

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

Liang Choon Wang[†], Michael Vlassopoulos[‡], Asad Islam[§], Hashibul Hassan^{**}

November 2023

Abstract

The Covid-19 pandemic caused prolonged school closures worldwide. Children in resource-poor settings were particularly affected due to their limited access to remedial distance learning opportunities through the internet, television, and radio. To address this poor access to formal education, we designed an educational intervention consisting of a set of audio lessons that were delivered via mobile phones to primary school students using Interactive Voice Response (IVR) technology. During the 15-week program period, parents had free access to these lessons by dialing a designated phone number and listening to a lesson with their child at any time. We evaluate the impact of our randomized intervention on 1,763 primary school children across 90 villages in Bangladesh during the Covid-19 school closures in 2021. Our findings show that the intervention significantly improved the literacy and numeracy test scores of participating children by 0.60 Standard Deviations (SD). Furthermore, the intervention led to an increase in the amount of time that parents devoted to homeschooling. The intervention was particularly beneficial for academically weaker students and those with less-educated caregivers. Our results highlight the potential of this scalable and low-cost intervention to address the 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 reviewers and editor, as well as 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.

[†] Department of Economics, Monash University, Australia.

[‡] Department of Economics, University of Southampton, United Kingdom; and IZA Institute of Labor Economics, Germany.

[§] Centre for Development Economics and Sustainability (CDES) and Department of Economics, Monash University, Australia. Email: asadul.islam@Monash.edu (Corresponding author).

^{**} Department of Finance, Jagannath University, Bangladesh.

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, Doepke, Sorrenti, & Zilibotti, 2022; Engzell, Frey, & Verhagen, 2021; Moscoviz & Evans, 2022; Patrinos, 2022; Singh, Romero, & 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 (Rahman & Sharma, 2021; UNICEF, 2020). 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 (Beam, Mukherjee, Navarro-Sola, Ferdosh, & Sarwar, 2021; Hassan, Islam, Siddique, & Wang, 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 & Bosch, 2005; Ho & 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 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 & 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 (Billsberry, Vega, & Molineux, 2019; Murphy & Johnson, 2011).² Despite its potential importance, there has been limited research evaluating interventions aimed at enhancing leadership skills among children through randomized trials (Gutman & Schoon, 2013). To address this gap, we offered an additional leadership skill module in 15 audio lessons, adapted

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.

² Some studies have shown that leadership development activities in schools are associated with measures of children’s leadership trait (Ayman-Nolley & Ayman, 2005; Billsberry et al., 2019; Salmond & Fleshman, 2010).

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 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

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.

recommended sequence of lessons that progressively built up the difficulties of learning materials, differences in how caregivers accessed the lessons potentially explain the heterogeneous 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, & Matsheng, 2022](#); [Crawford, Evans, Hares, & Sandefur, 2023](#); [Hassan et al., 2021](#); [Lichand & Christen, 2021](#); [Schueler & Rodriguez-Segura, 2023](#)).⁴ One challenge facing these programs is that it can be

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 & Chan, 2021](#); [Berlinski, Busso, Dinkelman, & Martinez, 2016](#); [Hurwitz, Lauricella, Hanson, Raden, & Wartella, 2015](#); [Kraft & Dougherty, 2013](#); [Mayer,](#)

difficult to scale them up, especially in a low-income country context, as paid instructors are costly and volunteers are difficult to retain (Islam, Malek, Tasneem, & Wang, 2022). Importantly, our program distinguishes itself by offering flexible study hours, a feature absent 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, Halim, Lucas, & Saif, 2022; Bianchi, Lu, & Song, 2022; Cardim, Molina-Millán, & Vicente, 2023; Johnston & Ksoll, 2022; Muralidharan, Singh, & 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 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 & 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

Kalil, Oreopoulos, & Gallegos, 2015), or high school graduates (Bird et al., 2021; Castleman & Meyer, 2020), or adult learners (Aker & Ksoll, 2019; Ksoll, Aker, Miller, Perez, & Smalley, 2015), rather than primary graders.

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.

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

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.

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, & Kariuki, 2002). Facilitators play an important role in supervising the progression of the lessons (Ho & 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 Figure 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 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, Cole, Duflo, & Linden, 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

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.

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.

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).

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.

3.3 Outcomes

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

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 et al. \(2002\)](#). 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, & Kim, 2013](#)). Second, we measured the grit of the participants using an 8-item grit scale developed by [Angela Lee Duckworth and Quinn \(2009\)](#). Grit is defined as “perseverance and passion for long-term goals” ([Rimfeld, Kovas, Dale, & Plomin, 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 & 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 ([Angela Duckworth, 2016](#); [AL Duckworth & Eskreis-Winkler, 2013](#); [Angela L Duckworth, Peterson, Matthews, & Kelly, 2007](#); [Eskreis-Winkler, Shulman, Beal, &](#)

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.

Duckworth, 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, Rieffe, Broekhof, Crone, and Güroğlu (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) Goodman (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:

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

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 & 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 30th percentile of the treatment group distribution corresponds to the 60th 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 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.

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).

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 & 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 [Table 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 heterogeneous 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, & Roldán, 2022), a five-week volunteer-based online tutoring (N=1,059; 17 hours) in Italy (Carlana & La Ferrara, 2021), and a 12-week pilot program of online tutoring (N=560; 4 hours) by college-volunteers in the US (Kraft, List, Livingston, & Sadoff, 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 et al., 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 (Crawford 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

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.

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 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 et al. \(2022\)](#) 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 & 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 et al., 2022; Lichand & Christen, 2021; Muralidharan & 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

- Agnafors, S., Barmark, M., & Sydsjö, G. (2021). Mental health and academic performance: a study on selection and causation effects from childhood to early adulthood. *Social psychiatry and psychiatric epidemiology*, *56*, 857-866.
- Agostinelli, F., Doepke, M., Sorrenti, G., & Zilibotti, F. (2022). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of public economics*, *206*, 104574.
- Aker, J. C., & Ksoll, C. (2019). Call me educated: Evidence from a mobile phone experiment in Niger☆. *Economics of Education Review*, *72*, 239-257.
- Alan, S., Boneva, T., & Ertac, S. (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, N., Bergman, P., & Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour*. doi:10.1038/s41562-022-01381-z
- Anzalone, S., & Bosch, A. (2005). Improving educational quality with interactive radio instruction: a toolkit for policymakers and planners.
- Ayman-Nolley, S., & Ayman, R. (2005). Children's implicit theory of leadership. *Implicit leadership theories: Essays and explorations*, 227-275.
- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The quarterly journal of economics*, *122*(3), 1235-1264.
- Beam, E. A., Mukherjee, P., Navarro-Sola, L., Ferdosh, J., & Sarwar, M. A. H. (2021). Take-Up, Use, and Effectiveness of Remote Technologies. In *Working Paper Series*. Washington, DC: The World Bank.
- Beg, S., Halim, W., Lucas, A. M., & Saif, U. (2022). Engaging teachers with technology increased achievement, bypassing teachers did not. *American Economic Journal: Economic Policy*, *14*(2), 61-90.
- Bergman, P., & Chan, E. W. (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, S., Busso, M., Dinkelman, T., & Martinez, C. (2016). Reducing parent-school information gaps and improving education outcomes: Evidence from high frequency text messaging in Chile. *Unpublished manuscript*.
- Bianchi, N., Lu, Y., & Song, H. (2022). The effect of computer-assisted learning on students' long-term development. *Journal of development economics*, *158*, 102919.
- Billsberry, J., Vega, C. E., & Molineux, J. (2019). Think of the children: Leader development at the edge of tomorrow. *Journal of Management & Organization*, *25*(3), 378-381.
- Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lambertson, C., & Rosinger, K. O. (2021). Nudging at scale: Experimental evidence from FAFSA completion campaigns. *Journal of Economic Behavior & Organization*, *183*, 105-128.
- Biswas, K., Asaduzzaman, T., Evans, D. K., Fehrler, S., Ramachandran, D., & Sabarwal, S. (2020). TV-Based Learning in Bangladesh.
- Bosch, A. (1997). Interactive Radio Instruction: Twenty-three years of improving education quality. *Washington, DC: World Bank Group*.
- Bosch, A., Rhodes, R., & Kariuki, S. (2002). *Interactive Radio Instruction: an Update from the Field*. Retrieved from
- Cardim, J., Molina-Millán, T., & Vicente, P. C. (2023). Can technology improve the classroom experience in primary education? An African experiment on a worldwide program. *Journal of development economics*, *164*, 103145.
- Carlana, M., & La Ferrara, E. (2021). *Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic*. Retrieved from
- Castleman, B. L., & Meyer, K. E. (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.

- Chen-Bouck, L., Patterson, M. M., Qiao, B., & Peng, A. (2023). Evaluation of the Effectiveness of an Empathy Training on Empathy Skills, Life Satisfaction, and Relationship Quality for Chinese Adolescents and Their Mothers: A Mixed Methods Study. *Journal of Adolescent Research*, 38(4), 591-631. doi:10.1177/07435584211064209
- Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences*, 113(31), 8664-8668.
- Crawford, L., Evans, D. K., Hares, S., & Sandefur, J. (2023). Live tutoring calls did not improve learning during the COVID-19 pandemic in Sierra Leone. *Journal of development economics*, 164, 103114. doi:<https://doi.org/10.1016/j.jdeveco.2023.103114>
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, 24(4), 349.
- DataReportal. (2021). Digital 2021: Bangladesh. Retrieved from <https://datareportal.com/reports/digital-2021-bangladesh>
- Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review*, 109(8), 2728-2765.
- Dobronyi, C. R., Oreopoulos, P., & Petronijevic, U. (2019). Goal setting, academic reminders, and college success: A large-scale field experiment. *Journal of Research on Educational Effectiveness*, 12(1), 38-66.
- Duckworth, A. (2016). *Grit: The power of passion and perseverance* (Vol. 234): Scribner New York.
- Duckworth, A., & Eskreis-Winkler, L. (2013). True grit. *Observer*, 26(4). Retrieved from <https://www.psychologicalscience.org/observer/true-grit>
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of personality assessment*, 91(2), 166-174.
- Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-control and academic achievement. *Annual review of psychology*, 70, 373-399.
- Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development*: Psychology press.
- Dweck, C. S., Chiu, C.-y., & Hong, Y.-y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological inquiry*, 6(4), 267-285.
- Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., . . . Gawron, P. (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, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17), e2022376118.
- Eskreis-Winkler, L., Shulman, E. P., Beal, S. A., & Duckworth, A. L. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in psychology*, 5, 36.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, 38(5), 581-586.
- Gortazar, L., Hupkau, C., & Roldán, A. (2022). Online tutoring works: Experimental evidence from a program with vulnerable children. *Working Paper*.
- Guo, L., Shi, G., Du, X., Wang, W., Guo, Y., & Lu, C. (2021). Associations of emotional and behavioral problems with Internet use among Chinese young adults: the role of academic performance. *Journal of Affective Disorders*, 287, 214-221. doi:<https://doi.org/10.1016/j.jad.2021.03.050>
- Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. A literature review.
- Hassan, H., Islam, A., Siddique, A., & Wang, L. C. (2021). Telementoring and homeschooling during school closures: A randomized experiment in rural Bangladesh.

- Ho, J., & Thukral, H. (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, L. B., Lauricella, A. R., Hanson, A., Raden, A., & Wartella, E. (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* [sample survey data (SSD)]. Retrieved from: <https://doi.org/10.7910/DVN/NXKLZJ>
- Islam, A., Malek, A., Tasneem, S., & Wang, L. C. (2022). Can public recognition reward backfire? Field experimental evidence on the retention and performance of volunteers.
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives.
- Johnston, J., & Ksoll, C. (2022). Effectiveness of interactive satellite-transmitted instruction: Experimental evidence from Ghanaian primary schools. *Economics of Education Review*, 91, 102315.
- Karagianni, D., & Jude Montgomery, A. (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, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1), 83-119.
- Kraft, M. A., & Dougherty, S. M. (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, M. A., List, J. A., Livingston, J. A., & Sadoff, S. (2022). *Online tutoring by college volunteers: Experimental evidence from a pilot program*. Paper presented at the AEA Papers and Proceedings.
- Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130), 297-300.
- Ksoll, C., Aker, J. C., Miller, D., Perez, K., & Smalley, S. L. (2015). *Learning without Teachers? Evidence from a Randomized Experiment of a Mobile Phone-Based Adult Education Program in Los Angeles*. Retrieved from
- Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., . . . Vakis, R. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. *Science advances*, 5(7), eaaw5226.
- Lam, K. K. L., & Zhou, M. (2019). Examining the relationship between grit and academic achievement within K-12 and higher education: A systematic review. *Psychology in the Schools*, 56(10), 1654-1686.
- LEAD. (2021). LEAD Learning. 2021(15 January). Retrieved from <https://leadedu.org/learning>
- Lichand, G., & Christen, J. (2021). *Behavioral nudges prevent student dropouts in the pandemic*. Retrieved from
- Mayer, S. E., Kalil, A., Oreopoulos, P., & Gallegos, S. (2015). *Using behavioral insights to increase parental engagement: The parents and children together (PACT) intervention*. Retrieved from
- Miller, P. H., Baxter, S. D., Royer, J. A., Hitchcock, D. B., Smith, A. F., Collins, K. L., . . . Vaadi, K. K. (2015). Children's social desirability: Effects of test assessment mode. *Personality and individual differences*, 83, 85-90.
- Moscoviz, L., & Evans, D. K. (2022). Learning loss and student dropouts during the covid-19 pandemic: A review of the evidence two years after schools shut down. In *Center for Global Development, Working Paper* (Vol. 609).
- Muralidharan, K., & Singh, A. (2021). India's new National Education Policy: Evidence and challenges. *Science*, 372(6537), 36-38. doi:10.1126/science.abf6655
- Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. *American Economic Review*, 109(4), 1426-1460. doi:10.1257/aer.20171112
- Murphy, S. E., & Johnson, S. K. (2011). The benefits of a long-lens approach to leader development: Understanding the seeds of leadership. *The Leadership Quarterly*, 22(3), 459-470.

- Overgaauw, S., Rieffe, C., Broekhof, E., Crone, E. A., & Güroğlu, B. (2017). Assessing empathy across childhood and adolescence: Validation of the Empathy Questionnaire for Children and Adolescents (EmQue-CA). *Frontiers in psychology*, 8, 870.
- Patrinos, H. A. (2022). Learning loss and learning recovery. *DECISION*, 1-6.
- Rahman, H. Z., Matin, I., Rahman, A., Das, N. C., Zillur, U., Ahmed, M. S., . . . Shameem, N. (2021). PPRC-BIGD COVID-19 Livelihoods & Recovery Panel Survey. Retrieved from <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, T., & Sharma, U. (2021). A Simulation of COVID-19 School Closure Impact on Student Learning in Bangladesh. *World Bank*.
- Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., & Westberg, K. L. (2002). *Scales for rating the behavioral characteristics of superior students. Technical and administration manual*: ERIC.
- Reynolds, W. 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, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, 111(5), 780.
- Salmond, K., & Fleshman, P. (2010). Gender and Women's Leadership: A Reference Handbook. In K. O'Connor (Ed.). doi:10.4135/9781412979344
- Schueler, B. E., & Rodriguez-Segura, D. (2023). A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya. *Journal of Research on Educational Effectiveness*, 16(3), 442-472. doi:10.1080/19345747.2022.2131661
- Singh, A., Romero, M., & Muralidharan, K. (2022). COVID-19 Learning Loss and Recovery: Panel Data Evidence from India. In: National Bureau of Economic Research.
- Sisk, D. A. (1993). Leadership education for the gifted. In K. A. Heller, F. J. Mönks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 491–505). New York, NY: Pergamon.
- Tsukayama, E., Duckworth, A. L., & Kim, B. (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. Retrieved from <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. In.
- Westfall, P. H., & Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment* (Vol. 279): John Wiley & Sons.
- Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The quarterly journal of economics*, 134(2), 557-598.

