IARIW 2025

IARIW 2025

Thursday, October 2 & Friday, October 3

Delivering Remote Learning in Developing Countries Using a Low-tech Solution: Evidence from Nepal and Pakistan

Paul Glewwe
(University of Minnesota)
Zhigang Li
(Asian Development Bank)
Prashant Loyalka
(Stanford University)
Khandker Rahman
(BRAC University and Oxford University)
Uttam Sharma
(Independent Consultant)

Paper prepared for the IARIW–World Bank–UEB/VNU Conference on "Improving Well-being Measurement in Data-challenged Environments in Developing Countries for Better Evidence-based Policies" October 2-3, 2025

Parallel Session 4B Measuring Mobility
Time Slot: Friday, October 3, 8:00-9:30 AM

Delivering Remote Learning in Developing Countries Using a Low-tech Solution: Evidence from Nepal and Pakistan. *

September 2, 2025

Abstract

We conducted two randomized controlled trials in Nepal and Pakistan to test the effectiveness of an educational intervention designed to address learning losses among secondary school students resulting from prolonged school closures during the COVID-19 pandemic. Secondary school students were provided access to a series of pre-recorded math and English audio lessons delivered via basic-feature mobile phones using Interactive Voice Response (IVR) technology. One treatment arm (T1), the self-help group, offered access to IVR lessons for 24 weeks. A second treatment arm (T2), the assisted group, also offered access to IVR lessons for 24 weeks and, in addition, offered biweekly phone calls from a tutor to discuss the content covered in the IVR lessons. We find that the utilization of the IVR lessons was low in T1 and moderate in T2. The impact of the intervention on learning outcomes was small and statistically insignificant in both treatment arms in Nepal, but moderate to large impacts were found for the T2 arm in Pakistan. Turning to students' effort and non-cognitive skills, we again find little or no significant effects in Nepal, but for Pakistan, we find statistically significant impacts on students' effort for both arms and significant impacts on non-cognitive skills for the T2 arm.

JEL Classification Codes: I25, O15

*This study was pre-registered at the American Economic Association RCT Registry (ID: AEARCTR-0010507 and AEARCTR-0011078). Ethics approval for this project was received from Monash University Human Research Ethics Committee in 2022 (project number: 34569 (Nepal, August 2022) and 36624(Pakistan, January 2023)).

1 Introduction

Learning losses resulting from school closures are among the most severe global challenges to the medium and long-term recovery from COVID-19. On average, COVID-19-related school closures reduced learning by 0.16 SD, which is equivalent to 41% of a year of schooling lost in developed countries and 54% lost in developing countries (Dela Cruz, Adona, Molato-Gayares and Park, 2024).

The duration of school closures was particularly long in South Asia. In Nepal, schools were closed for 40 weeks and partially closed for another 48 weeks.[1] With the existing structural fault lines in its educational system, learning losses may be particularly acute in South Asia. The challenge of addressing learning loss lies in identifying interventions that are effective and scalable.

This study examines the efficacy of a low-tech remedial education intervention in a developing country context with the goal of producing a rigorously validated and scalable solution to address learning losses among secondary school students. The insights from this study will have broad applicability as the program can be implemented in any region where access to basic-feature mobile phones is widespread. In most of South Asia, including Pakistan and Nepal, basic feature phones are much more prevalent than smartphones, which presents an opportunity to create a cost-effective intervention that takes advantage of existing ICT networks in the region'.

We designed an educational intervention to address learning losses among secondary school students that have resulted from prolonged school closures during the COVID-19 pandemic. The intervention offered secondary school students access to a series of pre-recorded math and English audio lessons delivered via basic feature mobile phones using Interactive Voice Response (IVR) technology.

We conducted a randomized controlled trial to test the effectiveness of the intervention for three families of outcomes: student learning, time and effort spent on educational activities, and

[1] Source: UNESCO Institute for Statistics, "UNESCO map on school closures." https://en.unesco.org/covid19/educationresponse: (Last accessed: 3 June 2024).

aspirations and other non-cognitive skills. The study was carried out in 24 rural municipalities in Central and Eastern Nepal and 5 districts (Karachi Central District, Karachi East District, Karachi South District, Karachi West District and Korangi District) in the Karachi Division of Pakistan, all of which had limited access to educational resources.

The intervention offered 10th grade students access to 1440 minutes of pre-recorded audio lessons, covering remedial math and English material from grades 7 and 8.[2] One treatment arm (T1), the self-help group, was offered access to the IVR lessons for 24 weeks. A second treatment arm (T2), the assisted group, was also offered access to the IVR lessons for 24 weeks, but in addition received biweekly phone calls from a tutor to discuss the content covered in the IVR lessons.

After six months of the intervention, the use of IVR lessons was low in T1 and moderate in T2. On the extensive margin, 41.4 percent and 28.3 percent of the treated students listened to at least one IVR lesson in Nepal and Pakistan, respectively. In Nepal, 32.2 percent of the students in T1 and 50.6 percent in T2 listened to at least one IVR lesson. In Pakistan, participation was even lower: 17.1 percent in T1 and 39.5 percent in T2.

We estimate the intention-to-treat (ITT) effects of interventions. For both arms, we find small and statistically insignificant impacts on learning outcomes in Nepal. The treated students demonstrated math and English achievement similar to those in the control group, both in our assessment tests conducted shortly after the intervention and in the high-stakes school leaving exam taken six months later. Furthermore, the intervention had no measurable impact on the time and effort of the students devoted to educational activities, and no significant impacts on the non-cognitive skills of the students'.

In contrast, in Pakistan, we find a significant improvement in learning for the T2 group. There is a considerable 0.49 standard deviation (SD) increase in math and a 0.18 SD increase in English learning. We also find a significant increase in study effort in both arms. In addition, we find that students in the T2 group have higher aspirations and a higher locus of control.

[2] In Pakistan, we also included 9th grade students, particularly in schools with fewer 10th grade students attending schools.

Our findings have several policy implications. As low and middle-income countries (LMICs) grapple with addressing the learning crisis, further exacerbated by the COVID-19 pandemic, technology-based, low-cost, and scalable lesson delivery methods could be a potential solution. Our results suggest that, for secondary school students with consistent access to the Internet and sufficient time for remedial education, IVR-based delivery of pre-recorded lessons could be effective if combined with regular over-the-phone tutoring. In addition, IVR with the help of online tutors is also useful to enhance at least some non-cognitive skills.

Our paper contributes to two strands of literature. First, it contributes to the literature on interventions to address learning loss. Existing studies identify several potential education interventions, for example, telementoring (Angrist, Ainomugisha, Bathena, Bergman, Crossley, Cullen, Letsomo, Matsheng, Panti, Sabarwal and Sullivan, 2023b; Crawfurd, Evans, Hares and Sandefur, 2023; Hassan, Islam, Siddique and Wang, 2024; Schueler and Rodriguez-Segura, 2021), prerecorded IVR lessons (Wang, Vlassopoulos, Islam and Hassan, 2024), software or app-based solutions (Bourassa, 2022; Meeter, 2021; Ökördi and Molnár, 2022; Marques De Souza, Weissheimer and Buchweitz, 2022), online learning (Angrist, Aurino, Patrinos, Psacharopoulos, Vegas, Nordjo and Wong, 2023a; Clark, Nong, Zhu and Zhu, 2021), online tutoring (Carlana and Ferrara, 2021), and after-school in-person remediation programs (Singh, Romero and Muralidharan, 2022). The evidence from these interventions is mixed and the jury on what works and under what conditions is still out.

Second, we contribute to the literature on the learning crisis at the secondary school level. Existing evidence on telementoring and educational technology (EdTech) predominantly focuses on primary or pre-school education. To the best of our knowledge, this is the first study to examine remote complementary tutoring specifically for secondary school students. This project, along with companion studies conducted in other contexts, demonstrates that IVR-based solutions, when combined with moderate remote human involvement, can serve as a viable approach in some settings.

The remainder of the paper is organized as follows. Section 2 describes the intervention and

Section 3 describes the study design. We discuss data in Section 4 and the econometric strategy in Section 5. Section 6 presents the results. Section 7 concludes the paper.

2 The Intervention

IVR is an automated phone system technology that enables incoming callers to access pre-recorded instructions without speaking to an attendant (tutor). The system also allows callers to select lessons from a list of menu options using the phone keypad selection. Unlike television and radio programs, IVR lessons require only basic feature mobile phones and can be accessed at any time, allowing students to study at their convenience and pace.

We have developed and recorded a series of audio lessons (podcasts) that provide math and English instructions for 10th-grade students. These pre-recorded lessons are delivered through a toll-free IVR-based line. The intervention aims at 10th grade students, who are scheduled to take high-stakes end-of-school exams (the Secondary Education Examination, or SEE, in Nepal, and the Secondary School Certificate, or SSC, in Pakistan) in 2024 after completing grade 10.

The intervention was implemented between May 2023 and November 2023 in Nepal and from January 2024 to May 2024 in Pakistan.[3] During the intervention period, students had access to weekly IVR lessons for both subjects (math and English). Each treated student had access to approximately 1,440 minutes of pre-recorded IVR lessons throughout the intervention period. In addition, students received two text messages through the IVR system every week with reminders, lesson plans, and information on how to use the IVR system.

To address the concern that students might struggle to use the IVR lessons independently - whether due to limited motivation, lack of support, or difficulty grasping concepts - we introduced a second treatment arm. In this group, students received a weekly phone call from a tutor in addition to having access to the IVR lessons. The calls alternated between math and English tutoring, ensuring that the students spoke with their tutor for only one subject each week. During

[3]There were unavoidable delays in the commencement of field activities in Pakistan, primarily due to extended school closures caused by floods and the additional time required to identify a suitable IVR service provider capable of collaborating effectively with our local implementation partner, the Institute for Educational Development at Aga Khan University.

these calls, the tutors reviewed the content of the IVR lesson, reinforced key concepts, and helped resolve any questions the students had.[4]

There were notable implementation challenges in both Nepal and Pakistan, which we describe in greater detail in Appendix B and Appendix C. In Pakistan, the enumerators conducted orientation sessions prior to the intervention to familiarize students with the IVR system. After registration, participants received confirmation text messages once access was granted. However, many students reported ignoring these texts, as four-digit number messages are commonly perceived as promotional. Further confusion arose because the four-digit sender ID used for texts differed from the IVR number that students had to dial to access audio lessons. To mitigate this, a dedicated staff member called students in the IVR-only arm to explain the system use, while mentors and tutors in the IVR-plus-tutor arm reinforced these instructions and provided additional support. In addition, resource packs were developed for both subjects in both Pakistan and Nepal, which contain tasks and exercises to complement audio lessons. Insights from qualitative interviews revealed that some students completed these tasks with the help of private tutors or family members, particularly in Pakistan, without necessarily engaging in IVR lessons themselves.

In Nepal, we made similar attempts to support student participation, but implementation challenges proved to be more severe. The difficult terrain and frequent student absences limited the feasibility of in-person orientation, while weak access to the household phone and intermittent connectivity further hindered communication between students and tutors. Parents were often unavailable or unable to make calls and headteacher skepticism in some schools reduced institutional support. As noted in our qualitative work, these barriers contributed to a lower overall use of IVR in Nepal despite the same system design and distribution of the resource packs.

These implementation experiences underscore the importance of considering both technical design features (e.g. sender ID consistency, text delivery) and local learning ecosystems (e.g., household support, parental involvement, and school buy-in) when evaluating the effectiveness of low-tech interventions.

[4] In addition, the tutors made a weekly reminder call to each student to reinforce engagement with the IVR system and study materials.

3 Study Design

In Nepal, we partnered with Edukhabar, a multimedia organization that specializes in the development and dissemination of educational materials through electronic and print media, to implement interventions and conduct surveys. Edukhabar first appointed subject experts in both math and English as subject coordinators to customize and monitor the content of the audio lessons and twice-weekly text messages. The coordinators and the Edukhabar team then selected experienced teachers from community schools (publicly funded schools managed at the community level) to record the audio lessons. For the second intervention (T2), the tutors were professional teachers with at least two years of teaching experience and were selected by Edukhabar. In Nepal, male and female tutors were recruited and randomly assigned to the treated students. Edukhabar recruited 8 English tutors (4 male and 4 female) and 8 math tutors (3 male and 5 female). In Nepal, male and female tutors were assigned to teach students without regard to the student's gender; male and female tutors taught both boys and girls. The study was carried out in 24 municipalities located in 9 districts in 3 different provinces, Bagmati, Koshi, and Madhesh, in central and eastern Nepal.

In Pakistan, we partnered with the Institute for Education (IED) of Aga Khan University. They are responsible for implementing data collection activities and delivering quality data according to expectations and protocols outlined in this TOR, within the timeframe defined by the World Bank (WB) Research Team (RT). The institute is well known in Pakistan for its research on education and has conducted numerous studies in that field, ranging from pedagogical innovations to large-scale program evaluations of education interventions.

The IED appointed subject experts in mathematics and English as coordinators to customize and monitor the content of the audio lessons, as well as the biweekly text messages. The coordinators, together with the IED team, selected qualified university graduates to record the audio lessons. Tutors were also recruited by the IED based on eligibility criteria that required a university degree and prior teaching experience. Male and female tutors were hired and then randomly assigned to the treated students according to the following rule: male tutors were assigned exclusively to male

students, while female tutors worked with both male and female students. In total, IED recruited more than 50 tutors, approximately half in mathematics and half in English, although only eight were men. The study was carried out in five districts of the Karachi Division in Pakistan. Karachi Central, Karachi East, Karachi South, Karachi West, and Korangi.

This study implemented a two-stage cluster randomized controlled trial with three experimental arms (two treatment arms and one control arm). In Nepal, all secondary schools in the 24 municipalities were included in the study, which includes 223 schools. All 223 schools are co-ed. In Pakistan, the study encompasses 225 single-sex secondary schools distributed in five districts in Karachi.

