Baseline Report

Play.Connect.Learn

Sesame Workshop India Trust (SWI)

Prepared by:
School-to-School International (STS) and
SWI
For All Children Reading: A Grand Challenge for Development

January 2017

Table of Contents

I. Executive Summary

Recognizing that literacy is fundamental to learning, skill acquisition, and success in primary school and beyond, education stakeholders are increasing their focus on the assessment of early grade reading skills. The Early Grade Reading Assessment (EGRA) is an oral student assessment designed to measure the most basic foundational skills for literacy acquisition in the early grades: recognizing letters of the alphabet, reading simple words, understanding sentences and paragraphs, and listening with comprehension.[1] The EGRA methodology was developed under EdData II and has been applied in more than 30 countries and 60 languages.[2]

All Children Reading: A Grand Challenge for Development (ACR GCD), a partnership between the United States Agency for International Development (USAID), World Vision, and the Australian Government, has adopted the standard EGRA to systematically assess reading skills across all Round 2 grantees. The instrument is adapted according to each grantee's project context.

Sesame Workshop India Trust (SWI), an ACR GCD Round 2 grantee, conducted an EGRA baseline assessment on students in the Indian state of Maharashtra. The EGRA was implemented in collaboration with Disha Research Group and School-to-School International (STS).

The results of the baseline data collection, conclusions, and recommendations are presented in this report. Below is a summary of the key findings as presented in Table 1.

Key Findings

1. **The proportion of students receiving zero scores was highest on the reading comprehension subtask and lowest on the letter name identification subtask.** This finding is consistent with the expected pattern of early reading grade skills acquisition by students. Letter name identification is generally seen as a precursor for reading tasks, and reading comprehension requires letter and word recognition, as well as comprehension.
2. **Students are unable to read words and passages with fluency.** Students read, on average, about five familiar words per minute and six words in the oral reading fluency passage. This is not sufficient to be able to read with comprehension.

[1] RTI International and International Rescue Committee. (2011). Guidance Notes for Planning and Implementing Early Grade Reading Assessments.
[2] USAID EdData II. Available at: https://www.eddataglobal.org/reading/

3. **Girls performed better than boys and had lower proportions of zero scores on all subtasks except for listening comprehension.** On listening comprehension, girls had lower scores and higher proportions of zero scores.

Table 1: Mean Results for EGRA Subtasks by Group

Subtask	Intervention (N=402)		Comparison (N=400)		All students (N=802)	
	Mean	Zero Scores (n)	Mean	Zero Scores (n)	Mean	Zero Scores (n)
Letter name identification (CLNPM)	14.3	191	16.2	172	15.2	363
Syllable identification (CSSPM)	6.8	258	7.0	240	6.9	498
Familiar word reading (CFWPM)	4.5	278	6.0	254	5.3	532
Oral reading fluency (ORF) (CWPM)	5.9	257	6.7	239	6.3	496
Reading comprehension (Number of questions correct out of five)	0.2	349	0.3	346	0.3	695
Listening comprehension (Number of questions correct out of four)	0.8	223	1.0	183	0.9	406

II. Project Description

SWI, the organization behind the locally-produced version of *Sesame Street* called *Galli Galli Sim Sim* (GGSS), develops and distributes high-quality and engaging content through television, radio, community radio, print, and digital outreach, to reach children up to eight years old. With support from ACR GCD, SWI is implementing the Play.Connect.Learn program to improve children's early grade reading skills in their mother tongue, Marathi. The program reaches Grade 1 and 2 students through a smartphone application (app) called Play.Connect.Learn. The app, a self-paced, story-based game that uses GGSS characters, is distributed to students with the support of community-based self-help groups (SHGs)[3] and is intended to serve as a supplementary learning tool for students who attend government schools. New content is distributed to the students in three packages, and each package is used for a period of two and a half to three months. Each package contains four story books, and each story book is followed by a game that asks comprehension questions related to the book. Books are unlocked in turn after students complete the game for the book they have finished reading. They must score at least 70 percent to unlock the next book; otherwise, they must re-read the book and try again. In addition to the app, students receive supplementary digital storybooks

[3] Students' families are either already a part of, or are willing to join, the SHGs.

in PDF. If needed, PDF readers are installed on phones to ensure that students have access to the supplementary books.

There have been significant changes in the original program design due to various challenges. The key challenge faced by the program was the lack of smartphone penetration in the original target areas of Osmanabad and Beed. Initially, a prerequisite for participation in the program was family ownership of, or access to a smartphone. To determine families who had possession of a smartphone, SWI conducted a needs assessment, a baseline assessment, and a recruitment process. In preliminary research, SWI estimated 80 percent smartphone penetration in the target areas. Results from the assessments and recruitment processes, however, indicated that only eight percent of students in the target grade range had access to a smartphone at home. As a result of this finding, the original baseline was stopped, and SWI refined its program and research design to respond to the lessons learned about target populations.

In order to guide recruitment for Play.Connect.Learn after the program redesign, SWI collaborated with Mahila Aartik Vikas Mahamandal (MAVIM), a local non-governmental organization (NGO) that provided guidance on which districts in Maharashtra to target for implementation. Based on MAVIM's district recommendations, SWI contacted and shortlisted prospective implementing partner NGOs in the new target districts whose role would be to facilitate SHGs to distribute the Play.Connect.Learn app to families. After discussions with prospective partners about the program and the program scale, including potential intervention locations and number of participants, SWI requested that they share registration, legal documents, and corporate capabilities documents including details on previous work and previous donors. Using this information, SWI selected four implementing partners and determined their intervention districts and participant targets. Implementing partner details are provided in Table 2.