Figures and Tables

Figure 1. Summary of Research Design

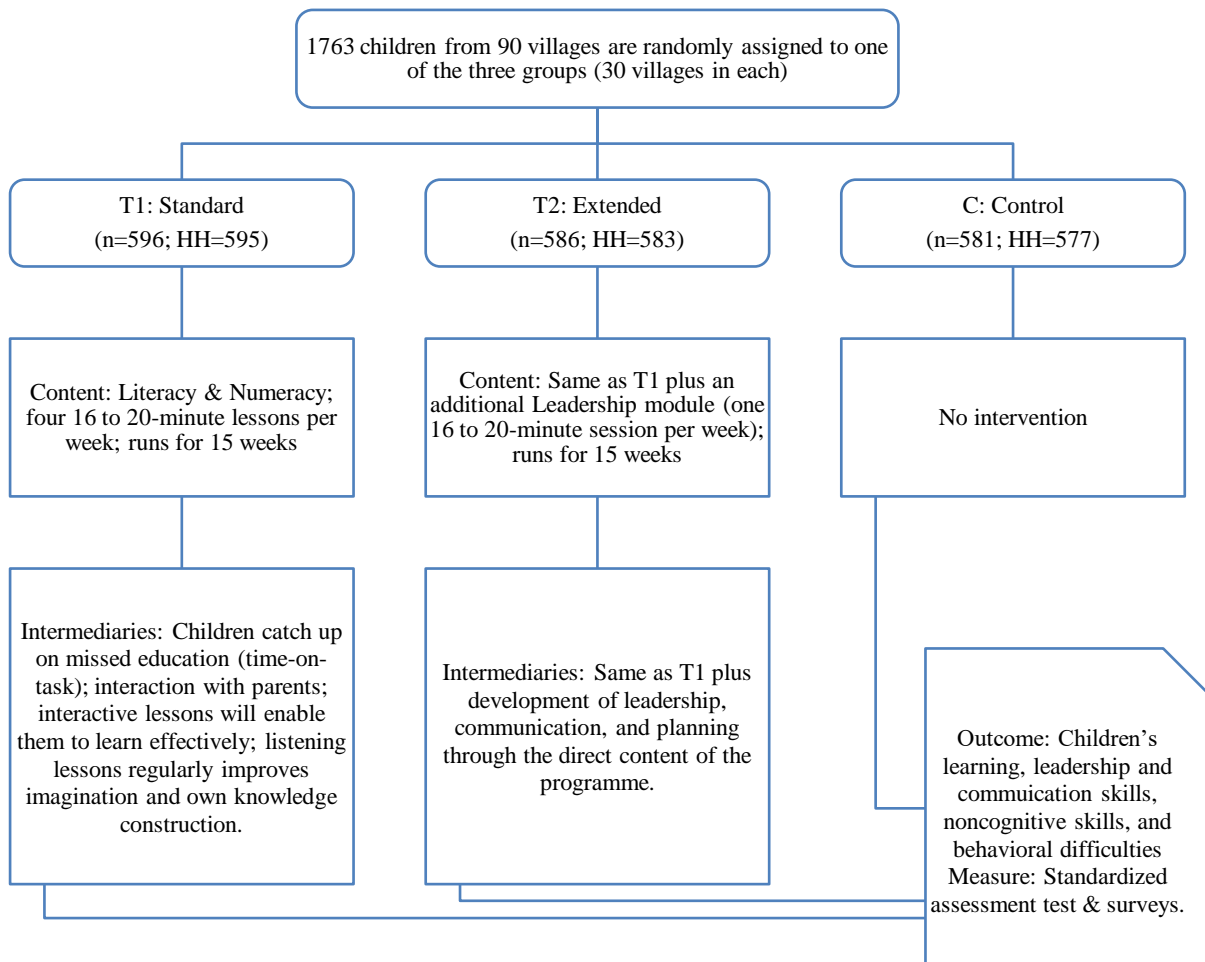
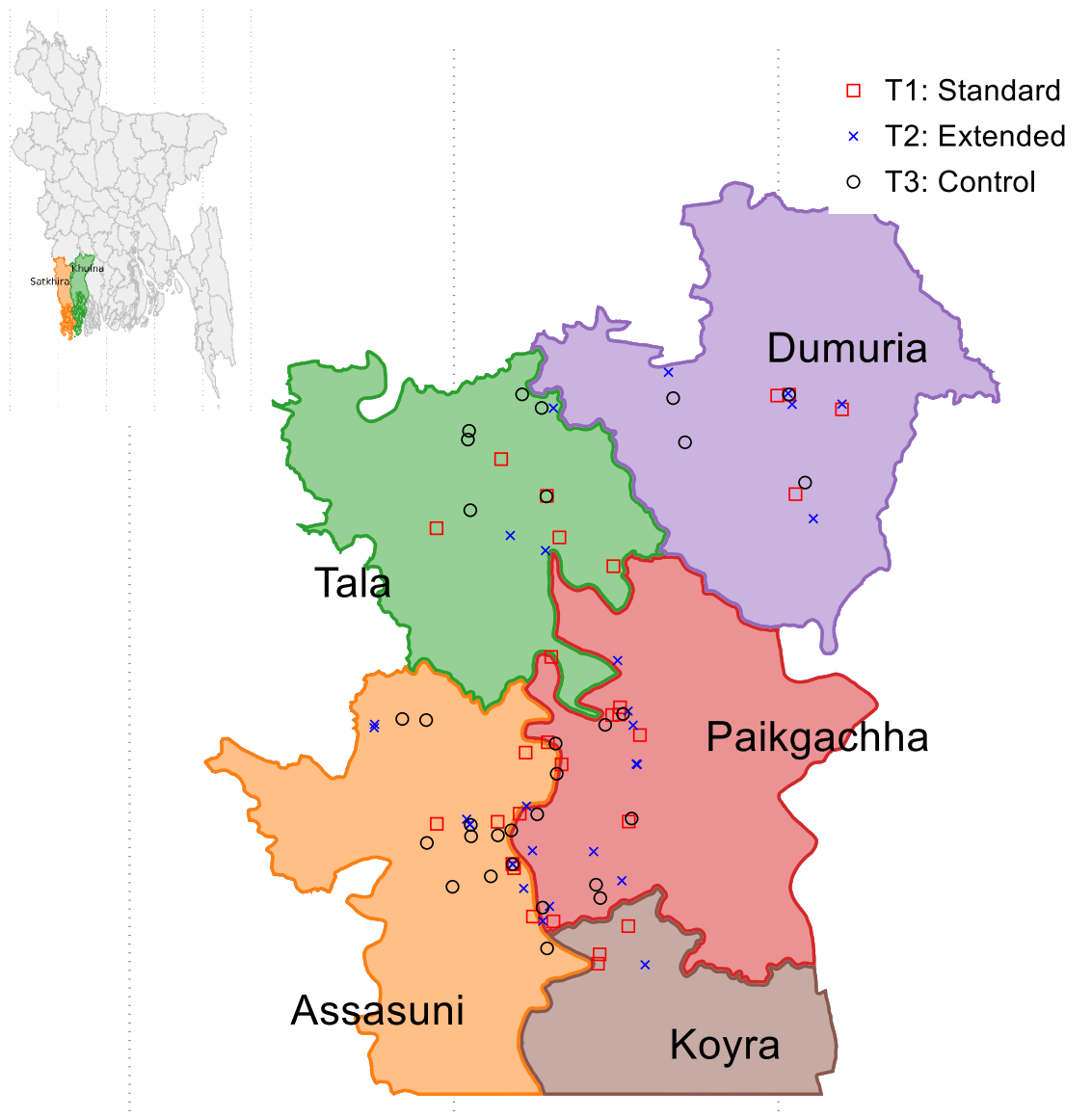


Figure 2. Bangladesh map, study area and treatment villages

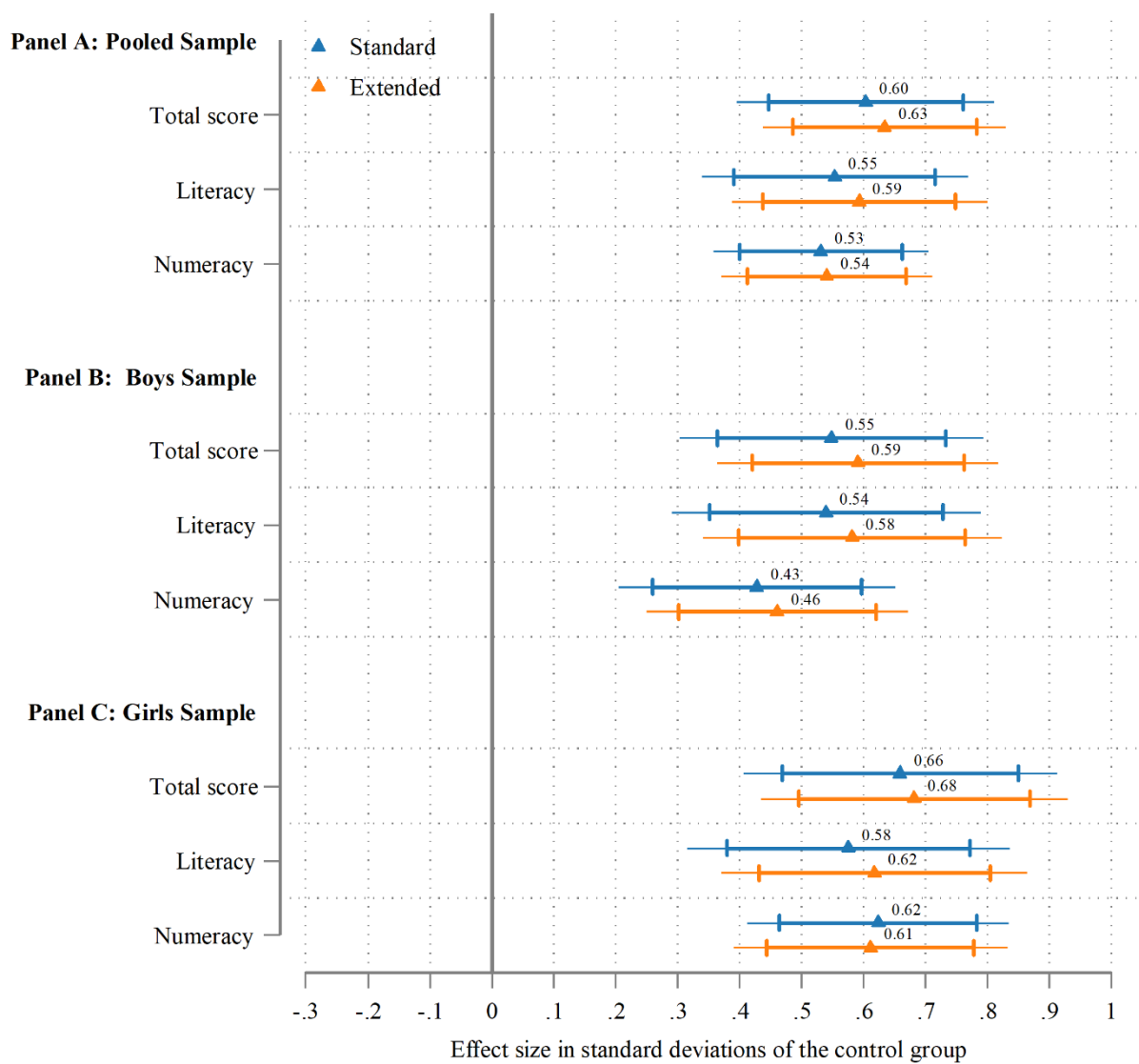


Note: This figure illustrates the study district and subdistricts of Bangladesh. Markers indicate the villages (clusters) involved in this intervention.

Figure 3. Project Timeline

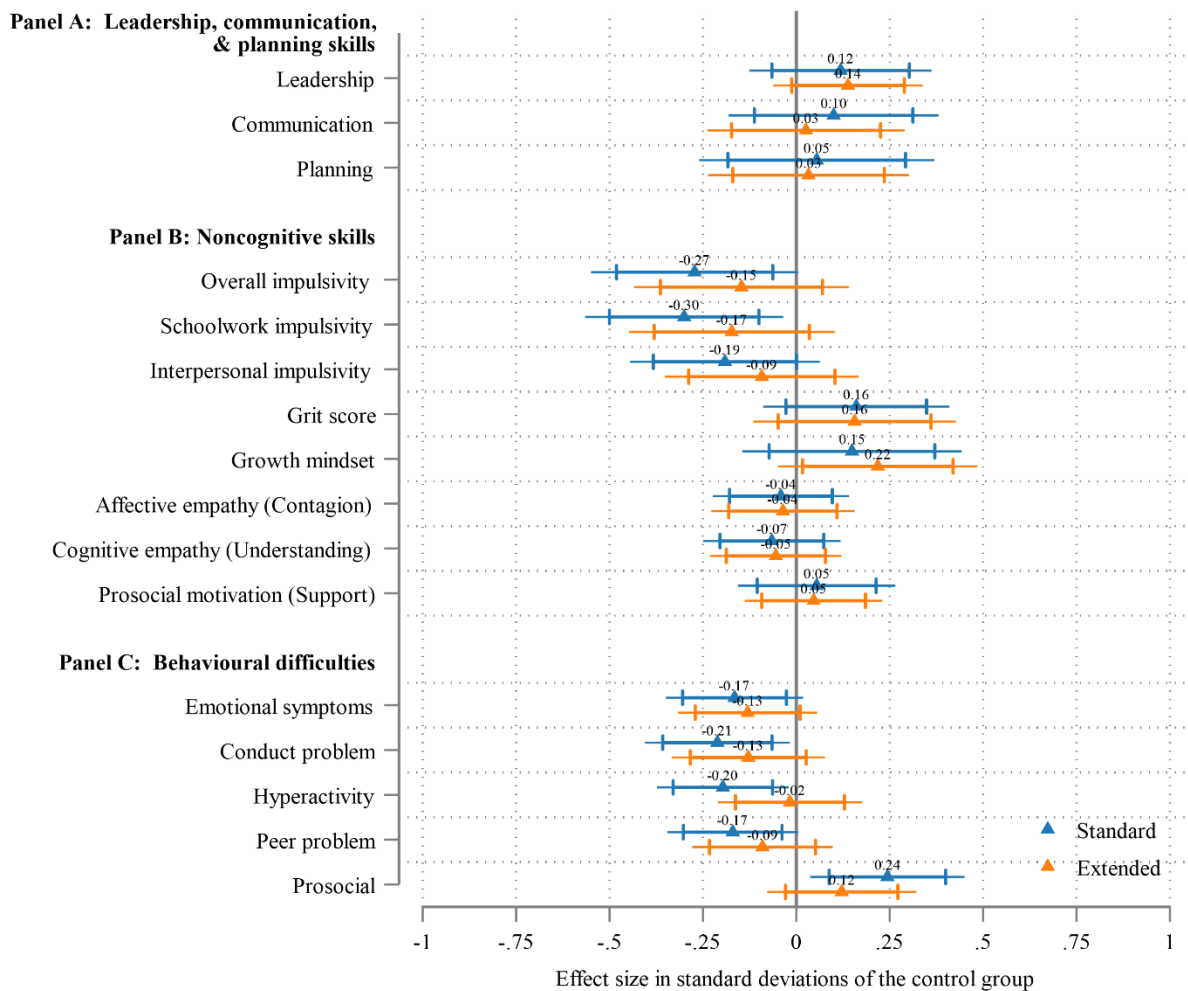
Preparation ↓	Baseline ↓	Random assignment ↓	Intervention start ↓	Intervention end ↓	Endline ↓
<i>Mar to May 21</i>	<i>May 21</i>	<i>Jun 21</i>	<i>Jun 21</i>	<i>Oct 21</i>	<i>Nov 21</i>
Content design; lesson recordings; IVR flow design & optimization; field test.	Baseline survey	Randomization; mothers' briefing	15-week intervention period		Children's assessment test; children's survey; parent's survey

Figure 4. Treatment effects on the standardized learning outcomes



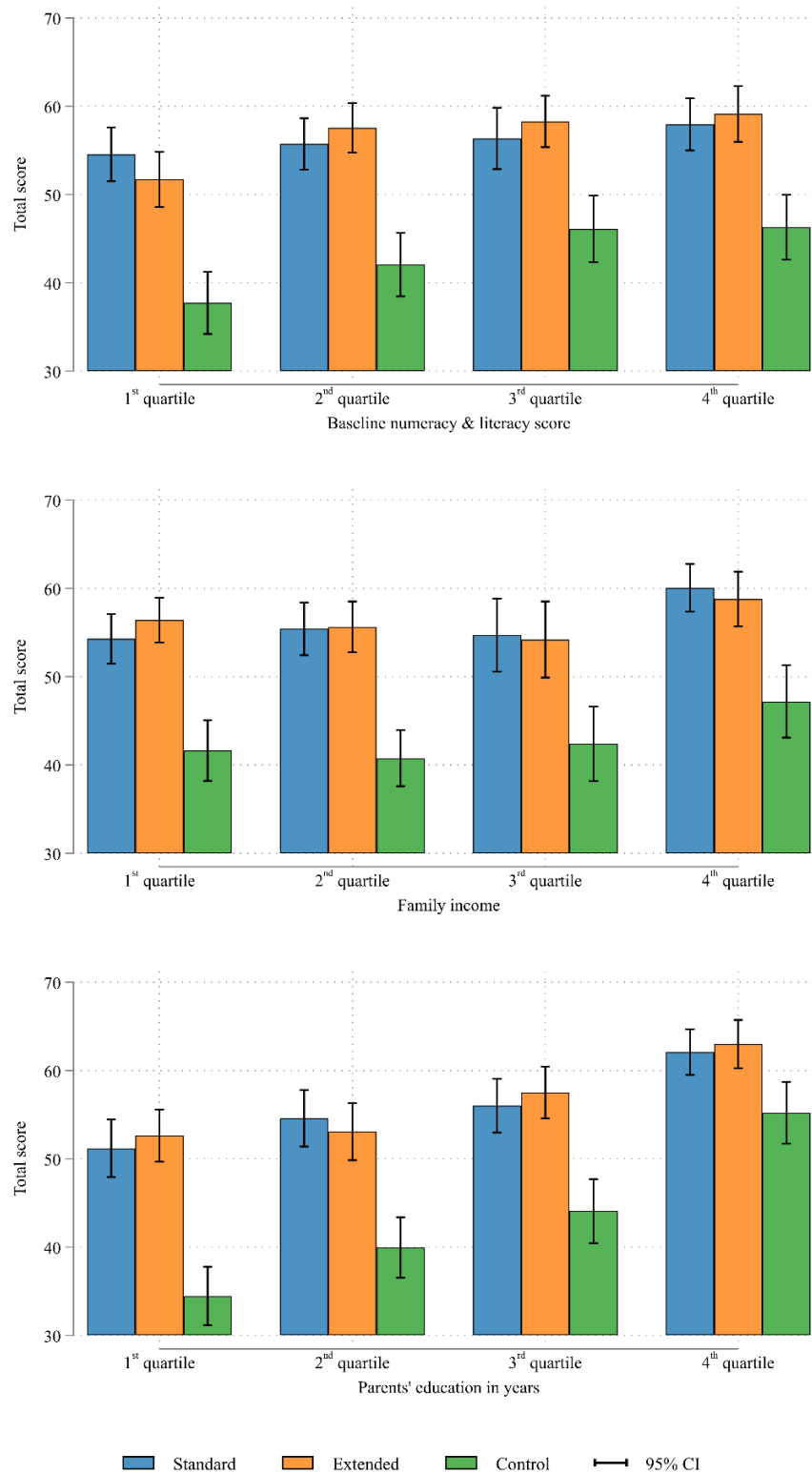
Note: All outcomes are standardized $[(y_i - \text{mean of the control group}) / \text{standard deviation of control group}]$ with the control group having a mean of 0 and SD of 1. This figure shows where the mean of the treatment groups lies in the distribution of the control group in standard deviation (SD) units, along with 95 and 99 confidence intervals. Coefficients are estimated using OLS regressions. Baseline controls include children's age, 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 are clustered at the village level.