We used a two-stage randomization design, illustrated in the appendix Figures A1 and A2. In the first stage, we randomly assigned the schools to one of the three experimental arms (without stratification). T1 (self-help group), T2 (assisted group) and T3 (control group).

In Nepal, in the second stage, we randomly selected a subgroup of 10th grade students within the T1 and T2 schools to receive the treatment, while surveying and collecting data from all 10th grade students in the 223 schools. The second stage of within-school randomization is stratified by gender to ensure that within each school, male and female students have similar chances of being selected for treatment. In general, 50% of the students in the 10th grade in the T1 and T2 schools were selected for treatment. In practice, we selected up to 27 students from each T1 and T2 school based on the total number of 10th grade students at the school.[5] This process created some variation in the share of treated students between schools (owing to variations in the total number of 10th grade students at each school), which we later used to estimate the effects of spillovers within the intervention schools.

In Pakistan, we randomly selected 20 students from the ninth and 10th grades to survey (in

[5]The randomization process ensured that around 50% of T1 and T2 students are selected for treatment and that the largest schools will have no more than 27 students selected. Specifically, for schools that had 12 or fewer 10th graders, all 10th graders were selected for treatment. We selected 13 students from schools that have 14 to 20 students in the 10th grade, 14 from schools with 21 to 29 students, 16 from schools with 30 to 39 students, 18 from schools with 40 to 49 students, 20 from schools with 50 to 59 students, 23 from schools with 60 to 69 students, and 27 from schools with 70 or more students in the 10th grade. The process resulted in an average of 14.3 students selected from each treatment group school. The stratified randomization process also ensured that, within each school, male and female students have similar chances of being selected for treatment.

all schools) and approximately 15 of them were assigned to receive treatment (in schools T1 and T2). Our local partner, IED, decided to include ninth-grade students due to fewer and irregularly attending grade 10 students in many schools.[6]

In T2, the tutors were randomly assigned to the treated students, resulting in random variations in the gender concordance between the tutors and the treated students. In secondary analyses, we will explore these variations to study the heterogeneous effects of interventions by gender concordance between the tutor and the student.

The three experimental arms are as follows:

a) **T1: The self-help group** consists of 74 schools, with 1,057 students selected to receive treatment, in Nepal, and 75 schools, with 1061 students selected to receive treatment, in Pakistan. The selected students received information on how to access IVR-based lessons and how IVR lessons could improve educational outcomes (such as higher scores on SSC exams). The selected students also received two text messages per week through the IVR system with reminders, lesson plans, and information on how to use the IVR system. Selected students could access the IVR system only by using a registered phone number. The rest of the students in the T1 schools did not receive any intervention or had access to the IVR system.

b) **T2: The assisted group** consists of 74 schools, with 1,060 students selected to receive treatment, in Nepal, and 75 schools with 1,058 students selected in Pakistan. The selected students received the same IVR-related information and text messages twice a week. In addition, students in T2 also received weekly phone calls from a tutor, alternating by subject: one week focused on math and the following week on English. This ensured that each student spoke to a tutor about one subject each week and covered both subjects over a two-week cycle. The rest of the students in the T2 schools did not receive any intervention and did not have access to the IVR system.

c) **T3: The control group** consists of 75 schools in each country with 2,170 students in Nepal and 1,590 students in Pakistan. This group of students did not receive any intervention. Since no student in this group received any intervention (unlike schools T1 or T2, where some students

[6]The content in grades 9 and 10 is similar to that of the students taking the SSC exam in grade 10, which includes materials from both grades 9 and 10.

received intervention while the rest remained untreated), this group constitutes the pure control group for this study.

We test the following hypotheses:

H1: Access to IVR lessons increases students' learning outcomes, increases students' time and effort spent on educational activities, and increases students' aspirations and other non-cognitive skills.

H2: Providing over-the-phone tutoring in addition to access to IVR lessons increases students' learning outcomes, increases students' time and effort spent on educational activities, and increases students' aspirations and other non-cognitive skills.

H3: The effects on all three categories of results (learning outcomes, time and effort spent on educational activities, aspirations, and other non-cognitive skills) are greater in T2 than in T1. That is, providing over-the-phone tutoring in addition to access to IVR lessons generates additional benefits for all outcomes.

4 Data

4.1 Data Collection

We conducted two rounds of face-to-face surveys with students and their caregivers. The baseline survey in Nepal was administered during November-December 2022, approximately six months before the start of the interventions. In Pakistan, the baseline survey was conducted in November-December 2023, immediately preceding the launch of the intervention in January 2024. The baseline survey included a student survey, a headteacher survey that collected basic school-level information, and student assessment tests for math and English. The headteacher survey collected information such as the qualifications and experience of the headteachers', the number of teachers and students in the school, and basic questions about school facilities such as the construction materials for the school buildings and the availability of libraries and latrines. The student survey included detailed questions on the time spent by students' in educational and other activities,

aspirations, hope, social-emotional and psychological status, as well as information on household assets, household composition, and the demographic, educational and occupational background of the parents. of the student'. The baseline assessment tests for math and English were customized for this study and designed by the subject experts on the national curriculum of the respective countries. These exams were administered in a regular classroom environment.

There was a significant time gap between our baseline survey and the start of the intervention in Nepal – the baseline survey was conducted when students were in 9th grade, while the intervention only started 6 months later after the start of a new academic year.[7] We therefore collected an updated list of students who were still enrolled in the sample schools in May 2023, shortly before the start of the intervention. The updated pre-intervention student roster includes 6,427 students in total (see Appendix Figure A1 for the distribution across treatment arms). We consider these 6,427 students who participated in our baseline survey and remained enrolled in schools before the start of the intervention as our baseline sample in the impact assessment.

In Nepal, caregivers were not surveyed at baseline. However, at the end of the study, we surveyed both the students and their principal caregivers about the time that the students' spent in educational and other activities, aspirations, hope, social-emotional and psychological status. The caregivers' survey also included questions about caregivers' participation in and aspirations for their children's education, parents' demographic, educational, and occupational background, and other household information such as household assets and household composition.

The end-line survey was conducted in December 2023 in Nepal, shortly after the intervention ended. In Pakistan, the end-line survey took place in July 2024. We also administered assessment tests for both.

In both countries, the enumerators for the baseline and endline surveys were recruited and trained by local partners. They were selected from a pool of professional data collectors. They received project-specific training prior to each round of data collection.

In addition to the assessment tests, we collected official results from the SEE exams in Nepal,

[7]We started the baseline survey early in anticipation of the potential difficulties of surveying schools and students in the mountainous regions and the various religious and school holidays between January and April.

which the students took in May 2024. We use the results of the SEE exam as an additional outcome to measure the effects of interventions on student' learning outcomes.

We also collected data on the amount of time each student spent listening to the IVR lessons from the IVR line provider. We also collected data on the actual amount of time the over-the-phone tutors spent talking to each student. These data allow us to construct measures of the utilization rate of IVR lessons and over-the-phone tutoring sessions. In addition, the end-of-course survey included questions asking the treated students to assess the usefulness of the interventions.

4.2 Baseline Characteristics and Randomization Balance

The random assignment achieved balance on the observable characteristics. We present the baseline balance tests in Tables A1 and A2 of Appendix A.

All Nepali schools are public and co-educational. On average, Nepalese students are slightly older than Pakistani students. In Nepal, the average years of formal schooling are 5.6 years for fathers and 3 years for mothers, while in Pakistan the average years of education of parents are 3.9 and 3.6 years, respectively. Among the students, 43% are male in Nepal and 37% are male in Pakistan. More students in Pakistan aspired to complete a BA degree (67%) than in Nepal (44%).

5 Empirical Strategy

5.1 Estimating the Direct Effects of the Interventions

We start by estimating the direct effects of the interventions, which are defined as the intent-to-treat (ITT) impact of the interventions on treated students (i.e. those who are offered access to the IVR or IVR plus phone tutor treatment). To estimate the direct effects, we will run the following OLS regression on a sample consisting of all *treated* students in treatment schools and all students in control schools (that is, we compare the treated students to the pure control students and *omit the untreated students* in treatment schools from this regression):

$$y_{ist} = \alpha_0 + \beta_1 T1_s + \beta_2 T2_s + \gamma y_{is0} + X'_{i0}\theta + Z'_{s0}\delta + \epsilon_{ist} \quad (1)$$

where y_{ist} is the outcome of interest of the individual i at school s measured at time t. $T1_s$ and $T2_s$ are indicators of being in a school in the T1 or T2 group. We condition on the baseline value of the outcome variable y_{is0}, when available, to increase the statistical power (McKenzie, 2012). X_{i0} is a vector of individual and household characteristics measured at baseline, including student's age, sex, father'and mother's years of education, household income, household religion (a dummy indicating religious minority), number of household members, and the number of children in the household. Z_{s0} is a vector of school characteristics measured at baseline, including class size, school size and its square, and district-fixed effects. Standard errors are clustered at the highest level of randomization, which is the school level.

The coefficients of interest are β_1 and β_2 which estimate the intent-to-treat (ITT) effects – an average of the causal effects of being offered treatment on the outcome variable.

In addition to the ITT effects, we also estimate treatment-on-the-treated (TOT) effects using program utilization rates. We construct two separate measures of the utilization rate for treated students in the T1 group. First, we create a dummy variable of whether the student called the IVR line at all. Second, we create another continuous measure: the percentage of IVR lessons that a student actually listened to.[8] We then run an IV regression using T1 and T2 assignments as instruments for these program utilization variables to estimate the TOT (ATT) effects of the interventions.

5.2 Estimating Within-School Spillover Effects

We also estimate the within-school spillover effects of interventions on untreated students in treatment schools (under the assumption of no cross-school spillovers). To estimate spillover effects, we run equation 1 on a sample consisting of all *untreated* students in treatment schools and all

[8] In the PAP, we stated that this percentage measure would be calculated as "the ratio between the actual minutes a student spent listening to the IVR lessons and the total length of the lessons." However, in practice, while most students did not access all 144 IVR modules, many listened to a selected number of modules multiple times. As a result, some students' listening time exceeded the total length of the lessons, even though they did not access all the modules. Hence, we created several modified measures to capture the intensive margin of usage of IVRs. In the main measure presented in Figure 1, the percentage of IVR lessons is the percent of modules the student has ever accessed. We also tested alternative measures such as the total number of minutes spent listening to IVR lessons and the number of times each student called the IVR system; the results of these measures are presented in Table A3.

students in control schools (that is, we compare untreated students with pure control students and omit treated students in treatment schools from this regression). With the alternative sample, the coefficients β_1 and β_2 estimate the indirect (spillover) effects of being in a treatment school in untreated students. This specification averages spillover effects for the two interventions; it is likely that there is variation in these effects across schools, in part due to variation in the proportion of treated students across these schools.

As the proportion of treated students at each school is determined by the total number of 10th graders in each school, this creates exogenous variation in treatment saturation rates across treatment schools. To allow spillover effects to vary by treatment saturation, we estimate a modified version of equation 1 replacing the indicators $T1_s$ and $T2_s$ with continuous variables $P1_s$ and $P2_s$ that measure the proportion of treated students in each treatment school (among the 10th graders). The coefficients on $P1_s$ and $P2_s$ measure the indirect effects of being an untreated student in a treatment school as the proportion of treated students increases from 0% to 100%.

5.3 Accounting for Multiple Hypotheses

As we consider a number of outcomes and our survey instruments often include several questions related to a single domain of outcomes, we account for multiple hypotheses by aggregating outcomes in the same domain into an aggregate index or by choosing a focal variable. We then used a family adjustment of *p-value* in the indices and focal variables.We construct inverse covariance-weighted index variables following the procedure proposed by Anderson (2008).[9]

The list of indices and focal variables will be included in the adjustment of the family p-value. Specifically, for the list of indices and focal variables, we calculate sharpened false discovery rate (FDR) q values following Benjamini, Krieger and Yekutieli (2006) as described in Anderson (2008). This method applies the Benjamini and Hochberg (1995) (BH) correction in two stages and provides better power than the standard BH correction method (Anderson, 2008).

[9]Specifically, the indices are constructed in three steps: (1) each individual variable in the domain is normalized by subtracting the mean of the control group and dividing by the standard deviation of the control group; (2) weights are generated from the inverted covariance matrix of all normalized variables in the domain; (3) the weighted sum of the normalized variables is then rescaled so that the control group is mean-zero and unit-variance.

6 Results

6.1 Take-up

The take-up of the IVR intervention was low in T1 and moderate in T2. Figure 1 illustrates the distribution of IVR lesson utilization rates, and the table in Appendix A3 presents the take-up statistics by treatment arm. Among the treated students, that is, the students who were offered the treatment, 41. 4% accessed the IVR lesson at least once (32.2% in T1 and 50.6% in T2) in Nepal. The average student listened to 2.7% of the lessons (just less than 4 lessons) and spent 15.2 minutes listening to math lessons and 13.5 minutes in English lessons. In Pakistan, acceptance rates were much lower, with 28. 3% of the students listening to at least one IVR lesson, 17. 1% of the students of T1, and 39. Only 5% of T2 at least listened to an IVR lesson. However, students who took up the IVR at least once in Pakistan spent more time listening to the IVR: 35.7 minutes in math and 27 minutes in English.

The take-up rates were higher in T2 than in T1 in both countries. The average student treated in T2 listened to twice the share of lessons and spent twice as much time as the average T1 student in both countries.

6.2 Intent to Treat Effects

Tables 1 and 2 present the direct effects of ITT of interventions on the list of prespecified primary outcomes. We find that in Nepal the effects of the interventions are small and statistically insignificant for both T1 and T2 (0.062 SD for math in T1 and 0.011 SD for math in T2). The impact on study time and effort and our primary non-cognitive skills results are all statistically insignificant[10] These null results are consistent with the low utilization of the IVR program in Nepal. However, in Pakistan, we find a significant and positive increase in T2 in both math and English: a 0.49 SD increase in math and a 0.18 SD increase in English. We also see a significant improvement in study effort in both arms (although this increase in effort does not translate into learning gain for T1).

