About the Guide

This data user guide briefly describes the sampling process, children's learning assessment tool, survey elements and timelines, variable information in data, and instructions for merging different data files.

About the Project

Our project explores the potential of community-based accountability relationships to raise children's foundational learning outcomes, with a focus on the most disadvantaged primary-school learners: namely those who are from poorer households and, within these, girls. We ask both whether and how changes occur when school actors are supported to view their accountability as being primarily to their local community and their goal as being to raise all children's learning. To do this, we evaluate a grassroots intervention that supports school actors to work directly with their communities to develop a shared understanding of children's learning levels, collaborate in planning how to raise them, and facilitate action both inside and outside the classroom. In addition to analysing changes in children's learning, we examine the intervention's capacity to create changes in school-community relations, teachers' attitudes perceptions, and actions in the classroom.

The intervention builds on two of *Pratham's* grassroots literacy campaigns to incorporate school-community accountability relationships. *Lakhon Mein Ek* (LME) was primarily an information campaign, supporting communities to undertake and discuss their own independent assessments of children's learning outcomes. *Reading Week* (RW) provided community volunteers with packages of simple materials with which to work with children to improve their foundational literacy and numeracy skills. The proposed intervention seeks to enhance the impact of these previous interventions by working directly with school actors to help them engage and work with their local communities.

Sampling Process

This randomized control trial study was conducted in 400 villages in *Sitapur* district in the northern state of *Uttar Pradesh, India*. The sample villages were randomly divided into three groups - 100 villages for the community-level intervention (i.e., first treatment *PAHAL[1]*), 200

---

[1] *PAHAL* stands for '*Pratham's Action in Home And Learning'*, with the word also meaning initiative in Hindi.

villages for the community-school level intervention (i.e., second treatment *PAHAL+*), and 100 villages were assigned to a control group. The sampled villages were selected such that each sampled village had at least two government primary schools. After this, we randomly sampled 30 children per school, 10 each from grades 2, 3, and 4.

In each sample school, 3 sample grades were included in the study – Std. 2, 3, and 4.

- First, using grade wise attendance register for the current survey month, we randomly selected 20 children from each sample class (grades 2, 3, and 4), for a total of 60 children from each sample school, with the exception of schools with fewer than 20 children enrolled in one or more classes.

For any given class, there were 2 sampling scenarios:

Case 1: 20 or fewer children enrolled in the sample class

Note down the names of all the children currently enrolled with the help of the current month's attendance register.

Case 2: More than 20 children enrolled in the sample class

Pick any number randomly between 1 and the total number of children enrolled in the class. This chosen serial number was the first sampled child in the class attendance register. After that, every fifth child was selected. This process was repeated until we selected a total of 20 children. Choose a new random starting number for each class.

- After selecting 20 children from the class, we tested all those who were present in school that day using ASER language and arithmetic tool, and recorded their highest level on the sampling sheet.
- After testing all children who were present, we identified children in the list who were absent. Starting from the top of the list, we visited each child's household and administer the ASER test at home. This process ended once we had 10 children from that class who were below story level in reading, starting the count from the top of the list of 20 randomly sampled children.
- These ten children were the final list of 10 sampled children in that class.

The tool used for sample identification is provided in a separate file named **Sample_Identification_Tool.pdf** in the **Quesntionnaire_Tools_Codebook.rar** folder.

Measuring Children's Foundational Learning

A children's foundational literacy and numeracy test - the *ASER* test developed by the *Pratham Education Foundation* - was administered to measure foundational literacy and numeracy levels. The children's *ASER* test was conducted in all in-person rounds of the survey, and the tool for measuring it is provided in a separate file named **ASER-Test_Hindi.pdf** in the **Quesntionnaire_Tools_Codebook.rar** folder.

For more information, please refer the *ASER Centre* **Tools & Testing** section through the following link: http://www.asercentre.org/p/141.html

Survey Elements and Timelines

The baseline survey was conducted between October 2018 and January 2019. In addition to children's foundational literacy and numeracy tests, information was collected on each selected student's household, and school-related information. In particular, the survey was done for the following units at the baseline: Child, Household, School, Head Teacher, Teacher[2], and Classroom observations[3]. Table 1 provides information on the data collection round (and timeline) and the respective tools used at that round for collecting information across different observation units. The data files in STATA format (by separate observation units and sections) for all rounds and their respective questionnaire/tool are provided in separate zip folders with their specific names. For instance, the **1.Baseline_Data.rar** folder contains baseline data (i.e., first survey round), collected during October 2018 - January 2019, with separate data folders for each observation unit, and their separate questionnaire/tool for each observation unit is provided in the **Baseline_Tools** folder. Similarly, the **2.Midline_Data.rar** folder contains midline data (i.e., second survey round), collected during November 2019 - March 2020, with separate data folders for each observation unit, and their separate questionnaire/tool for each observation unit is provided in the **Midline_Tools** folder. We have also tracked the sampled children households and teachers through a phone survey during the Covid-19 pandemic to

---

[2]An essential feature of this research was to understand teachers' perceptions and actions towards children, their families, the community and their own work. To do this the study design involved observing and administering a survey questionnaire to Hindi language teachers in Grades 2, 3, and 4.

[3] For more information about classroom observation, please see chapter 4 titled **'Inside the classroom'** of the *ASER Centre* report on **Inside Primary Schools.** See the following link:
http://img.asercentre.org/docs/Publications/Inside_Primary_School/Report/tl_study_print_ready_version_oct_7_2011.pdf

capture the status of education-related activities. The final survey only collected information on the sample children's enrolment status and their learning levels after the pandemic.

Table 1. Survey round, timeline, and tools used

Survey Round	Timeline	Tools Used
Baseline	Oct. 2018 – Jan. 2019	1. **A.** Child Assessment of Foundational Literacy and Numeracy Level **B.** Child Enrolment and Attendance 2. Household Survey 3. School Survey 4. Head Teacher Survey 5. Teacher Survey 6. Classroom Observations
Midline	Dec. 2019 – March 2020	1. **A.** Child Assessment of Foundational Literacy and Numeracy Level **B.** Child Enrolment and Attendance 2. Household Survey 3. Head Teacher Survey 4. Teacher Survey 5. Classroom Observations
COVID-19 Tracking	Oct. – Nov. 2021	1. Household Phone Survey 2. Teacher Phone Survey
Endline	Dec. 2021 – Jan. 2022	1. **A.** Child Assessment of Foundational Literacy and Numeracy Level **B.** Child Enrolment Status

Variables Names

The data was collected only in one district as mentioned earlier. Therefore, all the data files provide the state and district names. Also, a separate individual village (named *vid*) and school ID (named *schoolid*) is available in all the data files. The randomisation was done at the village level. Therefore, we have provided a treatment status of the village through a variable named *group_name.* This variable is divided into 3 categories based on the treatment status, i.e., *Control* for the control group, *PAHAL* for treatment 1, and *PAHAL+* for treatment 2. In treatment 1, the intervention was done only at the community level, while in treatment 2, the intervention was done both at the community and school levels. We have created a separate file, named **Data_Codebook.xlsx** file, that provides information on the variable's description and name used in all the data files.

Merging Data Files

We have constructed a separate identifier that is common across all the files, i.e., School ID. The name of this variable in all the STATA files is *schoolid.* Similarly, we have provided separate ID's for all the individual units, such as child, household, head teacher, and teacher. We have also provided the child ID, named *childid,* in all the files to link/merge other files at the individual child level. Therefore, we can merge these files using the common identifier *schoolid* (and *childid* for merging at the child level) and the separate ID's of the individual observation unit.