Figure 5. Treatment effects on secondary outcomes



Note: All outcomes are standardized $[(y_i - \text{mean of the control group}) / \text{standard deviation of control group}]$ with the control group having a mean of 0 and SD of 1. This figure shows where the mean of the treatment groups lies in the distribution of the control group in standard deviation (SD) units, along with 95 and 99 confidence intervals. Coefficients are estimated using OLS regressions. Baseline controls include children's age, 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 are clustered at the village level.

Figure 6. Heterogeneity in assessment test performance, by baseline numeracy & literacy score, family income, parents' education quartiles



Note: This figure displays average test performance (and 95% CI) for the two treatment groups and the control group, by quartiles of baseline numeracy & literacy score, family income, and parents' education for all children who completed the endline assessment tests. The standard errors used to construct the 95% CI are uncorrected for multiple hypothesis testing.

Table 1. Sample characteristics and balance at endline

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

Variables	T1: Standard			T2: Extended	
	(1) Control Mean	(2) Treatment effect (Raw)	(3) Treatment effect (SD)	(4) Treatment effect (Raw)	(5) Treatment effect (SD)
Planning	50.04 (1.05)	0.75 (1.63) {0.99} [0.66]	0.05 (0.12) {0.99} [0.66]	0.45 (1.39) {0.99} [0.73]	0.03 (0.10) {0.99} [0.73]
Emotional symptoms	2.73 (0.11)	-0.32* (0.14) {0.34} [0.02]	-0.17* (0.07) {0.34} [0.02]	-0.25 (0.14) {0.64} [0.07]	-0.13 (0.07) {0.64} [0.07]
Conduct problem	2.67 (0.12)	-0.41** (0.14) {0.12} [0.01]	-0.21** (0.07) {0.12} [0.01]	-0.25 (0.15) {0.75} [0.10]	-0.13 (0.08) {0.75} [0.10]
Hyperactivity	4.40 (0.11)	-0.42** (0.14) {0.10} [0.00]	-0.20** (0.07) {0.10} [0.00]	-0.04 (0.16) {0.99} [0.82]	-0.02 (0.07) {0.99} [0.82]
Peer problem	2.78 (0.07)	-0.23* (0.09) {0.23} [0.01]	-0.17* (0.07) {0.23} [0.01]	-0.12 (0.10) {0.89} [0.21]	-0.09 (0.07) {0.89} [0.21]
Prosocial	7.06 (0.12)	0.50** (0.16) {0.06} [0.00]	0.24** (0.08) {0.06} [0.00]	0.25 (0.15) {0.76} [0.11]	0.12 (0.08) {0.76} [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 $[(y_i - \text{mean of the control group}) / \text{standard deviation of control group}]$. 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*** (0.09)	0.65*** (0.10)	-0.09 (0.10)	0.48*** (0.10)	0.75*** (0.09)	-0.25* (0.11)	0.57*** (0.08)	0.65*** (0.11)	-0.08 (0.11)	0.53*** (0.08)	0.75*** (0.11)	-0.21* (0.10)
Literacy	0.54*** (0.10)	0.57*** (0.10)	-0.02 (0.10)	0.48*** (0.11)	0.65*** (0.09)	-0.17 (0.12)	0.50*** (0.09)	0.64*** (0.11)	-0.13 (0.11)	0.51*** (0.08)	0.64*** (0.12)	-0.13 (0.11)
Numeracy	0.43*** (0.09)	0.62*** (0.08)	-0.18 (0.10)	0.38*** (0.08)	0.73*** (0.09)	-0.32* (0.10)	0.54*** (0.08)	0.51*** (0.09)	0.02 (0.11)	0.42*** (0.07)	0.72*** (0.09)	-0.29** (0.09)
Student's study time	0.10 (0.11)	0.11 (0.12)	-0.02 (0.14)	0.13 (0.11)	0.04 (0.10)	0.08 (0.12)	0.04 (0.10)	0.18 (0.12)	-0.14 (0.13)	0.11 (0.11)	0.13 (0.14)	0.00 (0.15)
Caregiver's time in homeschooling	0.22* (0.11)	0.19 (0.11)	0.04 (0.13)	0.24* (0.11)	0.16 (0.10)	0.08 (0.11)	0.21* (0.11)	0.17 (0.11)	0.02 (0.12)	0.35*** (0.10)	0.03 (0.14)	0.30* (0.14)
Observations	551	577	1128	612	516	1128	623	505	1128	667	461	1128
Panel B: Extended treatment												
Total score	0.61*** (0.08)	0.69*** (0.09)	-0.09 (0.10)	0.60*** (0.09)	0.69*** (0.10)	-0.10 (0.10)	0.52*** (0.07)	0.78*** (0.10)	-0.26* (0.10)	0.56*** (0.08)	0.77*** (0.10)	-0.22* (0.10)
Literacy	0.60*** (0.09)	0.63*** (0.09)	-0.03 (0.10)	0.61*** (0.09)	0.61*** (0.10)	-0.00 (0.11)	0.46*** (0.08)	0.76*** (0.11)	-0.29** (0.10)	0.53*** (0.09)	0.72*** (0.10)	-0.20 (0.11)
Numeracy	0.48*** (0.08)	0.62*** (0.08)	-0.15 (0.10)	0.44*** (0.08)	0.65*** (0.09)	-0.22* (0.09)	0.49*** (0.07)	0.62*** (0.09)	-0.14 (0.10)	0.47*** (0.07)	0.66*** (0.09)	-0.19 (0.10)
Student's study time	0.14 (0.12)	0.02 (0.10)	0.11 (0.13)	0.04 (0.11)	0.12 (0.10)	-0.07 (0.11)	0.01 (0.11)	0.16 (0.11)	-0.16 (0.12)	0.12 (0.10)	0.03 (0.11)	0.12 (0.12)
Caregiver's time in homeschooling	0.03 (0.09)	0.12 (0.10)	-0.05 (0.13)	0.15 (0.09)	-0.00 (0.10)	0.13 (0.10)	0.07 (0.09)	0.08 (0.11)	-0.03 (0.11)	0.07 (0.09)	0.08 (0.11)	0.00 (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	T1: Standard			T2: Extended	
	(1) Control Mean	(2) Treatment effect (Raw)	(3) Treatment effect (SD)	(4) Treatment effect (Raw)	(5) Treatment effect (SD)
Student's study time	87.03 (3.70)	5.42 (4.83) {0.58} [0.27]	0.10 (0.09) {0.58} [0.27]	3.72 (4.66) {0.61} [0.43]	0.07 (0.09) {0.61} [0.43]
Extent of study	2.65 (0.04)	0.18* (0.07) {0.07} [0.01]	0.20* (0.08) {0.07} [0.01]	0.21** (0.07) {0.02} [0.00]	0.24** (0.08) {0.02} [0.00]
Caregiver's time in homeschooling	75.48 (3.41)	9.72* (4.33) {0.12} [0.02]	0.20* (0.09) {0.12} [0.02]	3.38 (3.93) {0.61} [0.40]	0.07 (0.08) {0.61} [0.40]
Extent of caregiver's homeschooling	2.63 (0.07)	0.21** (0.08) {0.04} [0.00]	0.23** (0.09) {0.04} [0.00]	0.12 (0.08) {0.36} [0.11]	0.13 (0.08) {0.36} [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 $[(y_i - \text{mean of the control group}) / \text{standard deviation of control group}]$. 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)	(2)	(3)	(4)	(5)	(6)
	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy
Less-educated parents	0.07* (0.03)	0.03 (0.04)	0.51 (0.85)	0.32 (0.89)	-0.14 (0.49)	-0.34 (0.46)
Lower baseline score	-0.05 (0.04)	0.08* (0.03)	-1.89** (0.67)	-1.38 (0.73)	-0.91* (0.44)	-0.48 (0.40)
Lower household income	0.04 (0.04)	0.08 (0.04)	0.46 (0.72)	0.08 (0.71)	0.27 (0.53)	0.24 (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.

Appendices

Appendix A: Additional Material

A-1.Examples of Learning Activities	A3
A-1.1. Literacy module	A3
A-1.2. Numeracy module	A3
A-1.3. Leadership module	A4
A-2.Learning content of different modules	A5
A-2.1. Literacy module	A5
A-2.2. Numeracy module	A7
A-2.3. Leadership module	A10
A-3.Outcomes and instruments for measurement	A11
A-3.1. Learning outcomes	A11
A-3.2. Leadership, communication, and planning skills.....	A12
A-3.3. Noncognitive skills	A13
A-3.4. Behavioral difficulties	A15
A-3.5. Validation of non-cognitive skill measures	A16
A-3.6. Homeschooling time (student)	A16
A-3.7. Homeschooling time (Mother or caregiver)	A17
A-4.Social desirability bias	A17
A-5.Assessment test reliability	A18

Appendix B: Additional Figures and Tables

Additional Figures

Figure B1. IVR flow diagram of T1: Standard treatment group.....	B1
Figure B2. IVR flow diagram of T2: Extended treatment group.....	B2
Figure B3. Distribution of total number of lessons completed by caregiver-child dyads, by module .	B3
Figure B4. Completion by caregiver-child dyads, by module and lesson number.....	B4
Figure B5. Total hours of audio lessons accessed, by program week.....	B5
Figure B6. Assessment-test score distribution, by treatment.....	B6
Figure B7. Dosage responses in standardized literacy and numeracy scores	B7
Figure B8. Differential Item Functioning (DIF).....	B8

Additional Tables

Table B1. Survey attrition rate	B9
Table B2. Attrition, by treatment	B10
Table B3. Comparison between study sample and national sample survey.....	B11
Table B4. Percentage of children who answered correctly, by question and treatment	B12
Table B5. Treatment effects on number of correct answers	B13
Table B6. Heterogenous treatment effects on learning outcome, by quartiles.....	B14
Table B7. Parents' social desirability bias analysis.....	B15
Table B8. Child's social desirability bias analysis.....	B16
Table B9. Relationships between percent of completed lessons ever repeated and baseline characteristics	B17
Table B10. Reliability of endline assessment test questions.	B18
Table B11. Differential Item Functioning (DIF) analysis.	B19
Table B12. Differential Item Functioning (DIF) adjusted learning outcome.....	B20
Table B13. Costing table.....	B21

Appendix A: Additional Material

A-1. Examples of Learning Activities

In the following subsections, we provide a few such examples or activities from various modules of the intervention.

A-1.1. Literacy module

In the literacy lessons, caregivers or mothers were provided with alphabet sheets (illustrated below) to practice or to follow the instructions given by the instructor during the audio lessons played over IVR. Additionally, we relied on the Bangla textbook provided by the government for them to practice various literacy exercises.

English Alphabet

Aa Bb Cc Dd
Ee Ff Gg Hh
Ii Jj Kk Ll
Mm Nn Oo Pp
Qq Rr Ss Tt
Uu Vv Ww Xx
Yy Zz

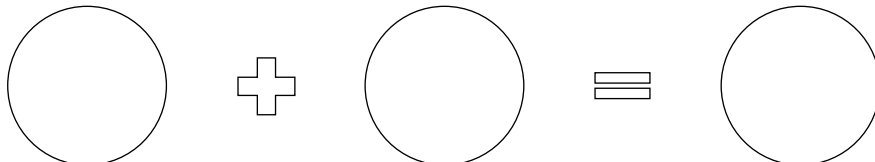
Bangla Alphabet

অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ	ট
ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন	প	ফ
ব	ভ	ম	য	র	ল	শ	ষ	স	হ	ড়
ঢ়	য়	ং	ঁ	ঃ	ৎ					

A-1.2. Numeracy module

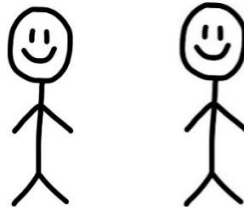
Counter: A counter is anything we can count. For example, sticks, marbles, pieces of bricks or colored pencils. The caregiver had to arrange the counters before the lesson and use them to help the child practice counting during the numeracy lessons.

Number partner: A number partner is two numbers when added together we get another new number. For example, 5 plus 5 equals 10. In this case, 5 and 5 are the number partners of 10. Similarly, 7 and 3 also add up to 10, so these two numbers are also number partners. A number can have many number partners. For example, the caregiver had to draw figures like the following one in a few lessons to help the child identify the number partners.



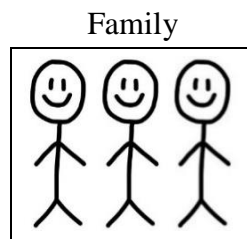
A-1.3. Leadership module

Stick figure: Stick figures are graphical representations of humans used for explaining various contexts to children. In leadership lessons, there were various situations where the child had to decide as per instructions given by the instructor during the audio lessons. For example, the caregiver had to draw stick figures to visually explain the situations, as follows:

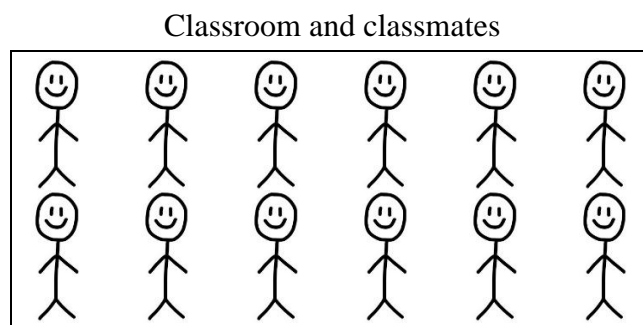


Oral saline: In lesson 5 of the leadership module, the caregiver taught their child how to make oral saline. The recipe for homemade oral saline was given to the caregiver beforehand in the guidebook. For example, *“The process of making oral saline is very simple and you all know it. 1 handful of sugar or molasses, 3 finger pinches of salt mixed with half a liter of clean water to make oral saline.”*

Family picture: In a few lessons, the caregiver had to draw a family figure. During the lesson, the instructor gave instructions to the caregiver to interact with their child using the family figure. A sample family figure is given as follows:



Classroom: The caregiver had to draw a classroom picture on a few occasions. An example of this picture is as follows:



A-2. Learning content of different modules

The literacy and numeracy modules were based on the lessons of the ‘Rising on Air’ program, while the leadership module was based on the ‘LEAD Learning’ program. These lessons were completely rewritten and contextualized for Bangladesh by local educators and educational researchers. These contents are briefly presented in the following sections.

A-2.1. Literacy module

The table below provides a brief overview of the contents of the literacy module:

Table A1. Literacy module’s lesson plan

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
1	Listening; Speaking; Reading.	Sentence; Words; Story	Sentence warm-up	Rice is healthy	Student will be able to (SWBAT) understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Boat; Cat	
			Brilliant blending words	Pen; Man; Rat; Nap	
			Story	The lost laugh (1)	
2	Listening; Speaking; Reading.	Sentence; Words; Story	Sentence warm-up	Cats have powerful eyes	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Tongue; Kitchen; Bird	
			Brilliant blending words	Pin; Fun; Tin	
			Story	The lost laugh (2)	
3	Listening; Speaking; Reading.	Sentence; Words; Story	Sentence warm-up	A mosquito is an insect	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Wave; Beach; Sky	
			Brilliant blending words	Pet; But; Wet; Map	
			Story	The bee & the elephant	
4	শোনা, বলা, পড়া, লিখা (In Bangla)	নতুন বাংলা শব্দ; সমার্থক শব্দ; বিপরীতার্থক শব্দ	ছড়া – শব্দ গঠন	আতা গাছে তোতা পাখি	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
			প্রতিশব্দ	পাখি- পক্ষী, বিহগ, খেচর	
			বিপরীত শব্দ	গাছ- বৃক্ষ, তরু, শাখী	
			গল্পের রাজ্য	ছোট- বড়, জয়- পরাজয়, বর-বধু, বেশি -কম	
5	Listening; Speaking; Reading.	Sentence; Words; Story	Sentence warm-up	Babies are born without teeth	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Three; Door; Rice	
			Brilliant blending words	Fish; Path; Bath	
			Story	The toothache (1)	
6	Listening; Speaking; Reading.	Sentence; Words; Story	Sentence warm-up	My dad is caring	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Hat; Fan; Bag	
			Listen and write words	Cat; Van; Man	
			Story	The toothache (2)	
7	Listening; Speaking; Reading.	Sentence; Words; Story	Sentence warm-up	Clouds are made up of tiny droplets of water	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Win; Sing; Dress	
			Listen and write words	Chop; Sad; Rock	
			Story	The red raincoat (1)	
8	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া	আমি হব-কাজী নজরুল ইসলাম	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
			প্রতিশব্দ	রাত- রাত্রি, নিশি, রজনী	
			বিপরীত শব্দ	সূর্য- রবি, তপন, ভানু	
			গল্পের রাজ্য	আলসে-কর্মঠ, সকাল-বিকাল, আগে-পরে, ঘুমিয়ে-জেগে	
9	Listening; Speaking; Reading; Writing	Sentence; Words; Story	Sentence warm-up	Fruit and vegetables are good for you	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Pin; Pen; Pan	
			Brilliant blending words	Beg; Fish; Ten	
			Story	The red raincoat (2)	
10	Listening; Speaking; Reading; Writing	Sentence; Words; Story	Sentence warm-up	Buses carry many passengers	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Bit; Red; Ten	
			Brilliant blending words	Fill; Hill; Sand	
			Story	The boy’s new bike (1)	