[10]We received study time data from Pakistan with a long delay due to administrative error by the data collection team, which we have not yet analyzed and therefore, this outcome is not presented in this report.

We also find an increase in aspiration and locus of control for the T2 arm.

Turning to secondary outcomes, in Nepal, we find a statistically significant and moderate impact on the psychological well-being index, which is the inverse of depression, stress, and anxiety levels. The results (column 17 of Table 3) suggest that the psychological well-being index improved by 0.147 SD in T1 and 0.155 SD in T2. In Pakistan, we find an increase in confidence in both arms, but no significant increase in psychological well-being (Table 4).

6.3 Treatment of the Treated Effects and Dose-Response Effects

We find similar results in the TOT analyses, which we present in Tables 5-10. We find that while there is no impact on learning in Nepal, both the use of IVR and the increased use of IVR improve the psychological well-being of students. The TOT impacts on schools of math achievement are positive, although still statistically insignificant. In Pakistan, the TOT analysis shows that the effects on T2 learning are more pronounced. Interestingly, the study effort for T1 is significantly greater than that of T2, and yet we do not see any impact of T1 on learning.

To further access the impact of more IVR use on learning outcomes, we perform simple and exploratory dose-response analyzes to assess the impact of the IVR intervention conditional on the amount of IVR use. These results are presented in Figure 2. The graph suggests that the intervention had a significant impact on learning outcomes among treated students who used more than 10% of math lessons and those who used more than 25% of English lessons in Nepal. However, as shown in Figure 1, only 7. 4% of the treated students listened to more than 10% of the math lessons and 1.9% listened to more than 25%.; for English less lessons, 6.8% of the treated students listened to more than 10% of the lessons and 1. 5% listened to more than 25%. In Pakistan, there is a significant impact in math for any nonzero usage. In English, more than 10% use leads to better learning[11] Because the dose was not randomized, the relationship is not causal. We presented this analysis for completeness, and the observed pattern, though it can not be attributed as causal, is consistent with existing evidence in the literature. This observed correlation suggests that low

[11] which could happen, for example, if students in T1 who know the problems going through the IVR materials or contents might try hard to spend more time, but maybe cannot help themselves.

utilization rates may be the main reason for the overall low effects of the intervention in Nepal. One potential implication of this relationship is that the IVR lessons could perhaps be effective in improving learning among students who have the time and resources to use the lessons. There are a number of reasons for low usage which include poor connectivity, lack of access to electricity or dedicated phones for educational purposes, and limited understanding of the method's benefits. We lack direct data on these variables, yet the heterogeneity analysis shows no effects by parental income or education, suggesting that barriers may be widespread across groups.

In Nepal, several interrelated barriers contributed to the low adoption of IVRs: weak institutional capacity of the implementing partner to manage outreach and follow-up; limited trust and recognition at the community level; and coordination challenges with local schools, some of which viewed the program as competing with their existing activities. In addition, household-level constraints such as shared phone access, limited digital literacy, and post-COVID fatigue further reduced students' ability and motivation to engage. These restrictions were especially binding in the self-guided model (T1), where there was minimal external accountability or personalized support.

In Pakistan, IVR usage in an extensive margin was moderately higher than in Nepal, about 25-30% of students accessed meaningful content, but it was still insufficient to generate strong impacts on its own. Students in the assisted arm (T2) listened to more content than those in the self-help group (T1), suggesting that the presence of a tutor helped maintain engagement. However, several challenges remained: the implementing partner had limited experience with large-scale tech-based interventions; technical issues such as network interruptions and call delivery failures undermined reliability; and sociocultural barriers, particularly for girls, reduced access to phones and opportunities to participate. Many students, especially in rural areas and flood-affected areas, lacked consistent electricity or phone ownership, and parents often did not prioritize additional learning. In both countries, adoption was shaped by a combination of logistical, institutional, and social factors, highlighting the difficulty of scaling self-guided EdTech models in resource-constrained settings. In Appendix B and Appendix C, we provide a detailed explanation based on

the experience and insights of our local partners.

6.4 Within-School Spillover Effects

We find weak evidence of spillover to untreated students, which we report in Tables 11-12. As expected, the spillover effects on learning outcomes are all small and statistically insignificant in both countries. However, untreated students in Pakistan show a greater effort to study, a finding absent in Nepal. We also see that untreated students in T2 schools experience an improvement in psychological well-being in Nepal. The effect is larger and more precise when the concentration of treated students in the school increases. The spillover effect is 0.161 SD and statistically significant when we look at the extensive margin of treatment (that is, the impact of being in a school with classmates in the T2 arm). On the other hand, in Pakistan, untreated students in treatment school experience insignificant impacts on non-cognitive skills such as aspiration and hope; the corresponding coefficient estimates are mostly statistically insignificant. We observe a positive and statistically significant effect on growth mindset among T2 students in Pakistan.

6.5 Heterogeneity Analysis

In Nepal, we do not find any heterogeneity in the effect of treatment on learning outcomes with respect to sex (Table 13). Although we find that male students in the T1 group exert less effort, the total effect size cannot be distinguished from zero. There is weak evidence that T2 had a positive (but not statistically significant) impact on the aspiration of male students, but not on female students. We also find that the impact of interventions on psychological well-being is greater in women than in men. The net effect of interventions on the psychological well-being of male students is statistically insignificant.

We do not find any evidence of heterogeneity with respect to family income or father's education in Nepal (Tables 14 and 15). However, there is some heterogeneity according to the level of education of the mother. The impacts of T2 on study time, study efforts, and the hope index are negative (although not statistically significant) among students whose mothers had no formal schooling, but positive (also not statistically distinguishable from zero) among students whose

mothers had some schooling. However, the impacts of T2 on the locus of control and psychological well-being are positive and at least marginally significant among students whose mothers did not receive formal schooling, but smaller in magnitude among students whose mothers had some education (Table 16).

In Pakistan, improvement in learning in T2 is concentrated in male students. The total impact of T2 for men is significantly different from zero, unlike that for female students (Table 17). Interestingly, the study effort is significantly different from zero for all T1 students and female T2 students, but not for male T2 students. There is a concentration of impact by household income (Table 18), suggesting that endowment plays a role in mitigating learning loss. However, there is no discernible heterogeneity from parental education (Tables 19 and 20).

7 Conclusion

We designed and implemented a low-tech educational intervention using IVR technology to address learning losses among secondary school students in resource-constrained settings. Drawing on two randomized controlled trials conducted in Nepal and Pakistan, we tested the effectiveness of IVR with and without over-the-phone tutoring over a 24-week period.

In Nepal, we observed low IVR utilization and correspondingly limited improvements in learning outcomes, reflecting the challenges with student participation in the self-guided model. In Pakistan, while IVR usage was modest, the assisted treatment arm (T2), which combined IVR with remote tutor support, produced substantial improvements in academic performance, particularly in mathematics and English. This contrast underscores the importance of supplementing remote learning technologies with meaningful human interaction.

Beyond cognitive outcomes, estimates from both countries offer promising evidence of improvements in non-cognitive skills, such as aspirations, hope, growth mindset, and psychological well-being, especially among female students and those from disadvantaged backgrounds (for example, students whose mothers never received formal education). These effects were more pronounced in the assisted treatment arms, suggesting that the presence of a supportive adult, even

remotely, can foster student confidence and motivation.

The experience in both contexts highlights the recurring barriers to effective implementation of EdTech. The low adoption in Nepal and parts of Pakistan reflected institutional inexperience in technology-enabled delivery, infrastructure constraints (e.g., unreliable electricity, phone access), socioeconomic hardship, and cultural norms that limit girls' participation with technology. These challenges were particularly acute for self-help models without sustained support. At the same time, the success of the Pakistan T2 intervention, facilitated by a strong implementing partner with the capacity to recruit high-quality tutors, demonstrates the potential of hybrid models that integrate technology with personalized motivational support.

These findings yield three broad lessons. First, the effectiveness of remote learning interventions is not only dependent on content access but also on delivery capacity, local credibility, and support structures that promote participation. Second, hybrid models, where even modest technological interventions are complemented by human engagement, can be effective in settings where fully in-person education is not feasible (in this case, in-person tutoring was not feasible). Third, our findings suggest a nuanced relationship between student motivation and learning. In Nepal, the intervention improved psychological well-being and, for some groups, increased study effort, yet these changes did not produce measurable learning gains. In contrast, in Pakistan, T2 produced substantial improvements in learning, suggesting that effective delivery and support structures can translate participation into measurable results. This pattern mirrors broader evidence, for example, 'Teaching at the Right Level' programs and recent remote learning RCTs (Angrist et al., 2023b; Banerjee, Banerji, Berry, Duflo, Kannan, Mukherji, Shotland and Walton, 2016; Hassan et al., 2024), showing that motivation alone is not enough; learning tools must also match the baseline skills of the students and be delivered in ways that allow comprehension and practice. For remote learning to succeed, interventions must, therefore, pair efforts to foster motivation with content and delivery mechanisms tailored to the specific needs of the target population.

Looking ahead, our results highlight the need for further research on the conditions under which remote learning technologies can scale effectively. This includes understanding what drives

student uptake and participation, especially when traditional in-person instruction is disrupted by seasonal factors, natural disasters, or chronic teacher shortages. Future experimentation should also examine the relative cost-effectiveness of IVR-tutor hybrids compared to standard face-to-face instruction by trained teachers. As education systems grapple with growing inequality and climate-related disruptions, designing resilient and inclusive learning models remains a critical global priority.

References

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495. Number: 484.

Angrist, Noam, Elisabetta Aurino, Harry A. Patrinos, George Psacharopoulos, Emiliana Vegas, Ralph Nordjo, and Brad Wong, "Improving Learning in Low- and Lower-Middle-Income Countries," *Journal of Benefit-Cost Analysis*, 2023, *14* (S1), 55–80. Number: S1.

_ , Micheal Ainomugisha, Sai Pramod Bathena, Peter Bergman, Colin Crossley, Claire Cullen, Thato Letsomo, Moitshepi Matsheng, Rene Marlon Panti, Shwetlena Sabarwal, and Tim Sullivan, "Building Resilient Education Systems: Evidence from Large-Scale Randomized Trials in Five Countries," Working Paper 31208, National Bureau of Economic Research, Cambridge, MA 2023.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton, "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of âTeaching at the Right Levelâ in India," Working Paper 22746, National Bureau of Economic Research, Cambridge, MA 2016.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive Linear Step-up Procedures That Control the False Discovery Rate," *Biometrika*, September 2006, *93* (3), 491–507. Number: 3.

_ and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, January 1995, *57* (1), 289–300. Number: 1.

Bourassa, Benjamin J., "Investigating the Impact of the COVID-19 School Closures and Computerized Reading Instruction in COVID and Pre-COVID Times on Elementary Student Literacy Achievement." Ed.D. Dissertation, Concordia University Wisconsin 2022.

Carlana, Michela and Eliana La Ferrara, "Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic," 2021. Publisher: EdWorkingPapers.com.

Clark, Andrew E., Huifu Nong, Hongjia Zhu, and Rong Zhu, "Compensating for Academic Loss: Online Learning and Student Performance During the Covid-19 Pandemic," *China Economic Review*, August 2021, *68*, 101629.

Crawfurd, Lee, David K. Evans, Susannah Hares, and Justin Sandefur, "Live Tutoring Calls Did Not Improve Learning During the Covid-19 Pandemic in Sierra Leone," *Journal of Development Economics*, September 2023, *164*, 103114.

Cruz, Nina Ashley Dela, Ann Jillian Adona, Rhea Molato-Gayares, and Albert Park, "Learning Loss and Recovery from the COVID-19 Pandemic: A Systematic Review of Evidence," Technical Report, Asian Development Bank March 2024.

Hassan, Hashibul, Asad Islam, Abu Siddique, and Liang Choon Wang, "Telementoring and Homeschooling During School Closures: a Randomised Experiment in Rural Bangladesh," *The Economic Journal*, March 2024, p. ueae014.

McKenzie, David, "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, November 2012, *99* (2), 210–221. Number: 2.

Meeter, Martijn, "Primary school mathematics during the COVID-19 pandemic: No evidence of learning gaps in adaptive practicing results," *Trends in Neuroscience and Education*, December 2021, *25*, 100163.

Ökördi, Réka and Gyöngyvér Molnár, "Computer-Based Intervention Closes Learning Gap in Maths Accumulated in Remote Learning," *Journal of Intelligence*, August 2022, *10* (3), 58. Number: 3.

Schueler, Beth E. and Daniel Rodriguez-Segura, "A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya," 2021. Publisher: EdWorkingPapers.com.

Singh, Abhijeet, Mauricio Romero, and Karthik Muralidharan, "Covid-19 Learning Loss and Recovery: Panel Data Evidence from India," Technical Report w30552, National Bureau of Economic Research, Cambridge, MA October 2022. Issue: w30552.

Souza, Juliana G. Marques De, Janaina Weissheimer, and Augusto Buchweitz, "Well Played! Promoting Phonemic Awareness Training Using EdTechâGraphoGame BrazilâDuring the COVID-19 Pandemic," *Brain Sciences*, November 2022, *12* (11), 1494. Number: 11.

Wang, Liang Choon, Michael Vlassopoulos, Asad Islam, and Hashibul Hassan, "Delivering Remote Learning Using a Low-Tech Solution: Evidence from a Randomized Controlled Trial in Bangladesh," *Journal of Political Economy Microeconomics*, July 2024, pp. 000–000.

