12}	{0.61}	{0.61}
		[0.02]	[0.02]	[0.40]	[0.40]
Extent of caregiver's homeschooling	2.63	0.21**	0.23**	0.12	0.13
	(0.07)	(0.08)	(0.09)	(0.08)	(0.08)
		{0.04}	{0.04}	{0.36}	{0.36}
		[0.00]	[0.00]	[0.11]	[0.11]

Note: This table presents the treatment effects on raw and standardized homeschooling variables i.e., Student's study time – daily study time in minutes provided by the children; Extent of study – a 5-point Likert-scale response; 'none' to 'a great deal'; Caregiver's time in homeschooling – daily homeschooling time provided by the caregivers in minutes; Extent of caregiver's homeschooling – a 5-point Likert-scale response; 'none' to 'a great deal'. Column 1 shows the mean of control group. Columns 2 and 4 show the treatment effects on the raw homeschooling variables for the Standard and Extended groups, respectively. Columns 3 and 5 present the treatment effects on the standardized homeschooling variables for the Standard and Extended groups, respectively. Outcome variables are standardized [(y_i – mean of the control group)/standard deviation of control group]. Coefficients are estimated using OLS regressions. The total number of observations is 1690. Baseline controls include children's age, gender, baseline literacy score, baseline numeracy score, access to private tuition, parents' education in years, family income, religion, access to TV & smartphone, homestead size, and the number of members in the household. Children's grade fixed effects are included in all regressions. Robust standard errors, clustered at the village level, are in parentheses. Westfall-Young FWER adjusted p-values are in the curly brackets and Randomized Inference (RI) P-values are in the squared brackets. These p-values are calculated based on 5,000 replications. *** p<0.005, ** p<0.01, * p<0.05.

Table 5. Relationships between types of lesson completion and baseline characteristics

Variables	X: Following sequence without skipping		Y: Total number of lessons completed		Z: Percent of completed lessons ever repeated	
	(1) Literacy	(2) Numeracy	(3) Literacy	(4) Numeracy	(5) Literacy	(6) Numeracy
Less-educated parents	0.07*	0.03	0.51	0.32	-0.14	-0.34
	(0.03)	(0.04)	(0.85)	(0.89)	(0.49)	(0.46)
Lower baseline score	-0.05	0.08*	-1.89**	-1.38	-0.91*	-0.48
	(0.04)	(0.03)	(0.67)	(0.73)	(0.44)	(0.40)
Lower household income	0.04	0.08	0.46	0.08	0.27	0.24
	(0.04)	(0.04)	(0.72)	(0.71)	(0.53)	(0.37)
Observations	1012	1003	1012	1003	1012	1003
R-squared	0.03	0.03	0.02	0.02	0.012	0.012

Note: This table presents the associations between different types of lesson completion and three key baseline characteristics among the treated individuals with information on the exact number of lessons completed. Each regression specification also controls for children's age, gender, access to private tuition, religion, access to TV & smartphone, homestead size, and the number of members in the household. Less-educated parents takes the value of one if combined parental education is below the median. Lower baseline (literacy or numeracy) test score takes the value of one if it is below the median. Lower household income takes the value of one if it is below the median. Following sequence without skipping takes the value of one if lessons were completed in a sequence without skipping any lessons in the sequence. Total number of lessons completed is the total number of lessons completed. Percent of completed lessons ever repeated is the share of completed lessons that were ever repeated. Columns 5 and 6 include individuals whose registered phone numbers cannot be matched to the IVR access data and their numbers of lessons repeated are coded as zero; the statistical significance of each key explanatory variable is similar if we include a control for them or exclude them from the sample (see Table B9 in Appendix B). *** $p<0.005$, ** $p<0.01$, * $p<0.05$.