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
11	Listening; Speaking; Reading; Writing		Sentence warm-up	When you are honest you are telling the truth	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Brilliant blending words	Crop; Swim; Skip	
			Listen and write words	Crop; Swim; Skip	
			Story	The boy's new bike (2)	
12	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া .শব্দ গঠন	কাজের আনন্দ- নবকৃষ্ণ ভট্টাচার্য আহরণ, কব, তপলতা, বৃনি, সঞ্চয়, পিপীলিকা	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
			প্রতিশব্দ	ভাই- ভ্রাতা, ভাইয়া, সহোদর পিপীলিকা- পিপড়া, পিপড়ে	
			বিপরীত শব্দ	শীত- গ্রীষ্ম, আসা- যাওয়া, দাঁড়ানো - বসা, আনা- নেওয়া	
			গল্পের রাজ্য	শীতের সকাল	
			নতুন শব্দ ও বাক্য গঠন	পোহান, নাশতা	
13	Listening; Speaking; Reading. Writing		Sentence warm-up	The girl was crying because she was sad	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Brilliant blending words	Ask; Text; Next	
			Listen and write words	Ask; Text; Desk	
			Story	My feelings (1)	
14	Listening; Speaking; Reading. Writing		Sentence warm-up	Just like humans, plants need water to survive	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Brilliant blending words	Grand; Crept; Past	
			Listen and write sentence	A thin rat crept past the fat cat	
			Story	My feelings (2)	
15	Listening; Speaking; Reading. Writing		Sentence warm-up	Plants can be used as medicine	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Brilliant blending words	Fish; Stretch; Net	
			Listen and write sentence	The men on the ship stretch a big fishing net	
			Story	The red plant (1)	
16	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া .শব্দ গঠন	ট্রেন- শামসুর রাহমান	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
			প্রতিশব্দ	নদী- তটিনী, তরঙ্গিনী, প্রবাহিণী বাড়ি- গৃহ, আবাস, নিবাস	
			বিপরীত শব্দ	ঘরে- বাইরে, দেশ- বিদেশ, ছোট- থামা	
			গল্পের রাজ্য	গাছ লাগানো	
			নতুন শব্দ ও বাক্য গঠন	গাছ	
17	Listening; Speaking; Reading.		Sentence warm-up	It is important to exercise every day	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Cat/Cart; Pot/Port; Shirt/Shit	
			Brilliant blending words	Sit; Birth; Tom	
			Story	The red plant (2)	
18	Listening; Speaking; Reading.		Sentence warm-up	Keep your mouth happy by brushing your teeth	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Gem/Germ; Hut/Hurt	
			Brilliant blending words	Ten; Term; Turn; Sun	
			Story	The magical lunchbox (1)	
19	Listening; Speaking; Reading.		Sentence warm-up	Colors help you remember memories	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Sound workout words	Sport; Spurt; Barn	
			Listen and write words	Corn; planner	
			Story	The magical lunchbox (2)	
20	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া .শব্দ গঠন	বড় কে- হরিশ্চন্দ্র মিত্র	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
			প্রতিশব্দ	সংসার- জগত, ভুবন, বিশ্ব বড়- বৃহৎ, প্রকাণ্ড	
			বিপরীত শব্দ	বড়- ছোট, সহজ - কঠিন, যার - তার, সুখ - দুঃখ, দোষ- গুণ	
			গল্পের রাজ্য	পাখিদের কথা	
			নতুন শব্দ ও বাক্য গঠন	সংসার; কঠিন; ব্যাপার প্রিয় পাখি	
21	Listening; Speaking; Reading. Writing		Sentence warm-up	Planting trees is good for the environment	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
			Write and read words	Strong; Swimmer; Faster; Longer	
			Listen and write sentence	That girl is a strong swimmer. She can swim faster and longer than me	
			Story	Water is important (1)	
22			Sentence warm-up	Drinking water is needed for humans to stay alive	SWBAT understand English sentence structure; basic

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
	Listening; Speaking; Reading. Writing		Sound workout words Brilliant blending words Story	Bit; Bite Rip; Ripe; Hid; Hide Water is important (2)	spelling; and the meaning of a short story.
23	Listening; Speaking; Reading. Writing		Sentence warm-up Sound workout words Listen and write words Story	People from different countries can wear different clothes June; But Bone; Kite; Sharp The hat seller (1)	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
24	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া -শব্দ গঠন প্রতিশব্দ বিপরীত শব্দ গল্পের রাজ্য বাক্য গঠন	মামার বাড়ি- জসীমউদ্দীন বাড়- ঝাঞ্জা, ঝাটিকা, ঝাপটা, পুষ্প- কুসুম, ফুল, কুঁড়ি কাঁচা- পাকা, উঠা- নামা, ছেলে- বুড়ো, পিঁপড়ে ও ঘুঘু পিঁপড়ে ও ঘুঘু	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
25	Listening; Speaking; Reading. Writing		Sentence warm-up Write and read words Listen and write sentence Story	Fingernails grow faster than toenails. Read; Tale; Page We like this book. It has tales. The hat seller (2)	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
26	Listening; Speaking; Reading. Writing		Sentence warm-up Sound workout words Write and read words Story	People who watch sports can be called spectators Plate; Play; Rain Train; Tray The lost ball	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
27	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া -শব্দ গঠন প্রতিশব্দ বিপরীত শব্দ গল্পের রাজ্য বাক্য গঠন	আদর্শ ছেলে মানুষ- মানব, নর, লোক ভয়- ডর, ভীতি, আতঙ্ক হাত- পা, বিপদ- আপদ, হাসি- কান্না, কল্যাণ- অকল্যাণ একজন পটুয়ার কথা কামরুল হাসান	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.
28	Listening; Speaking; Reading. Writing		Sentence warm-up Sound workout words Write and read words Story	A triangle is a shape with three sides Kite; Child; Sky Fit; Fried; Night When will mother be back? (1)	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
29	Listening; Speaking; Reading. Writing		Sentence warm-up Spelling sound out Write and read words Story	You can use your fingers to count to ten Glue; Fruit; New Cup; Value; Juice When will mother be back? (2)	SWBAT understand English sentence structure; basic spelling; and the meaning of a short story.
30	শোনা, বলা, পড়া, লিখা (In Bangla)	বাংলা নতুন বাংলা শব্দ, সমার্থক ও বিপরীতার্থক শব্দ	ছড়া -শব্দ গঠন প্রতিশব্দ বিপরীত শব্দ গল্পের রাজ্য বাক্য গঠন	আমাদের এই বাংলাদেশ- সৈয়দ শামসুল হক দেশ- রাষ্ট্র, স্বদেশ, জন্মভূমি বীর- সাহসী, নিভীক, অকুতোভয় পূর্ব- পশ্চিম, স্বাধীন- পরাধীন, আপন- পর, বীর- কাপুরুষ, প্রিয়- অপ্রিয় ভাষাশহিদের গল্প মাতৃভাষা	SWBAT understand the synonym and antonym of Bengali words and the meaning of a short story.

A-2.2. Numeracy module

The table below provides a brief overview of the contents of the numeracy module:

Table A2. Numeracy module's lesson plan

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
1	Counting	Counting 1 to 10; Identification 1 to 10	Mindful moment Number warmup Number workout Brain break Challenge problem	Counting with breathing Counting Counters Right-left Hand	Students will be able to (SWBAT) count out and write represent a quantity 1-10; SWBAT identify numerals 1-10
2	Counting; Addition	Counting 1 to 20; simple addition	Mindful moment Number warmup Number workout	Music Counting Compare	SWBAT count out and write represent a quantity 1-20; SWBAT identify

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
3	Addition	Simple two numbers addition	Brain break	Music	numerals 1-20; SWBAT add simple numbers
			Challenge problem	Addition	
			Mindful moment	Sound focus	
			Number warmup	Count and move	
			Number workout	Addition	
4	Subtraction	Simple two numbers subtraction	Brain break	Action	SWBAT solve joining problems (using counters, counting all, or counting on) and write a matching equation.
			Challenge problem	Count	
			Mindful moment	Balance exercise	
			Number warmup	Counting backwards	
			Number workout	Subtraction	
5	Addition and subtraction	Solve addition and subtraction problems	Brain break	Dance	SWBAT solve separating problems (using counters, counting all, or counting on) and write a matching equation when the part is unknown.
			Challenge problem	Subtraction	
			Mindful moment	Muscle squeeze exercise	
			Number warmup	Count and move	
			Number workout	Addition and subtraction	
6	Number Comparison	Decomposition, addition, subtraction and comparison	Brain break	Shakedown	SWBAT represent and solve addition and subtraction problems.
			Challenge problem	Addition & subtraction	
			Mindful moment	Deep breathing exercise	
			Number warmup	Count and move	
			Number workout	Comparison	
7	Ordinal Number	Ordinal number	Brain break	Role play	SWBAT review the concept of Writing numbers, decomposition, addition, subtraction and comparison of numbers from 1 to 10
			Challenge problem	Addition	
			Mindful moment	Balance exercise	
			Number warmup	Count and move	
			Number workout	Ordinal number	
8	Numerals	Different sets of groups, number bonds and write a matching number sentence	Brain break	Song	SWBAT identify and describe ordinal numbers (first to tenth)
			Challenge problem	Tricky numbers	
			Mindful moment	Sound focus exercise	
			Number warmup	Detectives	
			Number workout	Numerals	
9	Ordinal numbers	Identify and describe ordinal numbers (eleventh to twentieth)	Brain break	Freeze dance	SWBAT determine and write how many objects are in a set (10-15) and identify numerals 10-15
			Challenge problem	Number bond	
			Mindful moment	Balance exercise	
			Number warmup	Skip counting	
			Number workout	Ordinal numbers	
10	Teen numbers	Understand teen numbers are composed of tens and ones	Brain break	Freeze dance	SWBAT identify and describe ordinal numbers (eleventh to twentieth)
			Challenge Problem	Identify circle	
			Mindful moment	Sound focus exercise	
			Number warmup	Guess number	
			Number workout	Teen numbers	
11	Multiple addition equations	Understand and apply multiple addition equations	Brain break	As if	SWBAT understand, identify, and write teen numbers
			Challenge problem	Teen numbers	
			Mindful moment	Balance exercise	
			Number warmup	Beep counting	
			Number workout	Commutative property of addition	
12	Subtraction	Subtracting two values using different ways	Brain break	Creative dance	SWBAT add more than two numbers and write the equation in multiple ways
			Challenge problem	True and false	
			Mindful moment	Sound sense	
			Number warmup	count backwards	
			Number workout	different ways of subtracting two values	
13	Addition and subtraction	Solving addition and subtraction problems	Brain break	Dance party	SWBAT subtract two values using fingers, counting back, counting up, or known facts.
			Challenge problem	Subtraction problems	
			Mindful moment	Muscle squeeze exercise	
			Number warmup	Count and move	
			Number workout	Solving addition and subtraction problems together	
14	Addition and subtraction comparison	Solving addition and subtraction comparison problems	Brain break	Shakedown	SWBAT solve addition and subtraction problems using ten as a landmark.
			Challenge problem	Mystery number	
			Mindful moment	Balancing exercise	
			Number warmup	Count and move	
			Number workout	Comparison addition & Subtraction	
15	Equal sign		Brain break	Creative dance	SWBAT solve comparison problems using addition and subtraction.
			Challenge Problem	Number sentence	
			Mindful moment	Balance exercise	

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
		Understanding equal sign and comparing both sides of the equal sign	Number warmup Number workout Brain break Challenge problem	Beep counting Identifying true-false Alphabet mov Distinguishing between true-false equations	SWBAT understand the meaning and application of equal sign
16	Expanded form of numbers	Understand and write the expanded or standard form of numbers	Mindful moment Number warmup Number workout Brain break Challenge problem	Sounds workout Count forward by tens Standard form As if... Mystery Number	SWBAT represent numbers 10-99 in expanded form.
17	Place value to compare two-digit	Learn to use place value to help us compare two-digit	Mindful moment Number warmup Number workout Brain break Challenge problem	Deep breathing exercise Count forward by 2s Comparing numbers As If... Counting and addition, Mystery number- L.P 11.5	SWBAT compare two-digit numbers by reasoning about tens and ones.
18	Number & Place value	Digit, number, place value, comparison	Mindful moment Number warmup Number workout Brain break Challenge problem	Balancing exercise Count forward by 2s Make numbers using place value As if..." Compare big and small numbers	SWBAT model numbers with more than 9 ones
19	Different forms of numbers	Expanded form, word form, standard form, compare numbers	Mindful moment Number warmup Number workout Brain break Challenge problem	Muscle squeeze exercise Count forward by 2s Various forms of numbers As if... Comparing numbers	SWBAT order numbers in different forms
20	Composed numbers	Represent numbers with pictures of sticks and dots, Composed numbers	Mindful moment Number warmup Number workout Brain break Challenge Problem	Balance exercise Skip count forward and backward Composed numbers Freeze dance Expand the form of numbers	SWBAT model numbers with tens sticks and counters, pictures of sticks and dots, and/or counting on
21	Expanded notation (No regrouping)	Addition, number break	Mindful moment Number warmup Number workout Brain break Challenge problem	Deep breathing exercise Detectives Expanded notation As if... Expanded notation	SWBAT add 2 two-digit numbers (no regrouping) using pictures and expanded notation
22	Expanded notation (With regrouping)	Identify the tens and ones in a two-digit number	Mindful moment Number warmup Number workout Brain break Challenge problem	Sound focus exercise Addition Adding numbers Shakedown Addition strategies	SWBAT add 2 two-digit numbers (with regrouping) using pictures and expanded notation.
23	Subtract (No regrouping)	Addition, subtraction, identify the tens and ones in a single- and double-digit number	Mindful moment Number warmup Number workout Brain Break Challenge problem	Deep breathing Substrate sentence Subtract Freeze dance Subtract with expanded notation	SWBAT subtract one- and two-digit numbers from two-digit numbers (no regrouping) using the standard algorithm.
24	Subtract (With regrouping)	Addition, subtraction, and multi-step instruction to solve with a standard algorithm	Mindful moment Number warmup Number workout Brain break Challenge problem	Balance exercise Skip count Subtract (with regrouping) Dance party Subtract with expanded notation	SWBAT subtract one- and two-digit numbers from two-digit numbers (with regrouping) using the standard algorithm.
25	Addition and subtraction word problem	Addition, subtraction	Mindful moment Number warmup Number workout Brain break Challenge problem	Balance exercise 2's Addition and subtraction word problem As if... Word problem	SWBAT solve 2-digit addition and subtraction word problems using the standard algorithm.
26	Numbers with more than 9 ones	Expanded notation, unit form, numbers with stick and dots	Mindful moment Number warmup Number workout Brain break Challenge problem	Sound focus exercise Standard form Represent number Dance party Number with unit form	SWBAT model numbers with more than 9 ones
27	Place value	Greatest to latest, latest to greatest	Mindful moment Number warmup	Deep breath Mystery number	SWBAT orders 2-digit numbers in different forms

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
28	Order number in a different form	Expanded, standard and written form, largest to smallest and smallest to largest	Number workout	Place value	SWBAT compare numbers with more than 9 ones
			Brain break	As if...	
			Challenge problem	Mystery Number	
			Mindful moment	muscle squeeze exercise	
			Number warmup	Count and move	
			Number workout	Order number	
29	Mental math and comparison	Greatest to least, least to greatest, smallest to largest, largest to smallest of numbers	Brain break	As if...	SWBAT review expanded form, +-/10/100 mental math, and comparison
			Challenge problem	Mystery number	
			Mindful moment	Balance exercise	
			Number warmup	Count and move	
			Number workout	Mental math and comparison	
30	Order numbers in many forms	10 ones = 1 ten, 10 tens = 1 hundred, largest to smallest, smallest to largest, solve problems in multiple ways	Brain break	Shakedown	SWBAT order numbers in many forms
			Challenge problem	Mystery number	
			Mindful moment	Sound focus exercise	
			Number warmup	Count and move	
			Number workout	Number order	
			Brain break	Freeze dance	

A-2.3. Leadership module

The table below provides a brief overview of the contents of the leadership module:

Table A3. Leadership module's lesson plan

No	Competency	Content	Sub-content	Intervention lesson	Learning outcome
1	Leadership	What is leadership?	Exploration	Leadership; Leader	Introduce students to the concept of leadership
			Blast from the past	President; Prime Minister; Cricket Captain	
			Creative corner	Doctor; Nurse	
			Acts of leadership	2 hypothetical stories	
2	Confident; Creative	The 5 qualities of a leader, part 1	Memory kick	Leadership	Introduce students to the first two qualities – be confident and be creative.
			Exploration	Qualities; Be confident; Be creative	
			Blast from the past	Gold medal of BD women's cricket team in SAG 2019 (confident) Dr Rafiqul Islam, inventor, ORS (creative)	
			Acts of leadership	1 hypothetical story	
3	Visionary; Teamwork; Delegation of Authority	The 5 qualities of a leader, part 2	Memory kick	Be confident; Be creative	Introduce students to the last three qualities – set the example, work together, and recognize and applaud success
			Exploration	Qualities; set the example; work together; recognize and applaud success.	
			Blast from the past	Freedom fighter Mostafa Kamal	
			Creative Corner	Scoring in a football match	
4	Better listening	Active listening	Memory kick	Set the example; work together; recognize and applaud success	Introduce students to the Importance of listening
			Exploration	Active listening	
			Blast from the past	The landscape of the language movement	
			Creative corner	Father-child conversation	
			Acts of leadership	1 hypothetical short story	
5	communicate	Communication and presenting yourself	Memory kick	Active listening	Learn how to communicate well and be approachable
			Exploration	Communication	
			Blast from the past	Greta Thunberg	
			Creative corner	Making oral saline	
6	Planning	How to 'make a plan'	Acts of leadership	1 hypothetical short story	Introduce students to the concept of brainstorming and how to make a plan
			Memory kick	Communication	
			Exploration	Brainstorm; Choose the best idea	
			Blast from the past	An incident from the independence war	
			Creative corner	Going to the town	
7	Humble	Bragging vs. humility	Acts of leadership	1 hypothetical short story	Introduce students to the concept of humility and how to avoid bragging
			Memory kick	Planning	
			Exploration	Humility; Bragging	
			Blast from the past	Muhammad Yunus – Nobel Lecture	
			Acts of leadership	1 hypothetical short story	