Figures

Figure 1: Distribution of IVR lesson utilization rates among treated students in Nepal and Pakistan

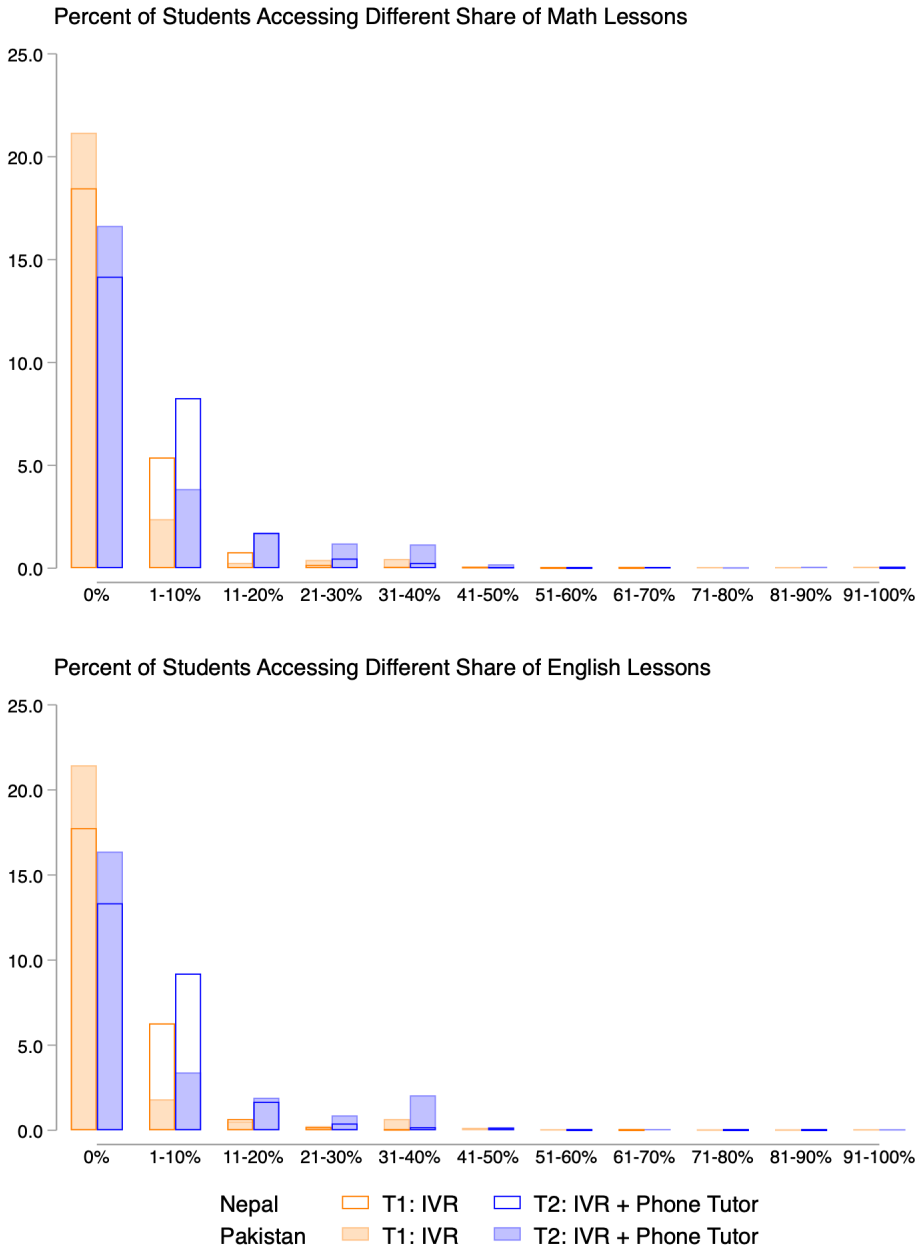
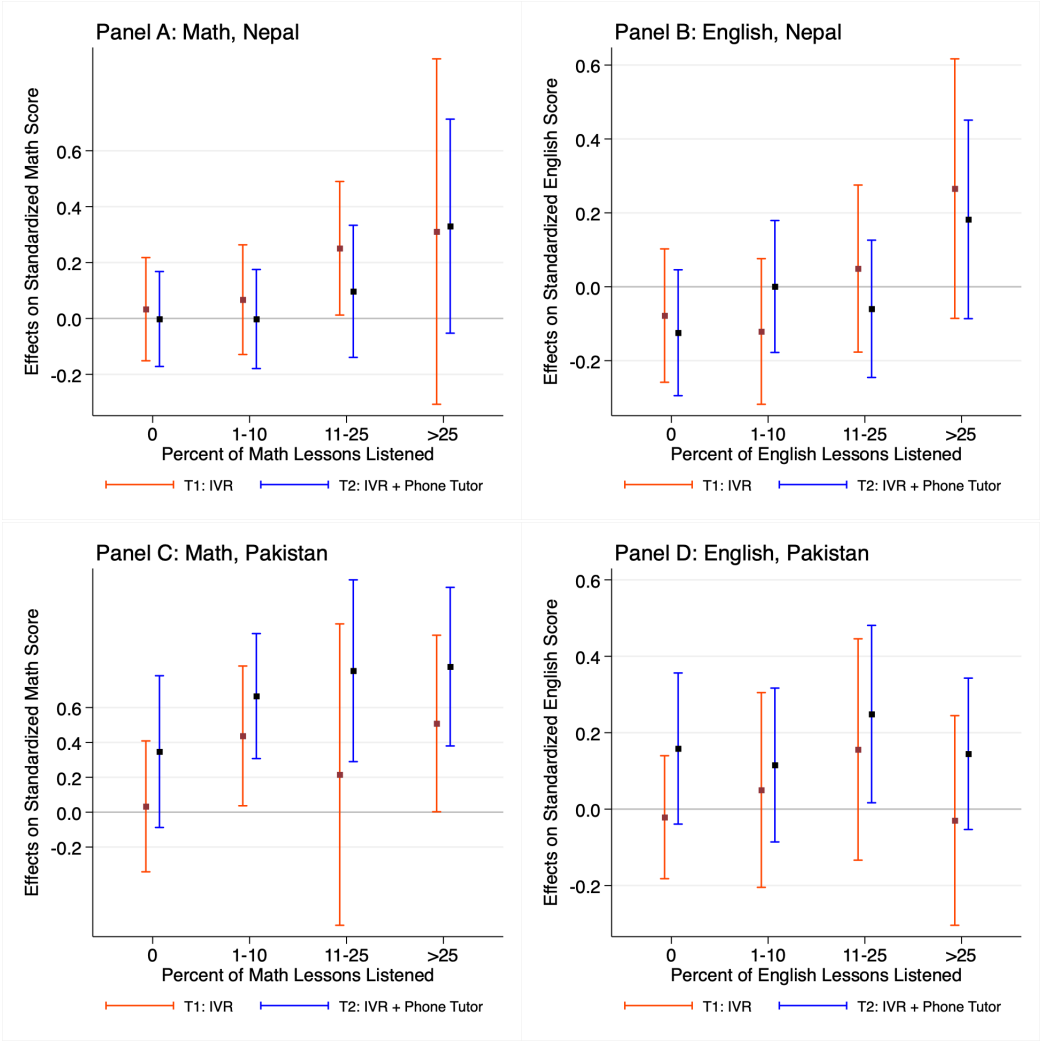


Figure 2: Dosage response in learning outcomes



Notes: Each whisker is a separate regression comparing a selected group of treated students to students in the pure control group. Treated students are divided into groups based on their treatment arm and the percent of math or English IVR lessons they listened to.

Tables

Table 1: Direct effects of the intervention on pre-specified **primary** outcomes in Nepal

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome family:	Learning Outcomes		Time and Effort Indices		Non-cognitive Skills Indices		
	Standardized Scores		Study	Study			Locus of
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control
T1: Self-help	0.062	-0.076	0.061	-0.025	0.069	-0.014	-0.048
	(0.090)	(0.087)	(0.105)	(0.070)	(0.071)	(0.068)	(0.053)
	{1.000}	{1.000}	{0.976}	{0.976}	{1.000}	{1.000}	{1.000}
T2: Assisted	0.011	-0.064	0.132	-0.037	0.063	-0.034	0.066
	(0.080)	(0.081)	(0.103)	(0.076)	(0.073)	(0.082)	(0.058)
	{1.000}	{1.000}	{0.976}	{0.976}	{1.000}	{1.000}	{1.000}
P-value: T1=T2	0.585	0.887	0.505	0.873	0.943	0.805	0.070
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). FDR-adjusted p-values (i.e., sharpened FDR q-values) that control for the FDR across primary outcomes within each family of outcomes are reported in curly brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Direct effects of the intervention on pre-specified **primary** outcomes in Pakistan

Outcome family:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Learning Outcomes		Time and Effort Indices		Non-cognitive Skills Indices		
	Standardized Scores		Study	Study			Locus of
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control
T1: Self-help	0.110	0.001	N.A.	0.681***	-0.038	-0.052	0.003
	(0.181)	(0.079)		(0.070)	(0.036)	(0.070)	(0.009)
	{0.569}	{0.979}			{0.579}	{0.579}	{0.931}
T2: Assisted	0.487**	0.180**	N.A.	0.373***	0.080**	0.055	0.214***
	(0.198)	(0.088)		(0.066)	(0.038)	(0.075)	(0.018)
	{0.064}	{0.070}			{0.101}	{0.579}	{0.001}
P-value: T1=T2	0.016	0.021		0.000	0.002	0.075	0.000
Observations	3,347	3,347		3,347	3,094	3,094	3,094

Notes: Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 218). FDR-adjusted p-values (i.e., sharpened FDR q-values) that control for the FDR across primary outcomes within each family of outcomes are reported in curly brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Direct effects of the intervention on pre-specified **secondary** outcomes in Nepal

	(1)	(2)	(3)	(4)	(5)	
Outcome family:	Learning Outcomes – Standardized Math Scores for Each Content Area					
Outcome variable:	Numbers & operations	Geometry	Ratio & statistics	Algebra & equations	Set theory	
T1: Self-help	0.034	0.116	0.092	0.027	-0.097	
	(0.085)	(0.095)	(0.078)	(0.081)	(0.080)	
T2: Assisted	-0.086	0.046	0.052	0.047	-0.059	
	(0.087)	(0.080)	(0.075)	(0.079)	(0.085)	
P-value: T1=T2	0.186	0.473	0.630	0.807	0.646	
Observations	4,017	4,017	4,017	4,017	4,017	

	(6)	(7)	(8)	(9)	(10)	(11)
Outcome family:	Time and Effort					
Outcome variable:	Study time by subject (min. per day)			Household chores & work		Leisure time
	Math	English	Other Subjects	Self-report (min. per day)	Caregiver rpt. (hrs. per week)	Caregiver rpt. (hrs. per week)
T1: Self-help	3.853	3.267	3.044	-4.847	-0.546	-0.237
	(2.728)	(2.550)	(4.467)	(9.306)	(0.526)	(0.440)
T2: Assisted	3.475	3.849	0.281	-13.455	0.206	0.247
	(2.790)	(2.789)	(5.061)	(9.980)	(0.600)	(0.416)
P-value: T1=T2	0.909	0.858	0.613	0.425	0.212	0.299
Observations	4,017	4,017	4,017	4,017	3,984	3,984
Control mean	60.0	48.3	95.8	264.5	8.1	4.6
Control S.D.	38.0	33.8	63.3	172.4	6.1	5.7

	(12)	(13)	(14)	(15)	(16)	(17)
Outcome family:	Non-cognitive Skills Indices					
Outcome variable:	Self-discipline	Confidence	Emotional Strength	Caregiver's Aspiration	Growth Mindset	Psychological Well-being
T1: Self-help	-0.008	0.010	0.001	0.002	0.059	0.147**
	(0.074)	(0.073)	(0.100)	(0.081)	(0.084)	(0.064)
T2: Assisted	-0.096	-0.079	0.026	-0.005	-0.015	0.155**
	(0.072)	(0.067)	(0.090)	(0.073)	(0.082)	(0.070)
P-value: T1=T2	0.263	0.237	0.802	0.931	0.415	0.906
Observations	4,017	4,017	3,984	3,984	4,017	4,017

Notes: Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. "Study time by subject" variables (cols. 6-8) are self-reported and measured in minutes per day. "Household chores and work" includes time spent on household chores (such as cooking and cleaning) and helping parents with income-generating activities (such as farming and tailoring). All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Cols. 1-5 control for the standardized baseline math score. Cols. 6-9, 12-13, and 16-17 control for baseline value of the outcome variable. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Direct effects of the intervention on pre-specified **secondary** outcomes in Pakistan

	(1)	(2)	(3)	(4)	(5)	
Outcome family:	Learning Outcomes – Standardized Math Scores for Each Content Area					
Outcome variable:	Numbers & operations	Geometry	Ratio & statistics	Algebra & equations	Set theory	
T1: Self-help	0.108	0.122	0.106	-0.132	0.002	
	(0.101)	(0.103)	(0.115)	(0.097)	(0.093)	
T2: Assisted	0.144	0.434***	0.353***	0.028	0.185*	
	(0.111)	(0.102)	(0.121)	(0.110)	(0.104)	
P-value: T1=T2	0.676	0.000	0.017	0.069	0.032	
Observations	3,347	3,347	3,347	3,347	3,347	

	(6)	(7)	(8)	(9)	(10)	(11)
Outcome family:	Time and Effort					
Outcome variable:	Study time by subject (min. per day)			Household chores & work		Leisure time
	Math	English	Other Subjects	Self-report (min. per day)	Caregiver rpt. (hrs. per week)	Caregiver rpt. (hrs. per week)
T1: Self-help	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
T2: Assisted	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
P-value: T1=T2						
Observations						
Control mean						
Control S.D.						

	(12)	(13)	(14)	(15)	(16)	(17)
Outcome family:	Non-cognitive Skills Indices					
Outcome variable:	Self-discipline	Confidence	Emotional Strength	Caregiver's Aspiration	Growth Mindset	Psychological Well-being
T1: Self-help	0.278***	0.403***	-0.129**	0.063	0.112*	0.009
	(0.070)	(0.069)	(0.050)	(0.053)	(0.060)	(0.016)
T2: Assisted	-0.047	0.212***	-0.174***	0.115**	0.090	-0.009
	(0.068)	(0.057)	(0.044)	(0.057)	(0.064)	(0.015)
P-value: T1=T2	0.000	0.012	0.283	0.364	0.712	0.240
Observations	3,347	3,094	3,094	3,094	3,094	3,094

Notes: Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. "Study time by subject" variables (cols. 6-8) are self-reported and measured in minutes per day. "Household chores and work" includes time spent on household chores (such as cooking and cleaning) and helping parents with income-generating activities (such as farming and tailoring). All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Cols. 1-5 control for the standardized baseline math score. Cols. 13-17 control for baseline value of the outcome variable. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 218). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: TOT effects specification I & II (Nepal)

Outcome family:	(1) Learning Outcomes	(2)	(3) Time and Effort	(4)	(5) Non-cognitive Skills Indices	(6)	(7)	(8)	(9)
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 1 - "Treated" = listened to any IVR lesson									
T1: any IVR usage	0.185	-0.227	0.178	-0.075	0.206	-0.043	-0.145	0.178	0.439**
	(0.267)	(0.257)	(0.310)	(0.209)	(0.212)	(0.203)	(0.160)	(0.249)	(0.190)
T2: any IVR usage	0.023	-0.120	0.245	-0.070	0.119	-0.063	0.122	-0.027	0.290**
	(0.149)	(0.151)	(0.190)	(0.141)	(0.135)	(0.152)	(0.106)	(0.152)	(0.131)
P-value:									
T1: any = T2: any	0.506	0.637	0.811	0.976	0.660	0.914	0.0860	0.380	0.391
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017
Panel B: Specification 2 - "Treated" = percent of IVR lessons listened to (with 1=100%)									
T1: percent of IVR used	3.367	-4.105	3.158	-1.354	3.724	-0.758	-2.683	3.248	7.907**
	(4.825)	(4.691)	(5.592)	(3.773)	(3.839)	(3.672)	(2.895)	(4.547)	(3.476)
T2: percent of IVR used	0.337	-1.678	3.394	-0.972	1.664	-0.875	1.666	-0.360	4.053**
	(2.080)	(2.113)	(2.644)	(1.955)	(1.892)	(2.110)	(1.490)	(2.103)	(1.857)
P-value:									
T1: percent = T2: percent	0.487	0.555	0.962	0.908	0.559	0.972	0.112	0.388	0.214
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: Each column is a separate IV regression using the assignment to T1 and T2 as instrument for being in T1 and T2 and having listened to any IVR lesson (in panel A) or the percent of IVR lessons listened to (in panel B). Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: TOT effects specification III & IV (Nepal)

Outcome family:	(1) Learning Outcomes	(2)	(3) Time and Effort	(4)	(5) Non-cognitive Skills Indices	(6)	(7)	(8)	(9)
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 3 - "Treated" = listened to any IVR lesson or talked to tutor at least once									
Any IVR usage [1]	0.186	-0.227	0.178	-0.075	0.207	-0.043	-0.146	0.178	0.440**
	(0.267)	(0.258)	(0.310)	(0.209)	(0.213)	(0.203)	(0.160)	(0.249)	(0.191)
Any phone tutoring [2]	-0.118	0.078	0.048	0.004	-0.064	-0.015	0.193*	-0.149	-0.108
	(0.176)	(0.165)	(0.201)	(0.137)	(0.145)	(0.137)	(0.113)	(0.169)	(0.126)
P-value:									
[1] = [2]	0.476	0.449	0.789	0.806	0.425	0.930	0.189	0.411	0.066
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017
Panel B: Specification 4 - "Treated" = [1] listened to any IVR lesson; [2] listened to IVR lesson and talked to tutor at least once									
Any IVR usage [1]	0.186	-0.228	0.178	-0.075	0.207	-0.043	-0.147	0.179	0.440**
	(0.268)	(0.259)	(0.311)	(0.209)	(0.213)	(0.204)	(0.160)	(0.250)	(0.191)
Used IVR & phone tutoring [2]	-0.189	0.125	0.078	0.007	-0.102	-0.024	0.311*	-0.239	-0.173
	(0.284)	(0.265)	(0.324)	(0.219)	(0.232)	(0.221)	(0.181)	(0.272)	(0.202)
P-value:									
[1] = [2]	0.478	0.479	0.868	0.838	0.463	0.962	0.155	0.399	0.095
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: Each column is a separate IV regression using the assignment to any treatment (T1 or T2) as instrument for "any IVR usage" and using assignment to T2 as instrument for "any phone tutoring" or "used [both] IVR [and] phone tutoring". Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: TOT effects specification V & VI (Nepal)

Outcome family:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Learning Outcomes		Time and Effort		Non-cognitive Skills Indices				
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 5 - "Treated" = hours listening to IVR (capped at 24 hours)									
T1: hours of IVR usage [1]	0.234	-0.284	0.217	-0.093	0.257	-0.052	-0.188	0.226	0.546**
	(0.340)	(0.329)	(0.389)	(0.261)	(0.268)	(0.254)	(0.201)	(0.317)	(0.251)
T2: hours of IVR usage [2]	0.019	-0.088	0.176	-0.051	0.087	-0.045	0.085	-0.018	0.211**
	(0.108)	(0.111)	(0.138)	(0.101)	(0.099)	(0.109)	(0.078)	(0.109)	(0.099)
P-value:									
[1] = [2]	0.486	0.502	0.906	0.853	0.489	0.977	0.147	0.401	0.140
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017
Panel B: Specification 6 - "Treated" = frequency of IVR calls (capped at 144 times)									
T1: frequency of IVR calls [1]	0.007	-0.009	0.007	-0.003	0.008	-0.002	-0.006	0.007	0.017**
	(0.011)	(0.010)	(0.012)	(0.008)	(0.008)	(0.008)	(0.006)	(0.010)	(0.008)
T2: frequency of IVR calls [2]	0.001	-0.003	0.007	-0.002	0.003	-0.002	0.003	-0.001	0.008**
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
P-value:									
[1] = [2]	0.484	0.541	0.996	0.894	0.541	0.985	0.118	0.392	0.188
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: Each column is a separate IV regression using the assignment to T1 and T2 as instrument for being in T1 and T2 and the total hours spent listening to IVR lessons (in panel A) or the frequencing of IVR lesson usage (in panel B). Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 8: TOT effects specification I & II (Pakistan)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Outcome family:	Learning Outcomes		Time and Effort		Non-cognitive Skills Indices				
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 1 - "Treated" = listened to any IVR lesson									
T1: any IVR usage	0.652	0.034	N.A.	3.585***	-0.182	-0.264	0.054	0.598*	0.044
	(0.957)	(0.416)		(0.486)	(0.188)	(0.368)	(0.055)	(0.323)	(0.084)
T2: any IVR usage	1.170**	0.428**	N.A.	0.890***	0.193**	0.132	0.510***	0.215	-0.022
	(0.479)	(0.215)		(0.154)	(0.091)	(0.179)	(0.048)	(0.153)	(0.036)
P-value:									
T1: any = T2: any	0.472	0.242		0.000	0.021	0.156	0.000	0.157	0.358
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094
Panel B: Specification 2 - "Treated" = percent of IVR lessons listened to (with 1=100%)									
T1: percent of IVR used	4.428	0.224	N.A.	24.359***	-1.258	-1.822	0.368	4.129*	0.304
	(6.591)	(2.831)		(4.644)	(1.304)	(2.539)	(0.431)	(2.295)	(0.582)
T2: percent of IVR used	8.146**	2.970**	N.A.	6.351***	1.319**	0.894	3.536***	1.534	-0.151
	(3.395)	(1.514)		(1.167)	(0.656)	(1.270)	(0.463)	(1.074)	(0.251)
P-value:									
T1: percent = T2: percent	0.453	0.229		0.000	0.021	0.155	0.000	0.177	0.358
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094

Notes: Each column is a separate IV regression using the assignment to T1 and T2 as instrument for being in T1 and T2 and having listened to any IVR lesson (in panel A) or the percent of IVR lessons listened to (in panel B). Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: TOT effects specification III & IV (Pakistan)

Outcome family:	(1) Learning Outcomes	(2)	(3) Time and Effort	(4)	(5) Non-cognitive Skills Indices	(6)	(7)	(8)	(9)
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 3 - "Treated" = listened to any IVR lesson or talked to tutor at least once									
Any IVR usage [1]	0.621	0.012	N.A.	3.742***	-0.205	-0.285	0.027	0.619*	0.048
	(0.993)	(0.430)		(0.525)	(0.199)	(0.377)	(0.051)	(0.338)	(0.088)
Any phone tutoring [2]	0.288	0.219	N.A.	-1.503***	0.209**	0.220	0.255***	-0.213	-0.037
	(0.399)	(0.185)		(0.266)	(0.091)	(0.153)	(0.029)	(0.152)	(0.040)
P-value:									
[1] = [2]	0.806	0.729		0.000	0.143	0.331	0.002	0.081	0.502
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094
Panel B: Specification 4 - "Treated" = [1] listened to any IVR lesson; [2] listened to IVR lesson and talked to tutor at least once									
Any IVR usage [1]	0.650	0.033	N.A.	3.594***	-0.183	-0.266	0.052	0.599*	0.044
	(0.959)	(0.417)		(0.488)	(0.188)	(0.369)	(0.056)	(0.324)	(0.084)
Used IVR & phone tutoring [2]	0.533	0.406	N.A.	-2.770***	0.385**	0.408	0.469***	-0.393	-0.068
	(0.741)	(0.347)		(0.461)	(0.166)	(0.288)	(0.063)	(0.278)	(0.074)
P-value:									
[1] = [2]	0.943	0.613		0.000	0.098	0.290	0.000	0.089	0.468
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094

Notes: Each column is a separate IV regression using the assignment to any treatment (T1 or T2) as instrument for "any IVR usage" and using assignment to T2 as instrument for "any phone tutoring" or "used [both] IVR [and] phone tutoring". Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 10: TOT effects specification V & VI (Pakistan)

Outcome family:	(1) Learning Outcomes	(2)	(3) Time and Effort	(4)	(5) Non-cognitive Skills Indices	(6)	(7)	(8)	(9)
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 5 - "Treated" = hours listening to IVR (capped at 24 hours)									
T1: hours of IVR usage [1]	0.177	0.010	N.A.	0.959***	-0.049	-0.071	0.016	0.163*	0.012
	(0.260)	(0.112)		(0.189)	(0.052)	(0.100)	(0.018)	(0.091)	(0.023)
T2: hours of IVR usage [2]	0.333**	0.121*	N.A.	0.261***	0.054**	0.036	0.144***	0.063	-0.006
	(0.140)	(0.062)		(0.049)	(0.027)	(0.052)	(0.020)	(0.044)	(0.010)
P-value:									
[1] = [2]	0.423	0.214		0.000	0.020	0.150	0.000	0.193	0.352
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094
Panel B: Specification 6 - "Treated" = frequency of IVR calls (capped at 144 times)									
T1: frequency of IVR calls [1]	0.024	0.001	N.A.	0.134***	-0.007	-0.010	0.002	0.023*	0.002
	(0.036)	(0.016)		(0.025)	(0.007)	(0.014)	(0.002)	(0.013)	(0.003)
T2: frequency of IVR calls [2]	0.042**	0.015**	N.A.	0.033***	0.007**	0.005	0.018***	0.008	-0.001
	(0.017)	(0.008)		(0.006)	(0.003)	(0.007)	(0.002)	(0.005)	(0.001)
P-value:									
[1] = [2]	0.532	0.269		0.000	0.027	0.171	0.000	0.162	0.372
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094

Notes: Each column is a separate IV regression using the assignment to T1 and T2 as instrument for being in T1 and T2 and the total hours spent listening to IVR lessons (in panel A) or the frequency of IVR lesson usage (in panel B). Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 11: **Spillover effects (Nepal)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Outcome family:	Learning Outcomes		Time and Effort		Non-cognitive Skills Indices				
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 1 - Indicators of being in a T1 or T2 school									
T1 School	-0.000	-0.123	0.026	-0.006	0.041	-0.022	-0.024	0.068	0.049
	(0.086)	(0.085)	(0.105)	(0.066)	(0.064)	(0.073)	(0.048)	(0.082)	(0.074)
T2 School	0.017	-0.081	0.077	-0.107	0.105	-0.014	-0.029	-0.012	0.161**
	(0.093)	(0.091)	(0.098)	(0.089)	(0.083)	(0.083)	(0.057)	(0.085)	(0.074)
P-value: T1=T2	0.869	0.664	0.636	0.285	0.452	0.920	0.933	0.427	0.170
Observations	4,037	4,037	4,037	4,037	4,037	4,037	4,037	4,037	4,037
Panel B: Specification 2 - Proportion of students in school treated in T1 or T2 treatment arm									
Proportion in T1	0.055	-0.206	0.032	0.063	0.074	0.009	0.002	0.214	0.080
	(0.175)	(0.181)	(0.204)	(0.134)	(0.137)	(0.151)	(0.100)	(0.165)	(0.157)
Proportion in T2	0.035	-0.177	0.131	-0.162	0.204	-0.023	-0.059	0.070	0.293**
	(0.171)	(0.178)	(0.194)	(0.176)	(0.165)	(0.172)	(0.113)	(0.173)	(0.146)
P-value: PT1=PT2	0.918	0.878	0.641	0.211	0.445	0.855	0.602	0.467	0.200
Observations	4,037	4,037	4,037	4,037	4,037	4,037	4,037	4,037	4,037

Notes: Each column in each panel is a separate regression estimating the spillover effects of the interventions on untreated students in treatment schools. Sample includes all **untreated** students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 214; schools where all grade-10 students are treated are excluded from the spillover analysis). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 12: **Spillover** effects (Pakistan)