No 8	Competency Patience	Content Patience I	Sub-content	Intervention lesson	Learning outcome Introduce students to the concept of patience
			Memory kick	Humility; Bragging	
			Exploration	Patience	
9	Patience	Patience II	Blast from the past	Bangladesh vs Zimbabwe test 2005	Example of patience
			Memory kick	Patience	
			Creative corner	Things to do before play	
10	Empathy	Empathy I	Acts of leadership	3 hypothetical short stories	Introduce students to the meaning of empathy and how to be empathetic
			Memory kick	Patience	
			Exploration	Empathy	
11	Empathy	Empathy II	Blast from the past	Mother Teresa	Examples of empathetic acts
			Memory kick	Empathy	
			Creative corner	Showing empathy to classmates; Showing empathy to other	
12	Sympathy; Compassion	Sympathy and compassion I	Acts of leadership	3 hypothetical short stories	Introduce students to the concept of sympathy and how to be compassionate
			Memory kick	Empathy	
			Exploration	Sympathy; Compassion	
13	Sympathy; Compassion	Sympathy and compassion II	Blast from the past	Rohingya refugees 2017	Example of sympathy
			Memory kick	Sympathy; Compassion	
			Creative corner	Prosocial experimental scenario	
14	Perseverance	Perseverance	Acts of leadership	3 hypothetical short stories	Introduce students to the importance of perseverance
			Memory kick	Sympathy; Compassion	
			Exploration	Perseverance	
15	Module review	Key Ideas	Blast from the past	Nelson Mandela [link 1] [link 2]	Review the key ideas of leadership and the qualities of a leader
			Memory kick	2 hypothetical short stories	
			Acts of leadership		

A-3. Outcomes and instruments for measurement

A-3.1. Learning outcomes

Table A4. Endline assessment test questions

Subject	No	Level 1 (Grade 1)	Level 2 (Grade 2)	Level 3 (Grade 3 & 4)
Literacy	1.	Read aloud the following letters (the first 4 letters from the Bengali alphabet)	Make two words using the Bangla letter ----.	Read aloud this following paragraph (Bangla).
	2.	Fill in the gaps (5 Bangla letters with 2 gaps).	Fill in the gap (a line in Bangla from the textbook)	What is the antonym of the Bangla word (FREEDOM)?
	3.	Make a word with the Bengali letter ----.	What is the spelling of the word (Sundarbans)?	What is the spelling of the word (Bangla of freedom fighter)?
	4.	What is the spelling of (Bengali word)?	What is the antonym of the Bangla word (high)?	What is the meaning of this Bangla word (Bangla word from the textbook)?
	5.	Read the following word (CAP).	Read the following word (FARMER)?	Read aloud this following paragraph (English).
	6.	Answer this English question: What is your name?	Answer this English question: How old are you?	Answer this English question: What month is it now?
	7.	Say the English of Bangla word – (DOOR).	Say the English of Bangla word – (WINDOW).	Say the English of Bangla word – (FARMER).
	8.	Say the English of Bangla word – (BOOK).	Say the English of Bangla word – (UMBRELLA).	Say the English of Bangla word – (WEDNESDAY).
	9.	Say the English of Bangla word – (DOG).	Say the English of Bangla word – (BREAKFAST).	Say the English of Bangla word – (FLAG).
	10.	Spell your name in English.	Read and say the name of these shapes (picture of the square, circle, triangle, and rectangle).	Match the appropriate description with this picture (match from 4 options).
Num eracy	11.	Which number comes after 6? Does it even or odd?	Name the even numbers between 1 and 10.	Sort these three numbers, smallest to the largest (20, 73, 10, 78).

12.	What is the result of 3+4=?	Sort these three numbers, smallest to largest (23, 17, 38).	There are 6 notes of 20 BDT. How much money is there?
13.	What is the result of 8-3=?	In a class, there were 16 students. The teacher sends 5 of them for gardening. How many students are left in the classroom?	What is the result of 13+11=?
14.	How many minutes in 60 seconds?	How many sides a triangle has?	What is the result of 2/4+2/4=?
15.	What is the result of 6+0=?	There are three fruits on a plate. How many fruits there are in 4 plates?	The price of 5 eggs is BDT 30. How much does it cost to buy 2 eggs?

Note: The test was conducted on a one-on-one basis.

From the answers to these questions, three learning outcome variables were constructed. These are –

- **Total score:** All 15 literary and numeracy questions from Table A4. The test totals eighty points. This variable is continuous; $Y_i^{ts} \in [0,80]$.
- **Literacy:** 10 questions on literacy (English & Bangla). This variable is continuous; $Y_i^{lit} \in [0,50]$.
- **Numeracy:** Five questions on numeracy. This variable is continuous; $Y_i^{num} \in [0,30]$.

A-3.2. Leadership, communication, and planning skills

Table A5. Children’s leadership, communication, and planning skills assessment questions

Scale	No	Questions	Answer
[Every statement will start with] My child demonstrates . . .			
7-item Leadership Characteristics	1.	responsible behavior, can be counted on to follow through on activities/projects.	(1) Never
	2.	a tendency to be respected by classmates.	(2) Very rarely
	3.	the ability to articulate ideas and communicate well with others.	(3) Rarely
	4.	self-confidence when interacting with age peers.	(4) Occasionally
	5.	the ability to organize and bring structure to things, people, and situations.	(5) Frequently
	6.	cooperative behavior when working with others.	(6) Always
	7.	a tendency to direct an activity when he or she is involved with others.	
4-item Communication Characteristics	1.	uses voice expressively to convey or enhance meaning.	(1) Never
	2.	conveys information nonverbally through gestures, facial expressions, and “body language.”	(2) Very rarely
	3.	is an interesting storyteller.	(3) Rarely
	4.	uses colorful and imaginative figures of speech such as puns and analogies.	(4) Occasionally (5) Frequently (6) Always
15-item Planning Characteristics	1.	determines what information or resources are necessary for accomplishing a task.	
	2.	grasps the relationship of individual steps to a whole process.	
	3.	allows time to execute all steps involved in a process.	
	4.	foresees consequences or effects of action.	(1) Never
	5.	organizes his or her work well.	(2) Very rarely
	6.	takes into account the details necessary to accomplish a goal.	(3) Rarely
	7.	is good at games of strategy where it is necessary to anticipate several moves ahead.	(4) Occasionally
	8.	recognizes the various alternative methods for accomplishing a goal.	(5) Frequently
	9.	can pinpoint where areas of difficulty might arise in a procedure or activity.	(6) Always
	10.	arranges steps of a project in a sensible order or time sequence.	
	11.	is good at breaking down an activity into step-by-step procedures.	
	12.	establishes priorities when organizing activities.	

13.	shows awareness of limitations relating to time, space, materials, and abilities when working on group or individual projects.
14.	can provide details that contribute to the development of a plan or procedure.
15.	sees alternative ways to distribute work or assign people to accomplish a task.

From the answers to these questions listed in Table A5, three variables were constructed as follows:

- **Leadership:** Each question of the leadership subscale from the Renzulli scale has a 7-point Likert scoring option. By adding the score of individual questions, the leadership score is calculated. The variable is continuous; $Y_i^{lead} \in [7,42]$.
- **Communication:** Expressive communication subscale from the Renzulli scale has 4 items and each has a 7-point Likert scoring option. The formation of this variable is continuous and calculated as the sum of all sub-questions; $Y_i^{com} \in [4,24]$.
- **Planning:** The planning subscale from the Renzulli scale has 16 questions and each has a similar 7-point Likert scoring option. The variable is continuous; $Y_i^{plan} \in [15,90]$.

A-3.3. Noncognitive skills

Table A6. Children’s noncognitive skill assessment survey questions

Scale	No	Questions	Answer
8-item Impulsivity Scale for Children	1.	I forgot something I needed for class.	(1) Almost never (2) About once a month (3) About 2-3 times a month (4) About once a week (5) At least once a day
	2.	I interrupted other students while they were talking.	
	3.	I said something rude.	
	4.	I couldn’t find something because my desk, locker, or bedroom was messy.	
	5.	I lost my temper at home or at school.	
	6.	I did not remember what my teacher told me to do.	
	7.	My mind wandered when I should have been listening.	
	8.	I talked back to my teacher or parent when I was upset.	
8-item Grit Scale	1.	New ideas and projects sometimes distract me from previous ones. (R)	(1) Very much like me (2) Mostly like me (3) Somewhat like me (4) Not much like me (5) Not like me at all
	2.	Setbacks don’t discourage me.	
	3.	I have been obsessed with a certain idea or project for a short time but later lost interest. (R)	
	4.	I am a hard worker.	
	5.	I often set a goal but later choose to pursue a different one. (R)	
	6.	I have difficulty maintaining my focus on projects that take more than a few months to complete. (R)	
	7.	I finish whatever I begin.	
	8.	I am diligent.	
Growth Mindset Scale	1.	You have a certain amount of intelligence, and you can’t really do much to change it.	(1) Strongly agree (2) Agree (3) Mostly agree (4) Mostly disagree (5) Disagree (6) Strongly disagree
	2.	Your intelligence is something about you that you can’t change very much.	
	3.	You can learn new things, but you can’t really change your basic intelligence.	
EmQue-CA	1.	If my mother is happy, I also feel happy.	(1) Not true (2) Sometimes true (3) often true
	2.	I understand that a friend is ashamed when he/she has done something wrong.	
	3.	If a friend is sad, I like to comfort him.	
	4.	I feel awful when two people quarrel.	
	5.	When a friend is angry, I tend to know why.	
	6.	I would like to help when a friend gets angry.	
	7.	If a friend is sad, I also feel sad.	

8.	I understand that a friend is proud when he/she has done something good.
9.	If a friend has an argument, I try to help.
10.	If a friend is laughing, I also laugh.
11.	If a friend is sad, I understand mostly why.
12.	I want everyone to feel good.
13.	When a friend cries, I cry myself.
14.	If a friend cries, I often understand what has happened.
15.	If a friend is sad, I want to do something to make it better.
16.	If someone in my family is sad, I feel really bad.
17.	I enjoy giving a friend a gift.
18.	When a friend is upset, I feel upset too.

Note: R - reverse scoring.

From the answers to these questions listed in [Table A6](#), multiple outcome variables were constructed:

- **Overall impulsivity:** Each question of the 8-item Impulsivity Scale for Children (ISC) has a 5-point Likert scoring option. Thus, the variable is continuous; $Y_i^{ISC} = \frac{\sum_{q=1}^{q=8} ISC_q}{8} \in [1,5]$.
- **Schoolwork impulsivity:** Items 1, 4, 6, and 7 from the ISC survey are used to construct the schoolwork impulsivity variable. The variable is continuous; $Y_i^{sch} = \frac{\sum_{q=1}^{q=4} ISC_q}{4} \in [1,5]$.
- **Interpersonal impulsivity:** Items 2, 3, 5, and 8 from the ISC survey are used to construct the interpersonal impulsivity variable. The variable is continuous; $Y_i^{per} = \frac{\sum_{q=1}^{q=4} ISC_q}{4} \in [1,5]$.
- **Grit:** Each question of the 8-item grit scale has a 5-point Likert scoring option. The great variable is constructed by averaging these questions. Thus, the variable is continuous; $Y_i^{grit} = \frac{\sum_{q=1}^{q=8} GRIT_q}{8} \in [1,5]$.
- **Growth mindset:** Each question of the 3-item growth mindset scale has a 6-point Likert scoring option. The mean of these questions indicates the growth mindset of the children. Thus, the variable is continuous; $Y_i^{gms} = \frac{\sum_{q=1}^{q=3} GMS_q}{3} \in [1,6]$.
- **Affective empathy (contagion):** 18-item self-report questionnaire that examines the level of empathy in three domains. Each question has a 3-point Likert answering option. All questions are exhibited in [Table A6](#). Items 1, 4, 7, 10, 13, 16, and 18 from the EmQue-CA survey are used to construct the affective empathy of the children. The variable is continuous; $Y_i^{EmA} = \frac{\sum_{q=1}^{q=7} EmQue_q}{7} \in [0,2]$.
- **Cognitive empathy (understanding):** Items 2, 5, 8, 11, and 14 from the EmQue-CA survey are used to construct the cognitive empathy of the children. The variable is continuous; $Y_i^{EmC} = \frac{\sum_{q=1}^{q=5} EmQue_q}{5} \in [0,2]$.
- **Prosocial motivation (support):** Items 3, 6, 9, 12, 15, and 17 from the EmQue-CA survey the EmQue-CA survey is used to construct the cognitive empathy of the children. The variable is continuous; $Y_i^{EmP} = \frac{\sum_{q=1}^{q=6} EmQue_q}{6} \in [0,2]$.

A-3.4. Behavioral difficulties

We used the 25-item parent-report Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). The main motivation for using this scale was to assess whether school closure triggered any fatigue, abnormality or conduct problems among the students and whether participation in the program reduced these problems. Items of this scale are listed in the following Table A7.

Table A7. Children’s behavioral difficulties assessment survey questions

No	Questions	Subscale	Answer
	Every statement will start with] My child ...		
1.	Considerate of other people’s feelings	Prosocial	
2.	Restless, overactive, cannot stay still for long	Hyperactivity	
3.	Often complains of headaches, stomach-aches, or sickness	Emotional Symptoms	
4.	Shares readily with other children, for example toys, treats, pencils	Prosocial	
5.	Often loses temper	Conduct Problem	
6.	Rather solitary, prefers to play alone	Peer Problem	
7.	Generally, well behaved, usually does what adults request (R)	Conduct Problem	
8.	Many worries or often seems worried	Emotional Symptoms	
9.	Helpful if someone is hurt, upset, or feeling ill	Prosocial	
10.	Constantly fidgeting or squirming	Hyperactivity	
11.	Has at least one good friend (R)	Peer Problem	
12.	Often fights with other children or bullies them	Conduct Problem	
13.	Often unhappy, depressed, or tearful	Emotional Symptoms	1) Not true
14.	Generally liked by other children (R)	Peer Problem	2) Somewhat true
15.	Easily distracted, concentration wanders	Hyperactivity	3) Certainly true
16.	Nervous or clingy in new situations, easily loses confidence	Emotional Symptoms	
17.	Kind to younger children	Prosocial	
18.	Often lies or cheats	Conduct Problem	
19.	Picked on or bullied by other children	Peer Problem	
20.	Often volunteers to help others (parents, teachers, other children)	Prosocial	
21.	Thinks things out before acting (R)	Hyperactivity	
22.	Steals from home, school or elsewhere	Conduct Problem	
23.	Gets along better with adults than with other children	Peer Problem	
24.	Many fears, easily scared	Emotional Symptoms	
25.	Good attention span, sees chores or homework through to the end (R)	Hyperactivity	

Note: R - reverse scoring.

From the answer to these questions from Table A7, 5 outcome variables are constructed. These are –

- **Emotional symptoms:** Items 3, 8, 13, 16, and 24 from the SDQ survey are used to estimate the emotional symptoms score. Each question has a 3-point Likert answering option. By adding the scores of individual questions, the final sub-score is calculated. The variable is continuous; $Y_i^{ess} \in [0,10]$.
- **Conduct problem:** Items 5, 7, 12, 18, and 22 from the SDQ survey are used to estimate the conduct problem score. Each question has a 3-point Likert answering option. The formation of this sub-score is continuous; $Y_i^{cps} \in [0,10]$.
- **Hyperactivity:** Items 2, 10, 15, 21, and 25 are used to estimate the hyperactivity score. The variable is continuous; $Y_i^{hs} \in [0,10]$.
- **Peer problem:** Items 6, 11, 14, 19, and 23 from the SDQ survey are used to estimate peer problem scores. This variable is continuous; $Y_i^{pps} \in [0,10]$.

- **Prosocial:** Items 1, 4, 9, 17, and 20 from the SDQ survey are used to estimate the prosocial score. The variable is continuous; $Y_i^{ps} \in [0,10]$.

A-3.5. Validation of noncognitive skill measures

We have used validated psychometric tools to understand the impact of intervention on the noncognitive domain of children. However, Bengali translation of these scales are not validated, except for the strengths and difficulties questionnaire (SDQ). We took help from a professional psychologist to translate those measures where validated Bengali translations were not available. The following Table A8 exhibits validation studies of the tools and correlations of these measures with actual behaviors of the children.

Table A8. Noncognitive skill measures – Validation and Correlates

Scale	Validation study	Correlation with other factors
Leadership	Renzulli et al. (2002)	Leadership improves confidence and has a positive impact on communication, social problem solving, perseverance, cooperation, decision-making, conflict resolution, self-esteem, self-efficacy, cognitive development, etc. (See Karagianni & Jude Montgomery, 2018 for details).
Impulsivity Scale for Children (ISC)	Tsukayama et al. (2013)	Self-control is correlated with lower impulsivity and higher academic achievement (Angela L Duckworth, Taxer, Eskreis-Winkler, Galla, & Gross, 2019).
Grit	Angela Lee Duckworth and Quinn (2009)	Overall grit level and its two facets (consistency of interest and perseverance of effort) are positively associated with academic achievement (Lam & Zhou, 2019).
Growth mindset scale	(Dweck, 2013)	Growth mindset (the belief that intelligence is not fixed and can be developed) reliably predicts achievement across a national sample of students, including virtually all of the schools and socioeconomic strata in Chile (Claro, Paunesku, & Dweck, 2016).
Empathy Questionnaire for Children and Adolescents	Overgaauw et al. (2017)	Empathy significantly benefits mother-child relationship quality and life satisfaction (Chen-Bouck, Patterson, Qiao, & Peng, 2023).
Strengths and Difficulties Questionnaire (SDQ)	Goodman (1997)	SDQ has a negative impact on academic achievement (Agnafors, Barmark, & Sydsjö, 2021 ; Guo et al., 2021).

A-3.6. Homeschooling time (student)

Students' time investment in homeschooling because of the intervention was measured by asking the following questions to the mother:

- How much time in various academic activities (i.e., writing, reading, mathematics, etc.) does your child spend per day in minutes?
- To what extent does your child provide time for academic activities?

Two variables were constructed to underpin the homeschooling time of the students. These are-

- Student’s study time: This is the numeric value of daily study time in minutes. The variable is continuous; $Y_i^{ST} \in [0, Y_{max}^{ST}]$.
- Extent of study: This variable is constructed from a 5-point Likert-scale response; ‘none’ to ‘a great deal’. This variable is categorical; $Y_i^{STC} \in \{1..5\}$.

A-3.7. Homeschooling time (Mother or caregiver)

Mother’s or caregiver’s involvement in children’s educational activities was measured by asking the following questions to the mother:

- a. How much time do you or the caregiver give to your child in various academic activities (i.e., writing, reading, storytelling, mathematics, etc.) per day in minutes?
- b. To what extent do you or the caregiver provide academic support to your child?

Two variables were constructed to underpin the homeschooling time of the students:

- Caregiver’s time in home-schooling: This is the numeric value of daily time in minutes given by the mother or primary caregiver in education. The variable is continuous; $Y_i^{PI} \in [0, Y_{max}^{PI}]$.
- Extent of caregiver’s time: This variable is constructed from a 5-point Likert-scale response; ‘none’ to ‘a great deal’. This variable is categorical; $Y_i^{PIC} \in \{1..5\}$.

A-4. Social desirability bias

In impact analysis, social desirability bias may arise when the outcome variables rely on survey questions. More specifically, the experimenter demand effect can lead to over or underestimation of the treatment effect. Therefore, during the endline, we surveyed the parents using the short-form Crowne-Marlowe module (Crowne & Marlowe, 1960; Reynolds, 1982) and children using the Children’s Social Desirability (CSD-S) scale to measure their social desirability bias (Miller et al., 2015). Questions of these scales are presented in the following Table A9. To explore social desirability bias, we estimated the following interaction model:

$$Y_i = \alpha + \beta_1 T_i + \beta_2 SDB_i + \beta_3 (T \times SDB)_i + X_i' \theta + \varepsilon_i$$

where SDB_i is a dummy variable that takes a value of 1 if the social desirability score of parents or children is above the median.

Table A9. Social desirability scales – questionnaires

Scale	No	Questions	Desired answer	Answer
Marlowe-Crowne Scale (Reynolds)	1.	It is sometimes hard for me to go on with my work if I am not encouraged.	False	1) True 2) False
	2.	I sometimes feel resentful when I don’t get my way.	False	
	3.	On a few occasions, I have given up doing something because I thought too little of my ability.	False	
	4.	There have been times when I felt like rebelling against people in authority even though I knew they were right.	False	

	5.	No matter who I'm talking to, I'm always a good listener.	True	
	6.	There have been occasions when I took advantage of someone.	False	
	7.	I'm always willing to admit it when I make a mistake.	True	
	8.	I sometimes try to get even rather than forgive and forget.	False	
	9.	I am always courteous, even to people who are disagreeable.	True	
	10.	I have never been irked when people expressed ideas very different from my own.	True	
	11.	There have been times when I was quite jealous of the good fortune of others.	False	
	12.	I am sometimes irritated by people who ask favors of me.	False	
	13.	I have never deliberately said something that hurt someone's feelings.	True	
Children's Social Desirability Short (CSD-S) scale	1.	Have you ever felt like saying unkind things to a person?	No	
	2.	Are you always careful about keeping your clothing neat and your room picked up?	Yes	
	3.	Do you sometimes feel like staying home from school even if you are not sick?	No	
	4.	Do you ever say anything that makes somebody else feel bad?	No	
	5.	Are you always polite, even to people who are not very nice?	Yes	
	6.	Sometimes, do you do things you've been told not to do?	No	
	7.	Do you always listen to your parents?	Yes	1) Yes
	8.	Do you sometimes wish you could just play around instead of having to go to school?	No	2) No
	9.	Have you ever broken a rule?	No	
	10.	Do you sometimes feel angry when you don't get your way?	No	
	11.	Do you sometimes feel like making fun of other people?	No	
	12.	Do you always do the right things?	Yes	
	13.	Are there sometimes when you don't like to do what your parents tell you?	No	
	14.	Do you sometimes get mad when people don't do what you want them to do?	No	

Notes: This table lists the 13-item short form of the Crowne-Marlowe social desirability scale used for parents and the 14-item Children's Social Desirability (CSD-S) scale used for children.

A-5. Assessment test reliability

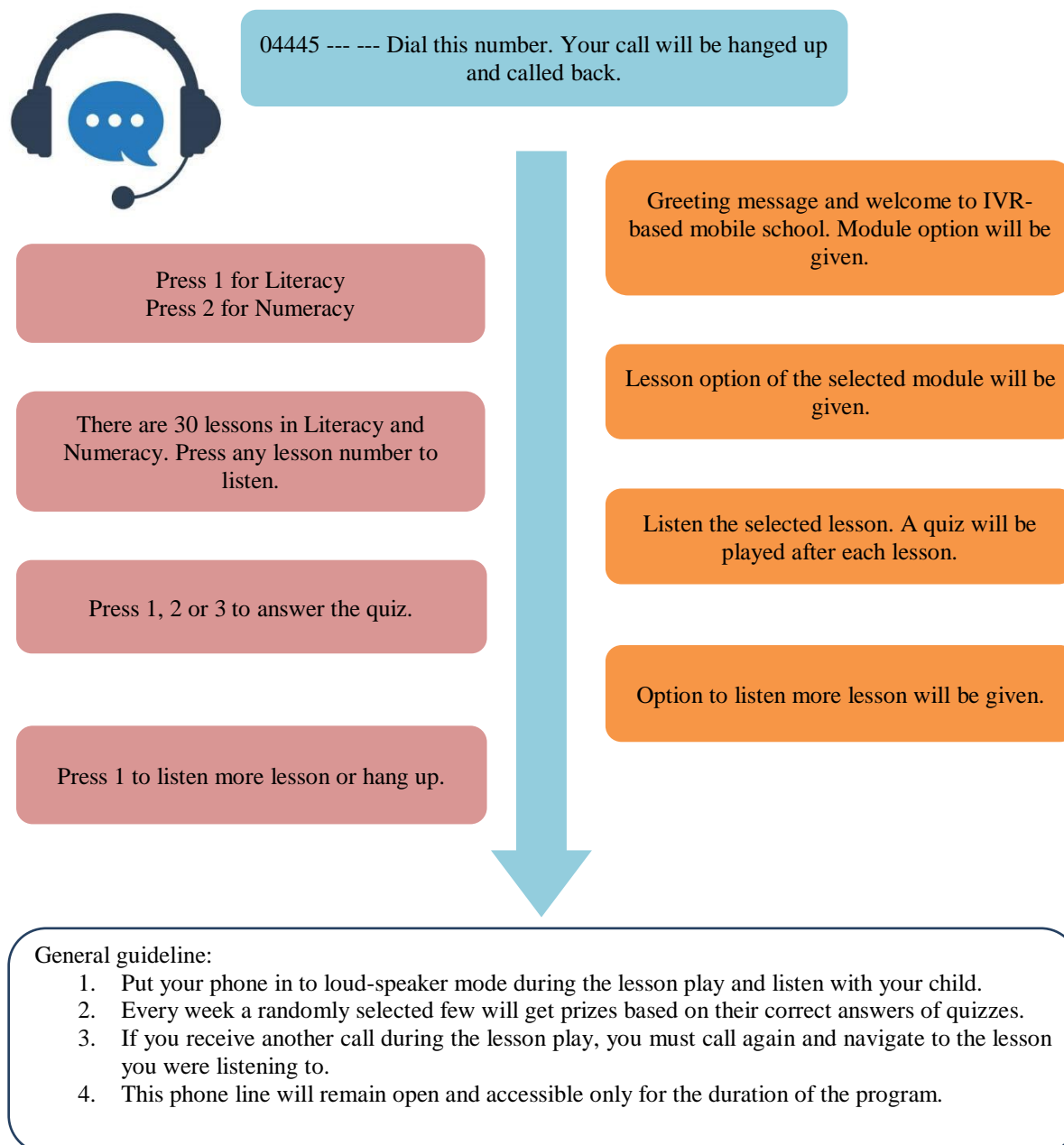
We designed our assessment test based on the curriculum designed by the National Curriculum and Textbook Board (NCTB) of Bangladesh. This curriculum has been rigorously tested and modified over the years based on the changing aptitude of the learners. We picked 15 questions from the latest version of the textbooks for the primary grades. These questions are exhibited in [Table A4](#). To verify the coherence of these selected items, we conducted some statistical tests. Firstly, we measured Cronbach's Alpha and McDonald's Omega. Results are exhibited in [Table B10](#) in the [Appendix B](#). Items of our assessment test have a high level of internal reliability and inter-item correlation. This indicates that assessment test questions are measuring the same underlying construct (literacy/numeracy).

Secondly, we constructed an adjusted assessment test score using a subset of items that appear to perform similarly between treatment and control groups. First, we converted all answers to binary choice (0 = incorrect and 1 = correct). Second, we fitted the answers to all questions to a 2-Parameter Logistic (2PL) Item Response Theory (IRT) model over the treatment and control groups, i.e., a constrained model. Third, we fitted the answers to a specific question to a hybrid 2PL IRT model where parameters can be varied across groups for that item, i.e., an unconstrained model. Fourth, we tested the two models (constrained vs. unconstrained) using

the Likelihood-Ratio test to check whether any item shows differential functioning across treatment and control groups. [Figure B8](#) exhibits the item characteristics curves for all 15 questions from the endline assessment tests and [Table B11](#) exhibits LR test statistics. Based on the LR test, it is evident that few items exhibit differential item functioning. Finally, we re-estimated the assessment test score by excluding these items. These adjusted treatment effects are presented in [Table B12](#), which indicates that the learning outcomes are improved after Differential Item Functioning (DIF) adjustment. Overall, our assessment test is stable, reliable, and coherent and the treatment was effective based on the DIF-adjusted test score.

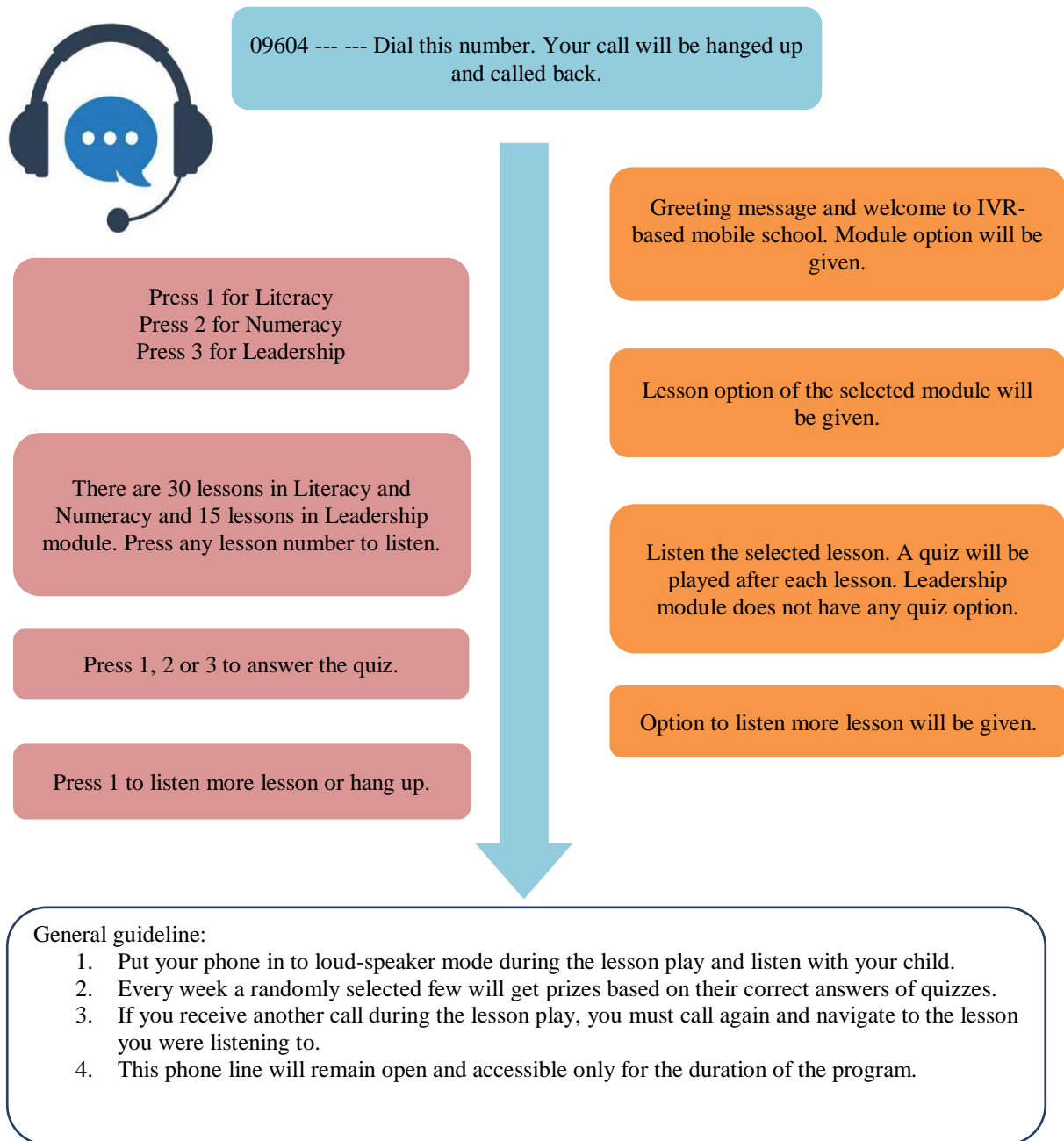
Appendix B: Additional Figures and Tables

Figure B1. IVR flow diagram of T1: Standard treatment group



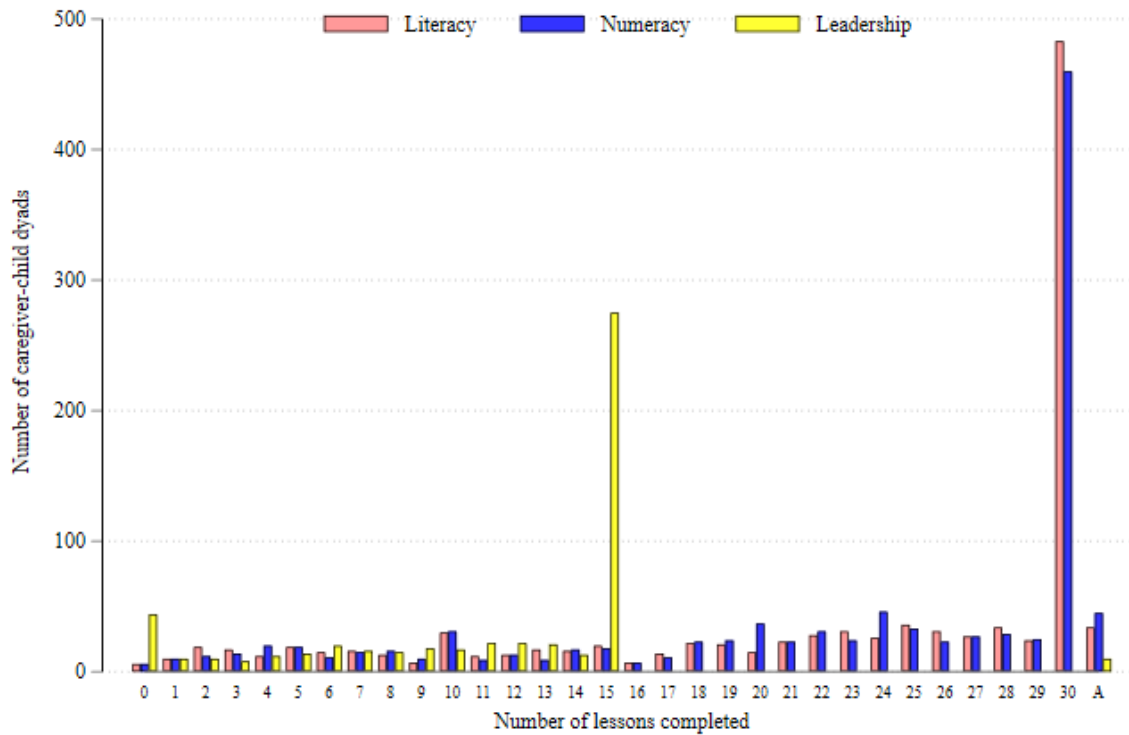
Note: This figure illustrates the IVR journey of the participating caregiver-child pairs in the Standard treatment group.

Figure B2. IVR flow diagram of T2: Extended treatment group



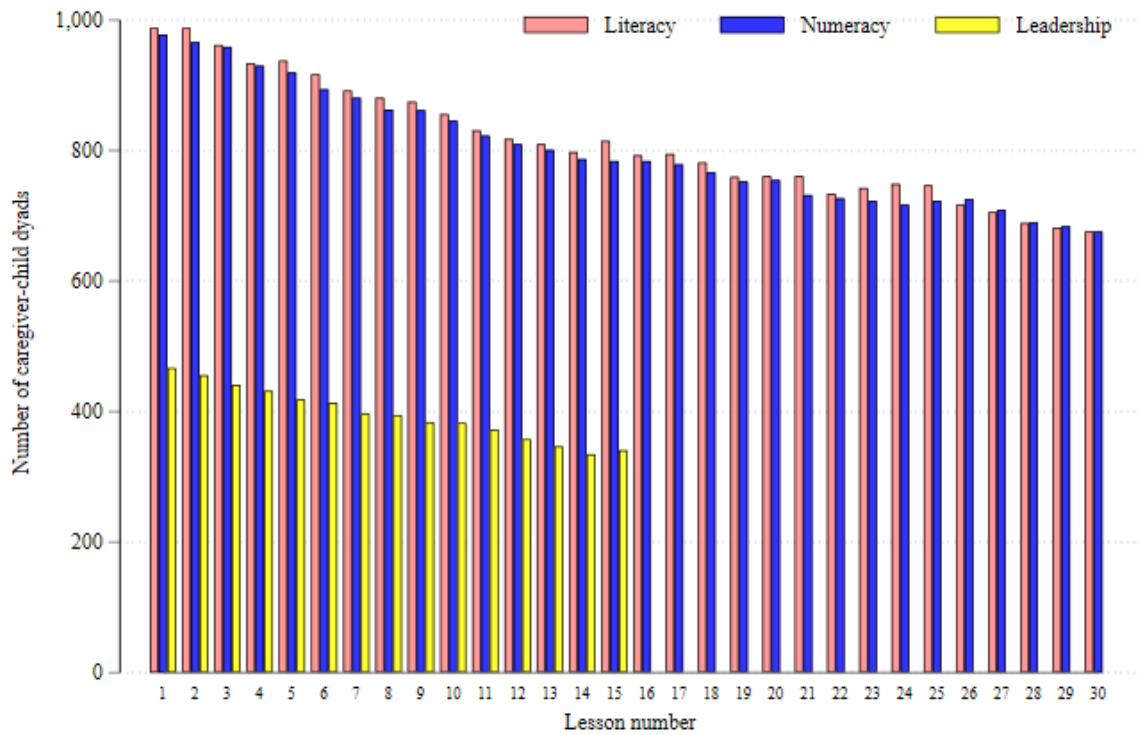
Note: This figure illustrates the IVR journey of the participating caregiver-child pairs in the Extended treatment group.