Outcome family:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Learning Outcomes		Time and Effort		Non-cognitive Skills Indices				
Outcome variable:	Math	English	Study Time	Study Effort	Aspiration	Hope	Locus of Control	Growth Mindset	Psychological Well-being
Panel A: Specification 1 - Indicators of being in a T1 or T2 school									
T1 School	-0.192	-0.120	N.A.	0.667***	-0.083*	-0.172*	0.001	0.003	-0.003
	(0.197)	(0.079)		(0.091)	(0.045)	(0.093)	(0.007)	(0.089)	(0.019)
T2 School	0.226	0.033	N.A.	0.631***	-0.043	-0.082	-0.006	0.125*	-0.009
	(0.234)	(0.093)		(0.085)	(0.044)	(0.098)	(0.007)	(0.072)	(0.019)
P-value: T1=T2	0.041	0.089		0.739	0.369	0.391	0.32	0.182	0.741
Observations	2,383	2,383		2,383	1,873	1,873	1,873	1,873	1,873
Panel B: Specification 2 - Proportion of students in school treated in T1 or T2 treatment arm									
Proportion in T1	-0.264	-0.196	N.A.	1.047***	-0.107	-0.223	0.004	-0.002	0.005
	(0.310)	(0.128)		(0.156)	(0.071)	(0.151)	(0.011)	(0.148)	(0.031)
Proportion in T2	0.532	0.096	N.A.	0.948***	-0.026	-0.052	-0.009	0.213*	-0.011
	(0.387)	(0.155)		(0.141)	(0.068)	(0.151)	(0.011)	(0.122)	(0.032)
P-value: PT1=PT2	0.026	0.057		0.602	0.302	0.321	0.301	0.178	0.659
Observations	2,383	2,383		2,383	1,873	1,873	1,873	1,873	1,873

Notes: Each column in each panel is a separate regression estimating the spillover effects of the interventions on untreated students in treatment schools. Sample includes all **untreated** students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 184; schools where all grade-10 students are treated are excluded from the spillover analysis). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Heterogeneous effects by gender (Nepal)

Outcome family:	(1) (2) Learning Outcomes		(3) (4) Time and Effort Indices		(5) (6) (7) (8) (9) Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.061	-0.096	0.112	0.063	0.075	0.018	-0.017	0.033	0.190**
	(0.094)	(0.091)	(0.108)	(0.088)	(0.079)	(0.079)	(0.066)	(0.096)	(0.085)
T2	-0.025	-0.086	0.125	-0.045	0.011	-0.044	0.066	0.039	0.184**
	(0.087)	(0.089)	(0.110)	(0.097)	(0.077)	(0.095)	(0.072)	(0.095)	(0.092)
T1 × Male	0.000	0.047	-0.121	-0.206*	-0.016	-0.075	-0.072	0.063	-0.100
	(0.079)	(0.074)	(0.106)	(0.115)	(0.090)	(0.092)	(0.098)	(0.101)	(0.092)
T2 × Male	0.082	0.051	0.016	0.017	0.119	0.022	0.002	-0.125	-0.067
	(0.080)	(0.077)	(0.100)	(0.116)	(0.091)	(0.104)	(0.097)	(0.100)	(0.098)
Male	0.121**	0.002	-0.006	-0.002	-0.170***	-0.015	0.131**	0.056	0.125**
	(0.054)	(0.050)	(0.061)	(0.072)	(0.062)	(0.055)	(0.066)	(0.063)	(0.062)
P-value:									
T1+T1×Male=0	0.549	0.615	0.950	0.128	0.514	0.505	0.268	0.337	0.199
T2+T2×Male=0	0.536	0.705	0.239	0.763	0.173	0.830	0.396	0.380	0.120
T1=T2	0.366	0.914	0.907	0.258	0.460	0.505	0.287	0.953	0.947
T1×Male=T2×Male	0.970	0.880	0.283	0.267	0.458	0.735	0.0690	0.0950	0.732
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 14: Heterogeneous effects by household income (Nepal)

Outcome family:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Learning Outcomes		Time and Effort Indices		Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.013	-0.070	0.077	-0.105	0.007	-0.038	-0.022	0.077	0.097
	(0.098)	(0.109)	(0.110)	(0.089)	(0.082)	(0.082)	(0.067)	(0.091)	(0.076)
T2	-0.001	-0.049	0.103	-0.029	0.045	-0.045	0.049	0.011	0.120
	(0.088)	(0.092)	(0.099)	(0.093)	(0.083)	(0.099)	(0.075)	(0.099)	(0.084)
T1 × High income	0.092	-0.012	-0.059	0.153	0.119	0.046	-0.048	-0.035	0.094
	(0.084)	(0.085)	(0.105)	(0.093)	(0.087)	(0.085)	(0.080)	(0.086)	(0.081)
T2 ×High income	0.027	-0.025	0.027	-0.016	0.035	0.020	0.036	-0.051	0.065
	(0.075)	(0.082)	(0.103)	(0.087)	(0.084)	(0.100)	(0.086)	(0.094)	(0.082)
High income	-0.008	0.042	-0.071	-0.044	-0.073	-0.066	-0.003	0.031	0.015
	(0.044)	(0.053)	(0.054)	(0.047)	(0.050)	(0.055)	(0.051)	(0.051)	(0.051)
P-value:									
T1+T1×High=0	0.292	0.329	0.889	0.541	0.138	0.917	0.296	0.665	0.0110
T2+T2×High=0	0.775	0.408	0.322	0.594	0.339	0.791	0.220	0.663	0.0170
T1=T2	0.884	0.839	0.819	0.434	0.662	0.943	0.406	0.529	0.765
T1×High=T2×High	0.467	0.935	0.394	0.286	0.641	0.715	0.0300	0.443	0.948
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: "High income" is an indicator equal to one if household's monthly income at baseline was above the sample median (22,000 Nepalese rupees per month). Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 15: Heterogeneous effects by father's education (Nepal)

Outcome family:	(1) (2) Learning Outcomes		(3) (4) Time and Effort Indices		(5) (6) (7) (8) (9) Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.087	-0.123	0.012	-0.031	0.069	0.037	-0.060	0.113	0.115
	(0.099)	(0.096)	(0.104)	(0.081)	(0.082)	(0.077)	(0.061)	(0.089)	(0.076)
T2	-0.007	-0.049	0.121	-0.037	0.064	-0.024	0.053	0.026	0.141*
	(0.081)	(0.085)	(0.106)	(0.086)	(0.077)	(0.092)	(0.070)	(0.088)	(0.081)
T1 × High edu. father	-0.060	0.117	0.082	0.019	0.001	-0.123	0.030	-0.129	0.078
	(0.077)	(0.071)	(0.110)	(0.088)	(0.077)	(0.076)	(0.079)	(0.104)	(0.083)
T2 ×High edu. father	0.045	-0.033	-0.005	0.006	0.001	-0.020	0.034	-0.096	0.035
	(0.070)	(0.060)	(0.094)	(0.084)	(0.077)	(0.090)	(0.085)	(0.097)	(0.081)
High edu. father	0.020	0.071*	0.098*	0.067	0.106**	0.067	0.030	0.068	0.004
	(0.042)	(0.037)	(0.056)	(0.059)	(0.045)	(0.052)	(0.049)	(0.055)	(0.052)
P-value:									
T1+T1×High=0	0.776	0.940	0.510	0.889	0.366	0.275	0.681	0.883	0.0130
T2+T2×High=0	0.693	0.357	0.361	0.732	0.455	0.643	0.239	0.502	0.0300
T1=T2	0.333	0.429	0.307	0.945	0.955	0.489	0.115	0.373	0.757
T1×High=T2×High	0.917	0.431	0.881	0.833	0.956	0.659	0.179	0.649	0.834
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: "High edu. father" is an indicator equal to one if father's education level is above sample median (i.e., attended at least some secondary school or completed 6 or more years of formal schooling). Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 16: Heterogeneous effects by mother's education (Nepal)

Outcome family:	(1) (2) Learning Outcomes		(3) (4) Time and Effort Indices		(5) (6) (7) (8) (9) Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.058	-0.067	-0.008	-0.016	0.107	-0.028	0.017	0.063	0.133
	(0.092)	(0.102)	(0.084)	(0.086)	(0.084)	(0.085)	(0.069)	(0.090)	(0.085)
T2	-0.008	-0.027	-0.111	-0.124	0.069	-0.121	0.111*	0.011	0.200**
	(0.085)	(0.089)	(0.083)	(0.082)	(0.079)	(0.088)	(0.066)	(0.095)	(0.086)
T1 × High edu. mother	0.003	-0.020	-0.017	-0.019	-0.076	0.024	-0.123	-0.011	0.023
	(0.082)	(0.091)	(0.098)	(0.097)	(0.085)	(0.086)	(0.078)	(0.093)	(0.086)
T2 ×High edu. mother	0.037	-0.069	0.154**	0.164**	-0.010	0.165**	-0.084	-0.047	-0.084
	(0.069)	(0.075)	(0.077)	(0.077)	(0.080)	(0.069)	(0.070)	(0.082)	(0.077)
High edu. mother	0.051	0.026	0.002	0.001	0.091*	0.008	0.097*	0.094**	0.090*
	(0.049)	(0.048)	(0.056)	(0.056)	(0.050)	(0.053)	(0.049)	(0.045)	(0.053)
P-value:									
T1+T1×High=0	0.556	0.362	0.766	0.682	0.705	0.957	0.098	0.606	0.027
T2+T2×High=0	0.750	0.285	0.624	0.652	0.493	0.628	0.692	0.682	0.117
T1=T2	0.471	0.693	0.255	0.233	0.672	0.322	0.215	0.623	0.454
T1×High=T2×High	0.766	0.922	0.439	0.403	0.767	0.58	0.078	0.396	0.606
Observations	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017	4,017

Notes: "High edu. mother" is an indicator equal to one if mother's education level is at or above sample median (i.e., attended any formal school). Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 223). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 17: Heterogeneous effects by gender (Pakistan)

Outcome family:	(1) (2) Learning Outcomes		(3) (4) Time and Effort Indices		(5) (6) (7) (8) (9) Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	-0.212	-0.061		0.916***	-0.017	-0.263***	0.019*	0.165**	0.020
	(0.194)	(0.088)		(0.087)	(0.052)	(0.087)	(0.012)	(0.076)	(0.022)
T2	0.278	-0.021		0.608***	0.149***	-0.264***	0.210***	0.083	-0.009
	(0.211)	(0.094)		(0.090)	(0.053)	(0.091)	(0.021)	(0.080)	(0.021)
T1 × Male	0.693**	0.120		-0.487***	-0.038	0.423***	-0.037**	-0.106	-0.034
	(0.343)	(0.146)		(0.135)	(0.064)	(0.150)	(0.015)	(0.110)	(0.028)
T2 × Male	0.443	0.426**		-0.486***	-0.144**	0.665***	0.006	0.027	-0.009
	(0.399)	(0.164)		(0.125)	(0.072)	(0.138)	(0.031)	(0.111)	(0.027)
Male	-0.037	0.007		-0.053	-0.001	0.083	0.010	-0.017	0.047**
	(0.304)	(0.113)		(0.083)	(0.048)	(0.117)	(0.012)	(0.077)	(0.019)
P-value:									
T1+T1×Male=0	0.118	0.646		0.000	0.177	0.173	0.126	0.496	0.477
T2+T2×Male=0	0.040	0.005		0.189	0.914	0.000	0.000	0.217	0.354
T1=T2	0.004	0.630		0.003	0.001	0.988	0.000	0.263	0.169
T1×Male=T2×Male	0.351	0.006		0.001	0.240	0.010	0.000	0.592	0.823
Observations	3,094	3,094		3,094	3,094	3,094	3,094	3,094	3,094

Notes: Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 18: Heterogeneous effects by household income (Pakistan)

Outcome family:	(1) (2) Learning Outcomes		(3) (4) Time and Effort Indices		(5) (6) (7) (8) (9) Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.054	-0.053	N.A.	0.689***	-0.023	-0.047	0.010	0.166**	-0.003
	(0.199)	(0.085)		(0.088)	(0.049)	(0.083)	(0.012)	(0.083)	(0.021)
T2	0.422*	0.115	N.A.	0.413***	0.131**	0.065	0.230***	0.158*	-0.031
	(0.222)	(0.096)		(0.084)	(0.052)	(0.089)	(0.026)	(0.088)	(0.021)
T1 × High income	0.101	0.095		-0.013	-0.024	-0.012	-0.012	-0.094	0.021
	(0.125)	(0.065)		(0.099)	(0.050)	(0.076)	(0.012)	(0.095)	(0.023)
T2 ×High income	0.113	0.112		-0.068	-0.088*	-0.019	-0.028	-0.117	0.037*
	(0.131)	(0.069)		(0.094)	(0.051)	(0.085)	(0.026)	(0.105)	(0.021)
High income	0.000	-0.022		0.025	0.029	-0.045	-0.007	0.051	-0.012
	(0.091)	(0.046)		(0.063)	(0.034)	(0.054)	(0.007)	(0.057)	(0.014)
P-value:									
T1+T1×High=0	0.408	0.621		0.000	0.221	0.441	0.828	0.305	0.354
T2+T2×High=0	0.008	0.016		0.000	0.290	0.579	0.000	0.601	0.711
T1=T2	0.028	0.044		0.002	0.003	0.125	0.000	0.928	0.185
T1×High=T2×High	0.025	0.030		0.000	0.031	0.155	0.000	0.679	0.533
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094

Notes: "High income" is an indicator equal to one if household's monthly income at baseline was above the sample median (28,000 Pakistani rupees per month). Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 19: Heterogeneous effects by father's education (Pakistan)