Figure B3. Distribution of total number of lessons completed by caregiver-child dyads, by module



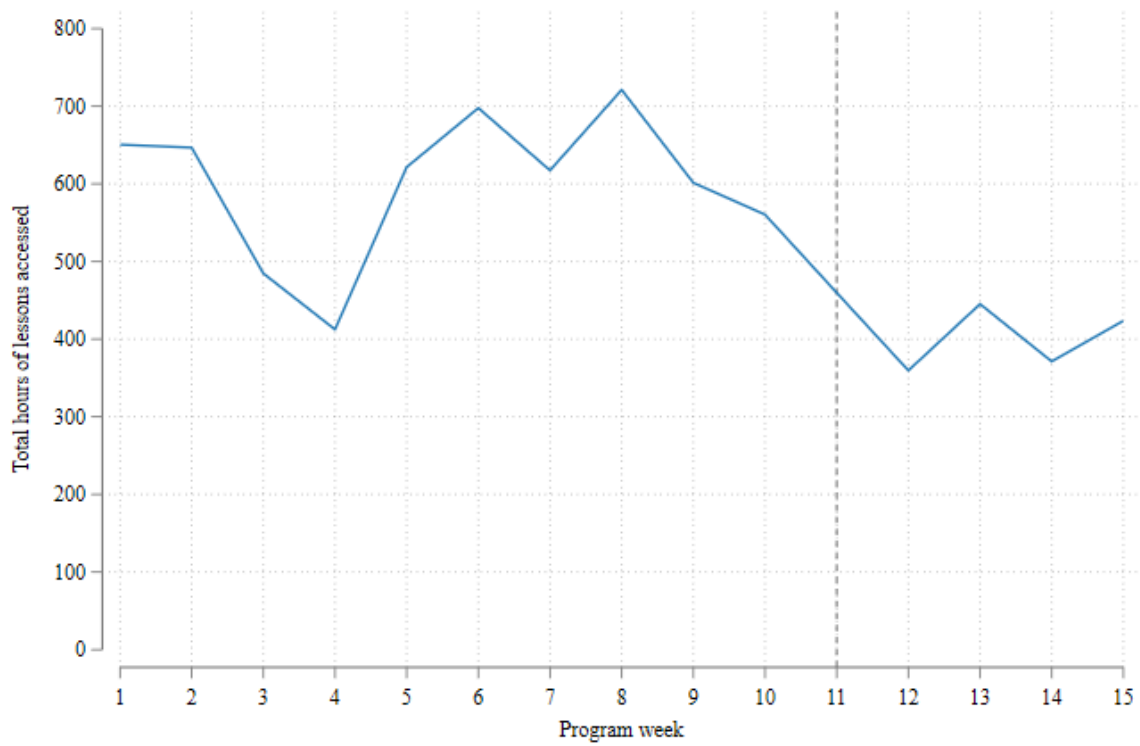
Note: This figure shows the distribution of total number of lessons completed by the caregiver-child dyads in different modules. In Literacy, Numeracy and Leadership, 6, 6 and 44 dyads did not listen to any of the lessons, respectively. On the contrary, 476, 454 and 276 dyads completed all the lessons for Literacy, Numeracy and Leadership, respectively. The Leadership module was offered only to the Extended group and there were 15 lessons. In our intervention, a total of 1,182 caregiver-child dyads received the treatment. However, a total of 1,107 dyads returned the lesson completion sheet. Among those who returned the lesson completion sheet, 35, 44, and 10 dyads indicated that they had ever completed the lessons for Literacy, Numeracy and Leadership, respectively, but they did not indicate which lessons they had completed and we could not match their registered phone numbers with the IVR-flow usage data to verify the lessons they completed. These dyads are reported in the “A” category. Among the remaining dyads who we also could not match their registered phone numbers with the IVR-flow usage data, 75 did not return the lesson completion sheet and 9 returned a blank sheet, and are therefore excluded from this figure.

Figure B4. Completion by caregiver-child dyads, by module and lesson number



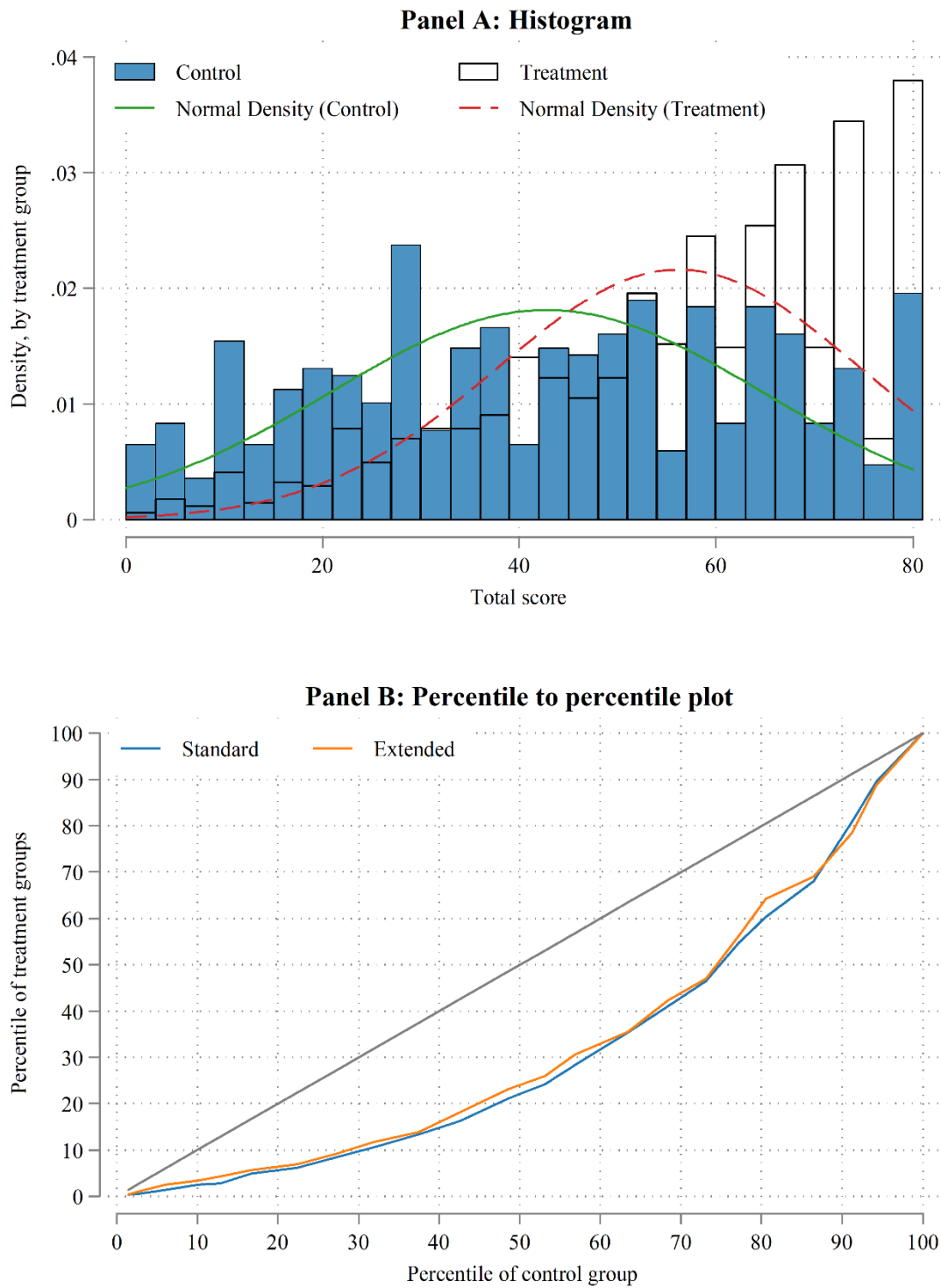
Note: This figure shows the attendance of caregiver-child dyads in various lessons across different modules. In the Leadership module, there were only 15 lessons.

Figure B5. Total hours of audio lessons accessed, by program week



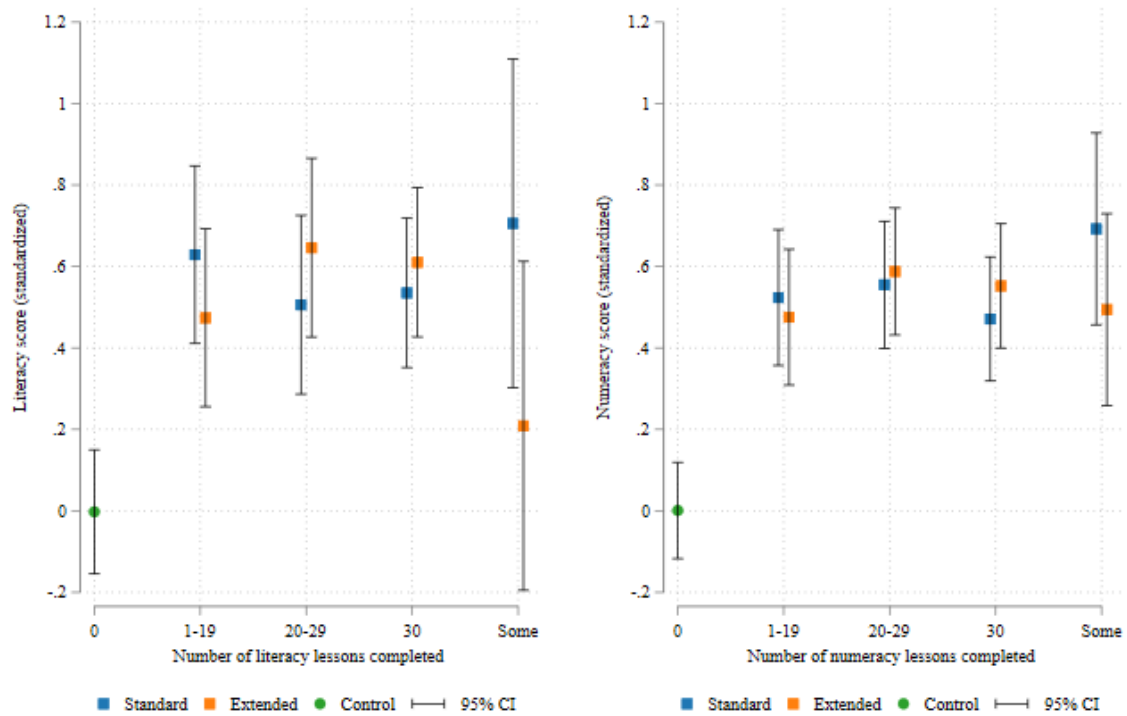
Note: This figure shows the total hours of audio lessons accessed by program week. In program week 10, the government announced that all schools would reopen for one to two days per week, starting from 12 September, 2021. This development occurred in the midst of program week 11.

Figure B6. Assessment-test score distribution, by treatment



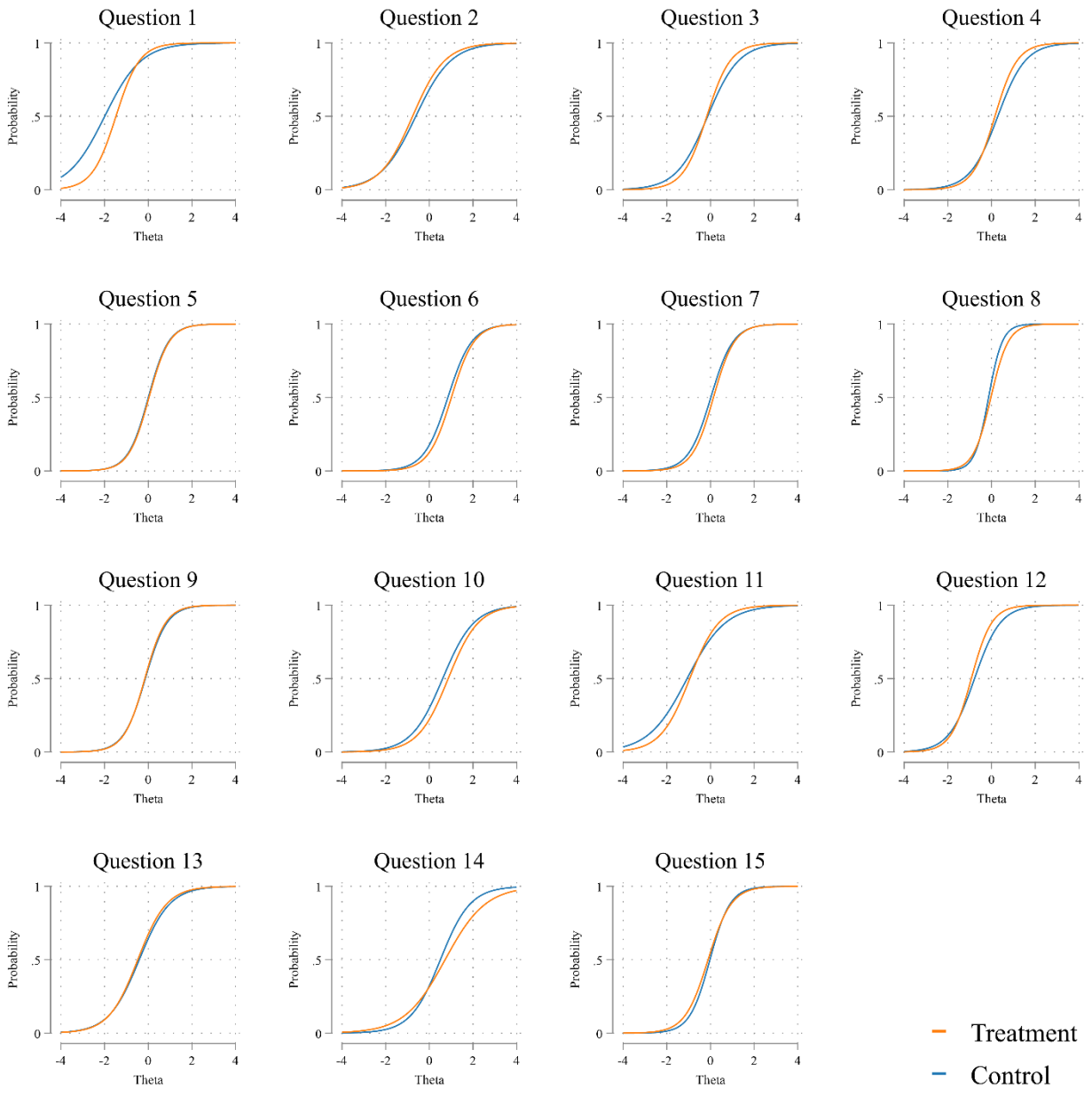
Note: This figure presents a comparison between test-score distributions. In Panel A, histograms of treatment and control groups are presented. Test score of the treatment group is left-skewed, which means treated children scored higher than control children. In Panel B, percentiles of treatment groups and control groups are displayed. This figure indicates that the 30th percentile of the treatment group distribution corresponds to the 60th percentile of the control group distribution.

Figure B7. Dosage responses in standardized literacy and numeracy scores



Note: This figure shows the standardized literacy scores and numeracy scores against the total number of lessons completed by the caregiver-child dyads by treatment status. In our intervention, a total of 1,182 caregiver-child dyads received the treatment and 581 caregiver-child dyads in the control group received no intervention at all. A total of 1,098 dyads in the treatment group returned the lesson completion sheet with usable information about the exact lessons completed. Among those who returned the lesson completion sheet indicating they had ever completed but without indicating the exact lessons completed are classified as having “Some” number of lessons completed. Among the remaining dyads in the treatment group, 75 did not return the lesson completion sheet, 12 did not complete the endline survey, 9 returned a blank completion sheet, and 6 reported to have never accessed any literacy or numeracy lessons, and are therefore excluded from this figure.

Figure B8. Differential Item Functioning (DIF)



Note: This figure exhibits the probability of answering each question correctly, by estimated ability (theta), and by treatment (pooled).

Supplementary Tables

Table B1. Survey attrition rate

Particulars	(1)		(2)		(3)		(4)	
	T1: Standard		T2: Extended		C: Control		Total	
	N	%	N	%	N	%	N	%
Not attrited	567	95.13%	562	95.90%	561	96.56%	1690	95.86%
Attrited	29	4.87%	24	4.10%	20	3.44%	73	4.14%
Total	596	100%	586	100%	581	100%	1763	100%

Note: This table reports the frequency of attrition at the endline survey. All 1763 children were approached for the endline survey and assessment. A total of 1690 child and mother dyads completed both the child assessment and parent survey. Pearson's Chi-squared test: $\chi^2 = 1.5063$, $P = 0.47$.