Outcome family:	(1) (2) Learning Outcomes		(3) (4) Time and Effort Indices		(5) (6) (7) Non-cognitive Skills Indices			(8)	(9)
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.088	-0.012	N.A.	0.710***	-0.038	-0.105	0.015	0.101	0.005
	(0.181)	(0.083)		(0.087)	(0.043)	(0.091)	(0.011)	(0.086)	(0.024)
T2	0.579***	0.204**	N.A.	0.403***	0.126***	0.054	0.198***	0.160*	-0.018
	(0.211)	(0.101)		(0.085)	(0.048)	(0.092)	(0.024)	(0.091)	(0.020)
T1 × High edu. father	0.038	0.022		-0.053	0.001	0.095	-0.022**	0.020	0.007
	(0.120)	(0.080)		(0.090)	(0.047)	(0.083)	(0.011)	(0.100)	(0.025)
T2 ×High edu. father	-0.163	-0.043		-0.053	-0.083	0.001	0.029	-0.126	0.015
	(0.143)	(0.084)		(0.095)	(0.051)	(0.083)	(0.029)	(0.107)	(0.025)
High edu. father	0.180**	0.070		0.051	0.029	-0.037	0.003	0.061	-0.010
	(0.078)	(0.050)		(0.064)	(0.033)	(0.056)	(0.008)	(0.063)	(0.018)
P-value:									
T1+T1×High=0	0.521	0.912		0.000	0.386	0.887	0.506	0.088	0.498
T2+T2×High=0	0.049	0.090		0.000	0.327	0.489	0.000	0.665	0.876
T1=T2	0.007	0.017		0.000	0.000	0.045	0.000	0.517	0.281
T1×High=T2×High	0.083	0.093		0.000	0.080	0.363	0.000	0.246	0.434
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094

Notes: "High edu. father" is an indicator equal to one if father's education level is above sample median (i.e., completed 5 or more years of formal schooling). Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 4) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 20: Heterogeneous effects by mother's education (Pakistan)

Outcome family:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Learning Outcomes		Time and Effort Indices		Non-cognitive Skills Indices				
	Standardized Scores		Study	Study			Locus of	Growth	Psychological
Outcome variable:	Math	English	Time	Effort	Aspiration	Hope	Control	Mindset	Well-being
T1	0.114	-0.006	N.A.	0.687***	-0.057	-0.111	0.002	0.072	0.007
	(0.185)	(0.085)		(0.094)	(0.044)	(0.086)	(0.011)	(0.084)	(0.019)
T2	0.543***	0.217**	N.A.	0.381***	0.073	0.044	0.189***	0.072	-0.013
	(0.203)	(0.098)		(0.086)	(0.048)	(0.092)	(0.025)	(0.092)	(0.020)
T1 × High edu. mother	0.001	0.019		-0.017	0.038	0.117	0.000	0.076	0.003
	(0.141)	(0.077)		(0.102)	(0.049)	(0.086)	(0.011)	(0.091)	(0.023)
T2 ×High edu. mother	-0.103	-0.070		-0.018	0.013	0.021	0.047	0.031	0.008
	(0.165)	(0.089)		(0.101)	(0.049)	(0.087)	(0.029)	(0.107)	(0.023)
High edu. mother	0.132	0.055		0.036	0.013	-0.045	-0.016*	-0.060	0.005
	(0.097)	(0.051)		(0.068)	(0.033)	(0.059)	(0.009)	(0.057)	(0.016)
P-value:									
T1+T1×High=0	0.564	0.881		0.000	0.646	0.939	0.831	0.027	0.620
T2+T2×High=0	0.051	0.141		0.000	0.048	0.427	0.000	0.165	0.743
T1=T2	0.009	0.012		0.001	0.009	0.048	0.000	0.999	0.268
T1×High=T2×High	0.089	0.148		0.001	0.011	0.417	0.000	0.571	0.428
Observations	3,347	3,347		3,347	3,094	3,094	3,094	3,094	3,094

Notes: "High edu. mother" is an indicator equal to one if mother's education level is at or above sample median (i.e., completed 5 or more years of formal schooling). Each column is a separate regression estimating the direct impact of the interventions on treated students using endline data. Sample includes all treated students in treatment schools and all students in control schools. All regressions (except the one in col. 7) control for the baseline value of the outcome variable. All regressions control for baseline individual, household, and school characteristics, as well as a set of district fixed effects. Unit of observation is at the student level. Robust standard errors clustered at the school level are shown in round parentheses (number of clusters is 217). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Appendix A Additional Tables and Figures

Figure A1: Randomization process in Nepal

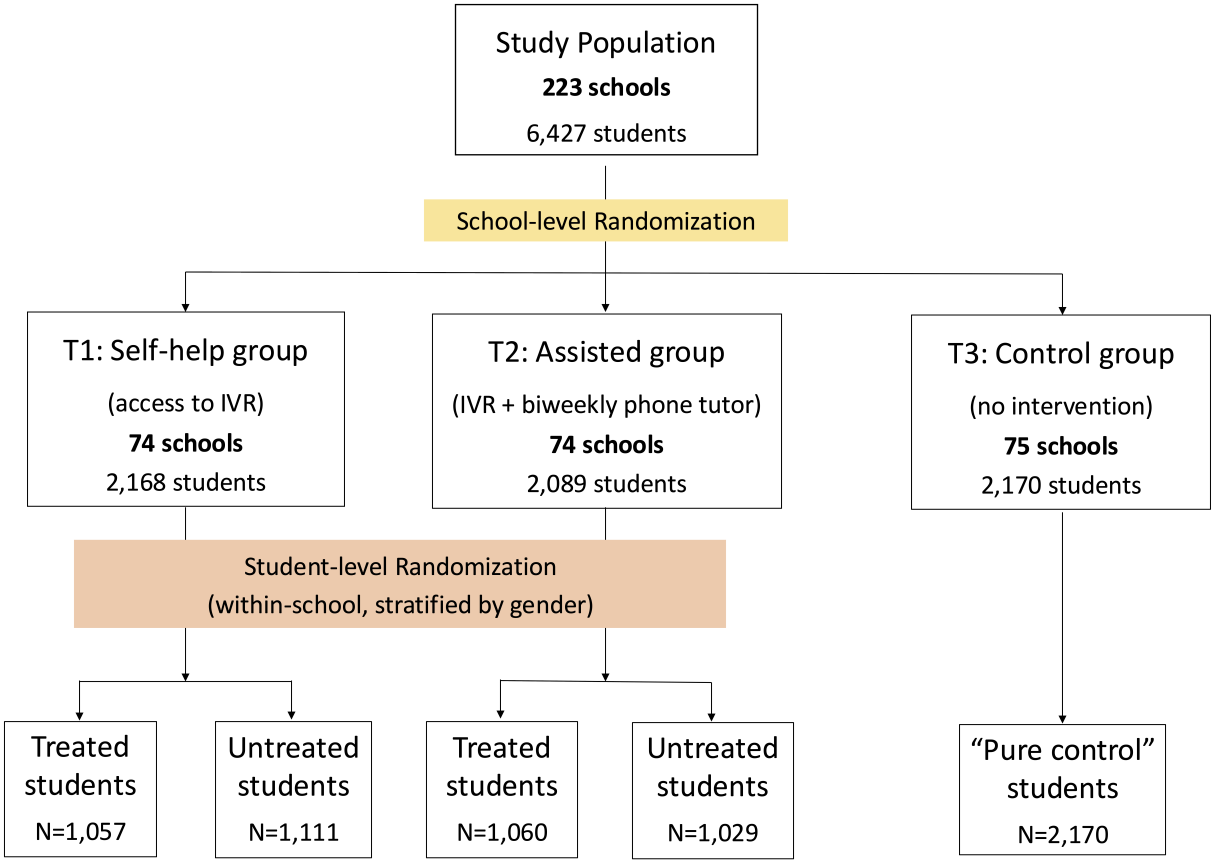


Figure A2: Randomization process in Pakistan

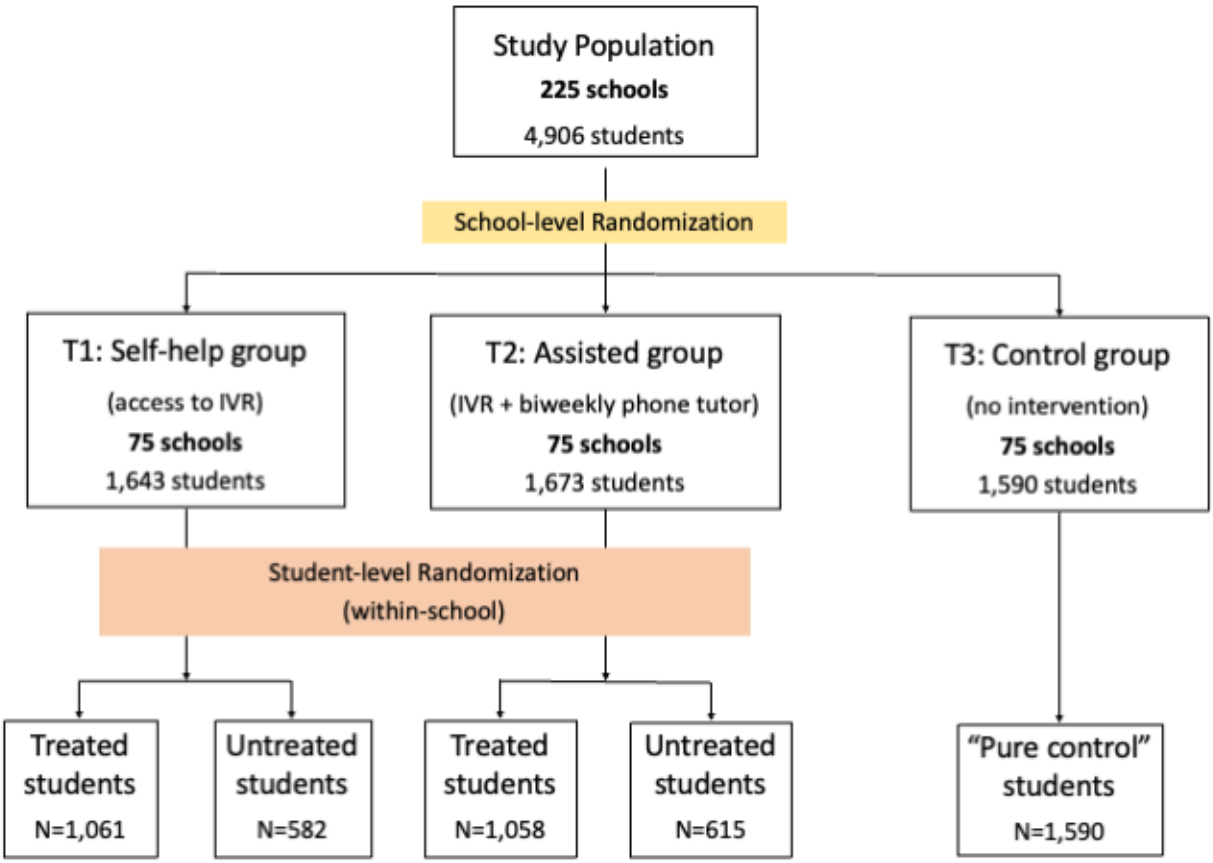


Table A1: Balance of baseline student and household characteristics across treatment and control schools in Nepal

	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(9)
	Full Sample		Mean by Treatment Arm			P-value for Testing Equality		
	Mean	S.D.	Control	T1	T2	C=T1	C=T2	All 3 Equal
Age	14.9	1.22	14.9	15	14.9	0.252	0.734	0.511
Male	0.43	0.496	0.43	0.44	0.43	0.824	0.898	0.975
Father's education (years)	5.6	4.00	5.8	5.5	5.6	0.283	0.320	0.522
Mother's education (years)	3.0	3.79	3.1	2.9	3.1	0.312	0.802	0.576
HH monthly income	26622	21406	26568	26119	27200	0.733	0.685	0.768
HH members	5.5	2.02	5.5	5.5	5.6	0.840	0.186	0.335
HH children under 18	3.2	2.09	3.1	3.3	3.2	0.459	0.858	0.714
HH minority religion	0.05	0.223	0.04	0.04	0.07	0.775	0.038	0.074
HH has phone	0.98	0.133	0.98	0.98	0.99	0.748	0.081	0.182
HH has smart phone	0.87	0.336	0.85	0.88	0.89	0.173	0.062	0.172
HH has TV	0.42	0.493	0.44	0.39	0.41	0.167	0.486	0.383
Math standardized score	0.040	0.942	0.000	0.028	0.095	0.802	0.389	0.663
English standardized score	0.048	0.967	0.000	0.058	0.087	0.611	0.364	0.660
Aspires BA degree or more	0.44	0.496	0.44	0.45	0.42	0.735	0.540	0.648
Has growth mindset	0.40	0.49	0.38	0.39	0.43	0.868	0.456	0.722
Engages private tutor	0.27	0.443	0.23	0.29	0.28	0.334	0.418	0.585
Monthly expenditure on tutor	264	616	238	275	280	0.595	0.563	0.813
Study minutes per day	166	93.7	171	163	166	0.206	0.439	0.439
Help HH chores minutes per day	262	200	252	285	248	0.101	0.768	0.147
Sample Size	6427	.	2170	2168	2089	.	.	.

Notes: Baseline student and household characteristics by treatment status of the school. The *p*-value for testing equality of means for a variable y comes from testing $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, and $\beta_1 = \beta_2 = \beta_3$ in regression $y = \beta_1 \cdot control + \beta_2 \cdot T_1 + \beta_3 \cdot T_2 + \epsilon$ with standard errors clustered at the school level (as the first-stage randomization was done at the school level). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Balance of baseline student and household characteristics across treatment and control schools in Pakistan

	Full Sample		Mean by Treatment Arm			P-value for Testing Equality		
	Mean	SD	C	T1	T2	C=T1	C=T2	All 3 equal
Age	13.6	5.46	12.6	13.9	14.2	0.008	0.001	0.003
Male	0.37	0.482	0.42	0.34	0.34	0.319	0.271	0.483
Baseline Survey Complete	0.87	0.335	0.81	0.89	0.91	0.007	0	0.002
Father's education (years)	3.9	3.6	3.7	3.9	4.1	0.373	0.046	0.133
Mother's education (years)	3.6	3.5	3.5	3.5	3.8	0.777	0.317	0.391
HH monthly income	30726	28409	29226	31101	31784	0.252	0.11	0.268
HH members	6.5	3.46	5.9	6.6	6.9	0.004	0	0
HH children under 18	3.4	1.89	3.2	3.4	3.5	0.034	0.003	0.012
HH minority religion	0.01	0.119	0.01	0.02	0.01	0.073	0.152	0.127
HH has phone	0.95	0.217	0.95	0.95	0.95	0.593	0.753	0.862
HH has smart phone	0.87	0.339	0.86	0.86	0.88	0.917	0.496	0.706
HH has TV	0.66	0.474	0.68	0.65	0.65	0.173	0.188	0.288
Math standardized score	-0.01	1.02	0	-0.067	0.036	0.471	0.679	0.517
English standardized score	-0.023	1.01	0	-0.063	-0.004	0.568	0.968	0.813
Aspires BA degree or more	0.67	0.47	0.61	0.69	0.71	0.006	0	0.001
Has growth mindset	0.39	0.487	0.35	0.41	0.41	0.14	0.098	0.194
Engages private tutor	0.45	0.497	0.44	0.44	0.46	0.965	0.555	0.775
Monthly expenditure on tutor	735	924	763	709	735	0.454	0.693	0.754
Study min per day	86	67.6	85	87	84	0.7	0.84	0.843
Help HH chores min per day	140	97	142	137	143	0.526	0.901	0.722
Sample Size	4906	.	1590	1643	1673	.	.	.

Notes: Baseline student and household characteristics by treatment status of the school. The p-value for testing equality of means for a variable y comes from testing $\beta_1 = \beta_2$, $\beta_1 = \beta_3$, and $\beta_1 = \beta_2 = \beta_3$ in regression $y = \beta_1 \cdot control + \beta_2 \cdot T_1 + \beta_3 \cdot T_2 + \epsilon$ with standard errors clustered at the school level (as the first-stage randomization was done at the school level). Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Utilization among treated students

	Nepal			Pakistan		
	All	T1	T2	All	T1	T2
Listened to any IVR lesson	41.4%	32.2%	50.6%	28.3%	17.1%	39.5%
% of math lessons listened	2.7%	1.8%	3.6%	4.0%	2.6%	5.5%
% of English lessons listened	2.7%	1.7%	3.6%	4.3%	2.6%	6.0%
Minutes on math lessons	15.2	8	22.5	35.7	24.0	47.5
Minutes on English lessons	13.5	6.6	20.4	27.0	17.0	37.1
Answered any tutor call			72.7%			77.5%
% of 12 math call answered			11.5%			17.0%
% of 12 English call answered			12.6%			19.7%
Minutes with math tutor			14.5			10.1
Minutes with English tutor			21.7			11.5

Appendix B Why Did IVR Not Work in Nepal

Despite the promise of low-tech learning interventions, the uptake of IVR in Nepal remained disappointingly low, with only about 7–10% of treated students accessing meaningful portions of the lessons. This stands in contrast to the somewhat higher participation in Pakistan, where around 25–30% of students accessed meaningful content. Our findings are based on detailed feedback from our implementing partner, Edukhabar, our Nepalese co-author Dr. Uttam Sharma, and local collaborators. These insights reflect on-the-ground experiences from running the program in rural, mountainous regions where post-COVID recovery has been uneven.

We partnered with Edukhabar because of their sectoral expertise and local policy engagement, while recognizing that their limited rural visibility could present challenges for large-scale community mobilization. Edukhabar is a local NGO with a strong presence in Nepalâs education policy and edtech space. They have deep grassroots connections in certain urban and peri-urban areas, prior experience in education technology initiatives, and a small but committed team. However, they are relatively small in scale and less familiar to communities in remote rural areas, especially in hill districts.

Factors Limiting IVR Uptake

1. **Limited Institutional Presence and Community Recognition:** In many intervention areas, Edukhabar lacked prior visibility or trust among communities. As noted by Dr. Uttam Sharma and Edukhabar managers, this unfamiliarity reduced community buy-in and student engagement, as households did not associate the program with a credible or established local provider. This contrasts with our implementing partner in Pakistan, IED, a premier institution with strong recognition in the education sector.

2. **Challenging Geographical and Technological Terrain:** Nepalâs mountainous terrain in the hill districts severely disrupted mobile network access. Edukhabar field teams reported erratic coverage, making it difficult for students to initiate or maintain IVR calls at conve-

nient times. Network disruptions are a known barrier in rural Nepal, where topography and infrastructure gaps hinder connectivity. Additionally, electricity shortages in remote households made phone charging unreliable, further impeding usage.

3. **Limited Household Access to Reliable Phones:** Many target households had only a single mobile phone, often controlled by adults who were absent during school hours. This limited studentsâ access to the IVR system, especially in households with multiple school-aged children. Some households still relied on basic feature phones with poor speaker quality, which further reduced the learning experience.

4. **Socio-Economic and Post-COVID Constraints:** Participating students often came from disadvantaged households still recovering from COVID-19's economic fallout. Dr. Sharma emphasized that families in these poor socioeconomic conditions showed low capacity to support children's learning, compounded by the pandemic's lingering effects on motivation and routines. We note that parental education levels were low, with the average years of formal schooling at only 3 years for mothers. Post-COVID school closures in Nepal exacerbated inequities, with rural areas reporting persistent disengagement and dropout risks due to economic vulnerability.

5. **Low Baseline Learning Levels and Perceived Curriculum Mismatch:** Public school students in the intervention had notably low academic proficiency, particularly in math and English, making self-directed IVR content less appealing or accessible. As observed by local partners, the low education levelsâworsened by COVID disruptionsâreduced students' motivation to engage with remedial material independently. Additionally, students expressed dissatisfaction with the IVR's structure, preferring content strictly aligned with the Grade 10 SEE curriculum over foundational skills. This mismatch, combined with low digital literacy in rural Nepal, hindered sustained use.

6. **Weak Belief in IVRâs Educational Value:** There was skepticism about IVR as a legitimate learning tool, rooted in Nepal's preference for traditional, in-person rote learning. Edukhabar

teams noted that audio-based distance education was seen as inferior in a context where classroom interaction dominates pedagogy. This perception gap, exacerbated by low digital literacy and unfamiliarity with EdTech, reduced motivation even among those with access.

7. **Limited School Involvement in Student Mobilization:** Due to the study design, school personnel were not actively involved in mobilizing students or monitoring participation. This compares with Pakistan, where IED's status as a premier educational institution may have given it more leverage with local school authorities. As Dr. Sharma and local collaborators pointed out, this lack of institutional anchoring reduced the program's perceived importance among students and parents.

Why T2 Did Not Work Well in Nepal

The T2 (self-help plus tutor support) program in Nepal wasn't as effective as hoped due to a number of interlinked reasons. While the T2 model offered a more intensive intervention, it faced significant challenges in execution, ultimately limiting its impact on learning outcomes.

- **Tutor Recruitment Challenges:** The shortage of qualified local tutors meant long delays in onboarding and inconsistencies in session quality. This was a key capacity constraint for Edukhabar, which is not an academic institution.
- **Limited Academic Oversight:** Edukhabarâs non-academic institutional structure limited their ability to provide continuous pedagogical support and quality assurance for the T2 program.
- **Logistical Mismatches:** Poor network connectivity in rural areas affected not just IVR but also tutorâstudent communication.
- **Competition for Attention and Time:** Motivating students and parents in these disadvantaged communities was a significant challenge. Anecdotal evidence suggests many parents were simply thrilled their children had completed past the eighth grade, and they didn't seem

interested in additional study through phone-based tutoring. This applied to both T1 and T2 and contributed to low overall engagement.

Lessons for Future Design

The Nepal experience offers a sobering yet insightful reminder that the success of low-tech education interventions depends heavily on local context. Issues of infrastructure, institutional trust, educational norms, and socioeconomic realities must not be overlooked. Future efforts may require more intensive community engagement, stronger alignment with curricular expectations, and greater technical and institutional scaffolding to succeed in similarly fragile settings.

For future design, it is essential to:

- **Strengthen Local Presence Before Rollout:** Allocate time and resources for extensive community sensitization in rural areas, especially when working with smaller organizations.
- **Plan for Infrastructure Gaps:** Incorporate offline or hybrid options in geographies with unstable network coverage.
- **Ensure Tutor Pipeline:** Partner with academic institutions or teacher training centers to secure a steady pool of qualified tutors for T2 models.
- **Stagger Rollout:** Begin with pilot clusters to establish credibility and refine operations before scaling.
- **Continuous Monitoring:** Set up real-time usage tracking to quickly detect and address barriers to uptake.

Appendix C Why Was IVR Take-Up Low in Pakistan

While IVR uptake in Pakistan was modest overall (approximately 25–30% accessing meaningful content, higher than Nepal's 7–10%), it remained insufficient to drive strong standalone effectiveness. By contrast, the T2 arm (IVR + tutor support) produced substantial learning gains (e.g., $+0.487$ SD in math, $+0.18$ SD in English) and improvements in non-cognitive skills (e.g., aspiration, locus of control). Our understanding draws on feedback from our implementing partner, the Institute for Educational Development (IED) at Aga Khan University (AKU), local collaborators, and qualitative observations.

IED is a premier education research institution in Pakistan with strong pedagogical expertise and credibility. Its established track record in teacher training and curriculum development made it a strong partner. Importantly, its graduates and networks provided a ready pool of skilled tutors for the T2 arm, which was central to its success. However, IED had limited prior experience with large-scale EdTech or IVR deployments. While this limited their readiness for the technical demands of IVR, their depth in educational practice made them a strong candidate for the hybrid T2 approach where human support was central.

Factors Limiting IVR Uptake

1. **Limited Institutional Experience with Large-Scale Tech Interventions:** IED's strong pedagogical expertise did not fully translate into seamless IVR delivery. The team faced gaps in call scheduling, troubleshooting, and mobilization strategies, particularly when scaling across diverse districts. This was exacerbated by the tech company's inexperience in delivering education via IVR, as this was its first such engagement.

2. **Technological and Infrastructure Constraints:** IVR delivery was hampered by weak mobile networks and frequent electricity shortages in rural and flood-affected areas, similar to the challenges faced in Nepal. Call drops, signal fluctuations, and limited charging capacity made regular engagement difficult. Many students also shared phones with family members,

limiting access during lesson windows.

3. **Low Baseline Proficiency and Curriculum Misalignment:** Students in public schools demonstrated low math and English skills. The IVR content, which focused on foundational learning from grades 7 and 8, felt misaligned with their immediate goal of preparing for the high-stakes board exams. This mismatch diminished the IVR's perceived utility, echoing global findings that low take-up is common when EdTech tools are not tailored to local curriculum pressures.

4. **Socioeconomic and Post-COVID Recovery Barriers:** Many households were recovering from pandemic-related economic stress and prioritized livelihood needs over supplementary education. Post-pandemic learning fatigue also dampened enthusiasm for IVR-based learning. For girls, entrenched gender norms restricted phone access and learning opportunities due to domestic obligations and digital gatekeeping by male family members.

5. **Cultural Skepticism Toward Audio-Based Learning:** In a traditional pedagogical system, face-to-face instruction is dominant, and students and parents often viewed IVR as a secondary, optional resource.

6. **Limited School Integration:** Without formal integration into school timetables or strong teacher endorsement, IVR participation depended entirely on individual or family initiative.

7. **Gender-Specific Barriers and Tutor Interaction:** Girls' lower IVR uptake stemmed from patriarchal norms restricting mobile access, along with limited autonomy and time due to domestic responsibilities. In the T2 arm, male students benefited more, potentially because tutorsâdue to cultural normsâengaged more readily with boys, even though nearly 85% of tutors hired were female. This aligns with feedback from IED and consistent findings on gendered gaps in educational access and attention in Pakistan.

Why T2 Worked Well in Pakistan (Even with Low IVR Take-Up)

Despite modest IVR engagement, the T2 intervention (combining IVR with tutor support) generated large positive impacts on both cognitive and non-cognitive outcomes. Key drivers included:

1. **High-Quality Tutor Recruitment:** IED's ability to recruit skilled tutorsâoften their own students or pedagogy-trained graduatesâensured high pedagogical quality, personalized guidance, and motivational support. This human element helped bridge gaps in foundational knowledge and IVR comprehension, which was a critical factor for students with low baseline proficiency.

2. **Enhanced Non-Cognitive Skills:** Tutors directly fostered aspiration and a higher locus of control through regular encouragement, feedback, and goal-setting. These non-cognitive boosts were particularly pronounced among boys, consistent with qualitative feedback and outcome data.

3. **Synergy Between IVR and Tutoring:** While IVR alone saw limited use, its pairing with regular tutoring created a reinforcing dynamic. Students in T2 accessed more IVR content (listening to twice as many lessons as T1) and benefited from tutors contextualizing and aligning IVR content with exam preparation. This hybrid approach improved both engagement and learning outcomes.

4. **Stronger Program Fidelity and Monitoring in T2:** Unlike the IVR-only arm, T2 benefited from tighter implementation oversight and higher accountability, supported by IED's traditional education management strengths. Tutors often assisted with IVR troubleshooting, increasing both access and effectiveness.

Lessons for Future Design

Pakistanâs experience highlights that EdTech interventions like IVR are highly sensitive to institutional readiness, infrastructural constraints, and cultural norms. Despite low take-up, T2âs success

affirms that well-supported hybrid modelsâwhere technology is complemented by high-quality human interactionâcan produce meaningful gains. This model offers a viable path for future remote learning programs in similarly constrained settings.

For future design, it is essential to:

- **Match Delivery Modality to Institutional Strengths:** Partnering implementing partners with a strong background in educational pedagogy with tech-strong implementers could improve IVR uptake and effectiveness.
- **Integrate with Schools and Curriculum:** Aligning IVR more closely with board exam content and embedding it in school routines could increase perceived relevance.
- **Anticipate Gender Barriers:** Proactive measures, such as family engagement campaigns, are needed to ensure girls can access digital tools.
- **Leverage Hybrid Models:** Pakistanâs T2 success shows that combining EdTech with strong human support can deliver substantial gains even in low-tech, resource-constrained environments.