Table B2. Attrition, by treatment

Variables	(1) T1 & C	(2) T1 & C	(3) T2 & C	(4) T2 & C	(5) T1, T2 & C	(6) T1, T2 & C
T1: Standard treatment dummy	0.01 (0.01)	0.06 (0.17)	-	-	-	-
T2: Extended treatment dummy	-	-	0.01 (0.01)	-0.24 (0.20)	-	-
Any treatment Dummy	-	-	-	-	0.01 (0.01)	-0.09 (0.17)
Constant	0.03*** (0.01)	0.30* (0.14)	0.03*** (0.01)	0.30* (0.14)	0.03*** (0.01)	0.30* (0.14)
Controls	No	Yes	No	Yes	No	Yes
Interaction terms (treatment dummy × controls)	No	Yes	No	Yes	No	Yes
Observations	1,177	1,177	1,167	1,167	1,763	1,763
R-squared	0.00	0.13	0.00	0.11	0.00	0.11
Joint F-test <i>p-value</i> on characteristics	-	0.10	-	0.10	-	0.08
Joint F-test <i>p-value</i> on interactions	-	0.70	-	0.14	-	0.38

Note: All columns present estimates using OLS regressions, where the dependent variable is a dummy variable for attrition (1 if the parent-child dyad did not participate in the endline survey and assessment). T1 – Standard treatment; T2 – Extended treatment; and C – Control. The sample in columns 1 & 2 is parent-child dyads in T1 & C, the sample in columns 3 & 4 is parent-child dyads in T2 & C, and the sample in columns 5 & 6 is the full sample. Control variables are – 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. Robust standard errors clustered at the village level are in parentheses. *** $p < 0.005$, ** $p < 0.01$, * $p < 0.05$.

Table B3. Comparison between study sample and national sample survey

Variable	(1) BIHS Sample Mean	(2) BIHS Observation	(3) Study Sample Mean	(4) Study Observation
Gender (1 if Boy)	0.51 (0.50)	2742	0.48 (0.50)	1763
Access to private tuition (1 if yes)	0.48 (0.50)	2742	0.55 (0.50)	1763
School Type (1 if Govt. School)	0.74 (0.44)	2742	0.92 (0.28)	1763
Father's Age	41.59 (8.30)	1713	37.75 (5.91)	1763
Father's education (in years of schooling)	4.63 (3.96)	1713	5.95 (4.11)	1763
Mother's Age	34.22 (6.61)	2130	30.05 (4.86)	1763
Mother's education (in years of schooling)	5.52 (3.58)	2130	7.07 (3.36)	1763
No. of household member	4.67 (1.47)	2142	4.83 (1.51)	1763
Family income (in BDT/ month)	10739.13 (10249.95)	2142	11021.64 (5445.70)	1763
Access to TV (1 if yes)	0.38 (0.49)	2142	0.46 (0.50)	1763
Access to phone (1 if yes)	0.97 (0.17)	2142	0.99 (0.11)	1763
Homestead land (in decimal)	7.68 (9.16)	2142	9.69 (13.41)	1762
Religion (1 if Islam)	0.90 (0.30)	2142	0.77 (0.42)	1763
Household head's occupation (1 if agriculture)	0.44 (0.50)	1782	0.40 (0.49)	1077

Note: The Bangladesh Integrated Household Survey (BIHS) survey is a nationally representative survey conducted by IFPRI. For comparability, we focus on households from the BIHS survey that have children studying in the similar grade level as in our study. We also limit our selection to households from rural background. We use the 2018-19 round for this comparison table.

Table B4. Percentage of children who answered correctly, by question and treatment

Module	Question no	Type	(1)	(2)	(3)	(4)	(5)
			Marks assigned	Standard	Extended	Control	Total sample
Bangla literacy	1	Letter identification/ Reading	5	0.94 (0.01)	0.96 (0.01)	0.87 (0.02)	0.92 (0.01)
	2	Vocabulary	5	0.80 (0.02)	0.84 (0.01)	0.64 (0.03)	0.76 (0.01)
	3	Vocabulary/ Spelling	5	0.74 (0.02)	0.73 (0.02)	0.53 (0.03)	0.67 (0.01)
	4	Vocabulary/ Spelling	5	0.65 (0.03)	0.66 (0.02)	0.42 (0.03)	0.57 (0.01)
English literacy	5	Letter identification/ Reading	6	0.70 (0.03)	0.70 (0.02)	0.50 (0.03)	0.63 (0.01)
	6	Answering verbally to an English question	6	0.41 (0.03)	0.38 (0.03)	0.27 (0.02)	0.35 (0.01)
	7	Vocabulary	4	0.67 (0.02)	0.66 (0.02)	0.49 (0.03)	0.61 (0.01)
	8	Vocabulary	4	0.73 (0.02)	0.74 (0.02)	0.54 (0.04)	0.67 (0.01)
	9	Vocabulary	4	0.73 (0.03)	0.76 (0.02)	0.54 (0.03)	0.68 (0.01)
	10	Writing/ Spelling	6	0.45 (0.03)	0.45 (0.03)	0.35 (0.03)	0.41 (0.01)
Numeracy	11	Basic numbers	6	0.87 (0.02)	0.85 (0.02)	0.73 (0.02)	0.82 (0.01)
	12	Addition/ Sorting	6	0.89 (0.01)	0.91 (0.01)	0.71 (0.02)	0.84 (0.01)
	13	Addition/ Subtraction	6	0.78 (0.02)	0.79 (0.02)	0.60 (0.03)	0.72 (0.01)
	14	Number conversion	6	0.49 (0.03)	0.49 (0.03)	0.37 (0.03)	0.45 (0.01)
	15	Addition/ Multiplication	6	0.73 (0.03)	0.72 (0.02)	0.50 (0.03)	0.65 (0.01)
All	Number of questions answered correctly	-	10.59 (0.23)	10.65 (0.20)	8.06 (0.31)	9.77 (0.09)	

Note: This table presents the percentage of children answering each question correctly in the endline assessment test, by treatment groups (columns 2-4). Column 5 reports the percentage for the full sample. Robust standard errors clustered at the village level are in parentheses. All reported differences are statistically significant.

Table B5. Treatment effects on number of correct answers

Outcomes	(1) Control means	(2) Treatment effects - Standard	(3) Treatment effects - Extended
Panel A: Number of correct answers			
Bangla Literacy [4 Questions]	2.47 (0.09)	0.65*** (0.09)	0.72*** (0.08)
English Literacy [6 Questions]	2.68 (0.15)	1.00*** (0.16)	1.05*** (0.16)
Literacy [10 Questions]	5.15 (0.23)	1.65*** (0.24)	1.77*** (0.23)
Numeracy [5 Questions]	2.91 (0.10)	0.87*** (0.11)	0.88*** (0.11)
Total [15 Questions]	8.06 (0.31)	2.51*** (0.33)	2.65*** (0.31)
Panel B: Answered correctly to all questions (0/1)			
Bangla Literacy [4 Questions]	0.26 (0.03)	0.22*** (0.04)	0.24*** (0.03)
English Literacy [6 Questions]	0.11 (0.02)	0.07** (0.02)	0.09*** (0.03)
Literacy [10 Questions]	0.09 (0.02)	0.08*** (0.02)	0.08*** (0.02)
Numeracy [5 Questions]	0.21 (0.02)	0.15*** (0.03)	0.15*** (0.03)
Total [15 Questions]	0.06 (0.01)	0.05** (0.02)	0.05*** (0.02)

Note: In Panel A, Column 1 presents the mean of the control group children. Columns 2 and 3 present estimates using OLS regressions, with the dependent variable being the number of correct responses given by the children. In Panel B, similar OLS regressions are estimated, where the dependent variable is the binary indicator for children who answered all questions correctly in each of the components of the test. Control variables 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. *** p<0.005, ** p<0.01, * p<0.05.

Table B6. Heterogenous treatment effects on learning outcome, by quartiles

Variables	Standard Treatment			Extended Treatment		
	(1) Baseline score	(2) Household income	(3) Parental education	(4) Baseline score	(5) Household income	(6) Parental education
Treatment dummy	16.82*** (2.84)	12.65*** (3.21)	16.71*** (2.99)	13.97*** (3.04)	14.77*** (3.01)	18.15*** (2.67)
2nd Quartile	4.34* (2.06)	-0.88 (2.12)	5.48** (1.96)	4.34* (2.06)	-0.88 (2.12)	5.48** (1.96)
3rd Quartile	8.37* (3.21)	0.76 (2.83)	9.60*** (1.71)	8.37* (3.21)	0.76 (2.83)	9.60*** (1.71)
4th Quartile	8.58** (2.86)	5.56 (3.26)	20.73*** (1.74)	8.58** (2.86)	5.56 (3.26)	20.73*** (1.74)
Treat*2 nd Quartile	-3.15 (2.89)	2.00 (2.82)	-2.06 (3.16)	1.51 (3.04)	0.10 (2.57)	-5.03 (2.65)
Treat*3 rd Quartile	-6.57 (4.39)	-0.33 (4.26)	-4.77 (3.16)	-1.79 (3.99)	-2.97 (3.94)	-4.71 (2.57)
Treat*4 th Quartile	-5.18 (3.51)	0.21 (3.88)	-9.84** (3.22)	-1.15 (3.76)	-3.18 (3.77)	-10.37*** (2.71)

Note: This table presents the heterogeneous treatment effects of the intervention on the learning outcomes, by baseline learning and household characteristics listed in the column heading. Coefficients are estimated with OLS regressions. The same list of control variables is used as before. Baseline controls included: children's age, 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. *** p<0.005, ** p<0.01, * p<0.05.

Table B7. Parents' social desirability bias analysis.

Variables	T1: Standard			T2: Extended		
	(1) Above median	(2) Below median	(3) Interaction	(4) Above median	(5) Below median	(6) Interaction
Leadership	0.08 (0.09)	0.15 (0.14)	-0.08 (0.13)	0.14 (0.09)	0.13 (0.12)	0.01 (0.13)
Communication	0.03 (0.10)	0.18 (0.16)	-0.14 (0.16)	-0.00 (0.12)	0.06 (0.12)	-0.08 (0.14)
Planning	-0.01 (0.12)	0.15 (0.16)	-0.15 (0.16)	-0.02 (0.13)	0.14 (0.13)	-0.17 (0.15)
Emotional symptoms	-0.14 (0.10)	-0.19* (0.09)	0.05 (0.13)	-0.11 (0.09)	-0.16 (0.10)	0.04 (0.13)
Conduct problem	-0.10 (0.07)	-0.32** (0.12)	0.24 (0.13)	-0.13* (0.06)	-0.08 (0.11)	-0.02 (0.11)
Hyperactivity	-0.02 (0.08)	-0.40*** (0.10)	0.40*** (0.13)	0.05 (0.07)	-0.04 (0.12)	0.12 (0.12)
Peer problem	-0.14 (0.09)	-0.18 (0.10)	0.06 (0.14)	-0.01 (0.08)	-0.17 (0.12)	0.18 (0.14)
Prosocial	0.11 (0.08)	0.44*** (0.11)	-0.33*** (0.11)	0.11 (0.09)	0.14 (0.09)	-0.04 (0.11)
Student's study time	0.07 (0.12)	0.18 (0.10)	-0.08 (0.15)	0.10 (0.11)	0.02 (0.11)	0.06 (0.13)
Caregiver's time in homeschooling	0.15 (0.10)	0.30* (0.12)	-0.15 (0.14)	0.05 (0.10)	0.10 (0.11)	-0.07 (0.13)

Note: This table presents treatment effects adjusted for the social desirability bias of the parents. Only those outcome variables that are dependent on parents' survey answers are included in this analysis. We estimated the social desirability bias of the parents using the 13-item short form of the Crowne-Marlowe social desirability scale. Above median – dummy variable for above-median value of social desirability bias scale; interaction – interaction term between SDB and treatment dummy. Coefficients are estimated with OLS regressions. The dependent variable for each regression is listed in the row heading. The same list of control variables is used as before. Children's grade fixed effects are included in all regressions. Robust standard errors clustered at the village level are in parentheses. *** $p < 0.005$, ** $p < 0.01$, * $p < 0.05$.

Table B8. Child’s social desirability bias analysis.

Variables	T1: Standard			T2: Extended		
	(1) Above median	(2) Below Median	(3) Interaction	(4) Above median	(5) Below Median	(6) Interaction
Overall impulsivity	-0.18 (0.09)	-0.07 (0.12)	-0.13 (0.11)	-0.11 (0.10)	0.13 (0.13)	-0.26* (0.12)
Schoolwork impulsivity	-0.27** (0.10)	-0.10 (0.12)	-0.19 (0.11)	-0.19* (0.10)	0.08 (0.12)	-0.30* (0.13)
Interpersonal impulsivity	-0.05 (0.08)	-0.03 (0.12)	-0.04 (0.11)	-0.02 (0.09)	0.14 (0.11)	-0.17 (0.12)
Grit	0.15 (0.11)	0.07 (0.11)	0.07 (0.12)	0.14 (0.12)	0.07 (0.10)	0.05 (0.12)
Growth mindset	0.19 (0.13)	0.11 (0.13)	0.10 (0.13)	0.25* (0.12)	0.20 (0.12)	0.05 (0.14)
Affective empathy (Contagion)	-0.08 (0.11)	-0.13 (0.08)	0.04 (0.14)	-0.05 (0.12)	-0.18* (0.08)	0.12 (0.13)
Cognitive empathy (Understanding)	-0.15 (0.10)	-0.13 (0.10)	-0.00 (0.14)	-0.11 (0.09)	-0.12 (0.09)	-0.00 (0.12)
Prosocial motivation (Support)	-0.07 (0.12)	-0.02 (0.09)	-0.03 (0.14)	-0.05 (0.11)	-0.06 (0.07)	-0.00 (0.13)

Note: This table presents treatment effects adjusted for the social desirability bias of the children. Only those outcome variables that are dependent on children’s survey answers are included in this analysis. We estimated social desirability bias using the Children’s Social Desirability (CSD-S) scale. Above median – dummy variable for above-median value of social desirability bias scale; interaction – interaction term between SDB and treatment dummy. Coefficients are estimated with OLS regressions. The dependent variable for each regression is listed in the row heading. The same list of control variables is used as before. Children’s grade fixed effects are used in all regressions. Robust standard errors clustered at the village level are in parentheses. *** p<0.005, ** p<0.01, * p<0.05.

Table B9. Relationships between percent of completed lessons ever repeated and baseline characteristics

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Literacy	Numeracy	Literacy	Numeracy	Literacy	Numeracy
Less-educated parents	-0.03 (0.66)	-0.37 (0.62)	-0.01 (0.46)	-0.23 (0.43)	-0.14 (0.49)	-0.34 (0.46)
Lower baseline score	-1.25* (0.62)	-0.78 (0.55)	-0.90* (0.43)	-0.55 (0.38)	-0.91* (0.44)	-0.48 (0.40)
Lower household income	0.54 (0.72)	0.45 (0.51)	0.40 (0.51)	0.33 (0.35)	0.27 (0.53)	0.24 (0.37)
Observations	715	707	1012	1003	1012	1003
R-squared	0.016	0.017	0.060	0.047	0.012	0.012

Note: This table presents the associations between percent of completed lessons ever repeated 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. Columns 1 and 2 exclude individuals whose registered phone numbers cannot be matched to the IVR access data. Columns 3 and 4 include individuals whose registered phone numbers cannot be matched to the IVR access data, where we code their dependent variable as zero and include an indicator for these observations as a control variable. Columns 5 and 6 include individuals whose registered phone numbers cannot be matched to the IVR access data and the dependent variable for them is coded as zero. *** p<0.005, ** p<0.01, * p<0.05.

Table B10. Reliability of endline assessment test questions.

Subject	Alpha	Omega	Correlation	No of questions	No of students
All questions	0.86	0.86	0.28	15	1690
Literacy	0.82	0.82	0.31	10	1690
Numeracy	0.70	0.70	0.32	5	1690

Note: This table shows Cronbach's Alpha, McDonald's omega, and the inter-item correlation for the assessment instruments used in the endline.

Table B11. Differential Item Functioning (DIF) analysis.

Subject	Question number	Chi-square	P-value
Literacy	1	4.76	0.09
	2	3.73	0.15
	3	5.27	0.07
	4	5.34	0.07
	5	0.17	0.92
	6	4.85	0.09
	7	2.75	0.25
	8	6.99	0.03
	9	0.24	0.89
	10	7.77	0.02
Numeracy	11	3.64	0.16
	12	11.79	0.00
	13	0.99	0.61
	14	5.95	0.05
	15	1.76	0.42

Note: This table presents the likelihood ratio (LR) test statistics for differential item functioning (DIF) in various assessment tests questions. The null hypothesis is no DIF in respective items across treatment and control groups. All questions are converted to a dichotomous variable, 0 – wrong answer, 1 – correct answer.

Table B12. Differential Item Functioning (DIF) adjusted learning outcome.

Variable	T1: Standard		T2: Extended	
	Coefficient on treatment dummy	t-test <i>P-value</i>	Coefficient on treatment dummy	t-test <i>P-value</i>
Total score	0.62*** (0.08)	0.00	0.65*** (0.07)	0.00
Literacy	0.58*** (0.08)	0.00	0.51*** (0.07)	0.00
Numeracy	0.51*** (0.07)	0.00	0.52*** (0.06)	0.00

Note: This table presents DIF-adjusted learning outcomes. It was evident that out of fifteen questions, there was DIF in four questions. We re-estimate our learning outcomes excluding these questions and then estimate the standardized treatment effects. Coefficients are estimated with 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 used in all regressions. Robust standard errors clustered at the village level are in parentheses. *** $p < 0.005$, ** $p < 0.01$, * $p < 0.05$.

Table B13. Costing table

Item	Type	Unit	Cost/unit in BDT	Total
IVR platform development	Fixed	1	300,000	300,000
Household reach out	Variable	1182	150	177,300
Program management & training	Fixed	4 Months	50,000	200,000
Voice charges	Variable	1182	975	1,152,450
SMS charges	Variable	1182	18	21,276
Content development	Fixed	75	10,000	750,000
Total cost in BDT				2,601,026
Total cost in USD				32,512.8
Per child cost in USD				27.5