Table 2: SWI Implementing Partners

Name of Organization	Target District	Estimated Participant Target*
Vikas Sahyog Pratishthan	Amravati, Buldhana	2,750
Sangli Mission Society	Sangli, Kohlapur	3,500
Yuva Rural Association	Amravati, Buldhana	3,000
Gramin Samassya Mukti Trust	Yavatmal, Chandrapur	2750
Estimated Participant Total		12,000
Actual Participant Total		13,291

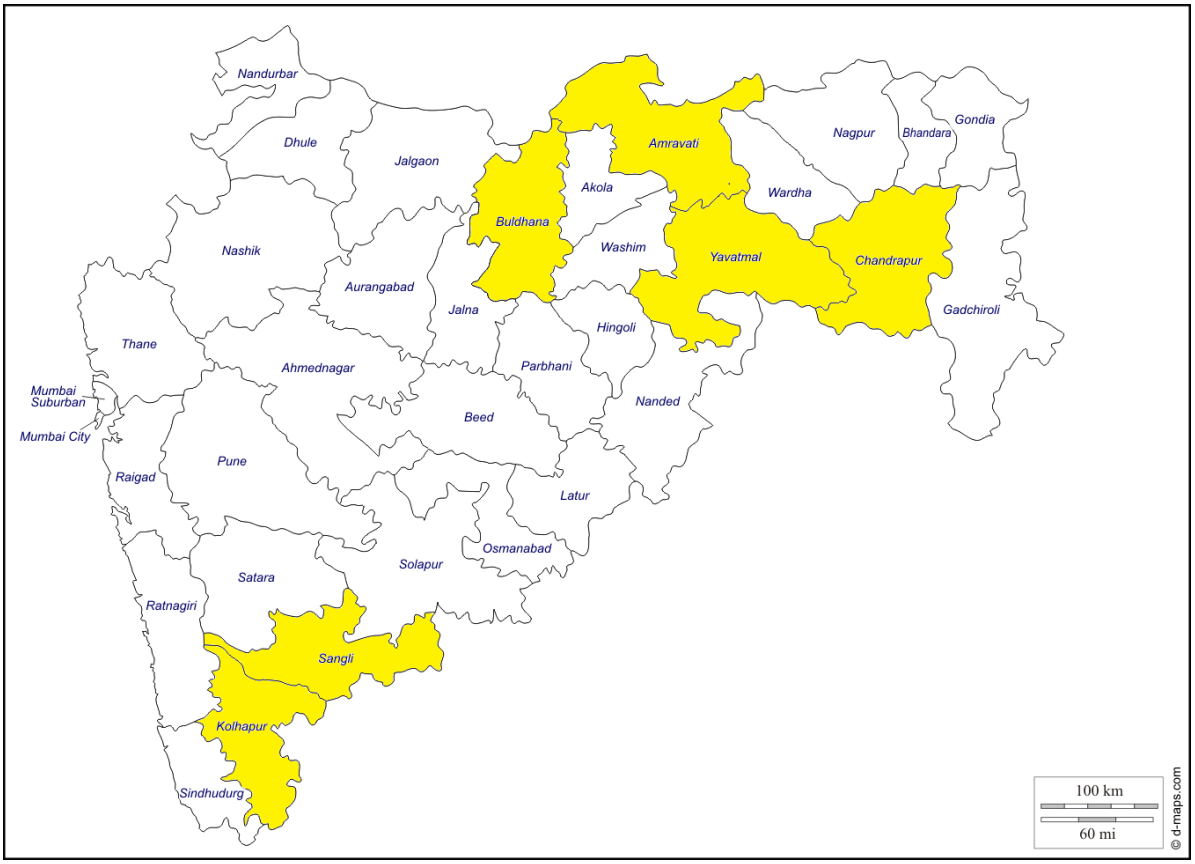
* The target is a subset of the population of participants that the partners reach.

In addition to the changes in target districts and total participants, changes were made to the research design. Because of the low smartphone penetration in target districts, STS and SWI determined that a smaller subset of students should be included in the research

study to be assessed with the EGRA. The smaller research sample made it possible for the program to provide phones to families who did not already have a smartphone. By removing the smartphone ownership prerequisite for incorporation into the research study, the program was able to have a sample that is representative of all families living in Maharashtra, rather than just those who already own smartphones.

Ultimately, the Play.Connect.Learn program's participant target was slightly reduced. As a result, the program will be administered to 12,000 Grade 1 and 2 students over a period of nine months in six target districts in Maharashtra: Sangli, Buldhana, Amravati, Yavatmal, Chandrapur, and Kohlapur (see Figure 1).

Figure 1: Target Districts in Maharashtra State



III. Purpose

The Play.Connect.Learn program seeks to improve children's early grade reading skills in their mother tongue, Marathi. To understand if the program reaches this goal, a research study will be conducted on program participants to answer one key research question specific to the Play.Connect.Learn program:

1. Does exposure to innovative, engaging, digital GGSS content delivered through a smartphone app result in higher improvements in Marathi early grade reading skills for Grade 1 and 2 students than their peers who do not receive exposure?

IV. Evaluation Design and Methodology

To measure the research question of Play.Connect.Learn, an EGRA will be conducted in two phases: a baseline assessment and an endline assessment. The research sample includes an intervention group of 402 Grade 1 and 2 students who will participate in the Play.Connect.Learn program, and a comparison group of 400 students who will not receive any benefits from the program. Baseline and endline assessment results will be compared across the two intervention groups to determine if there were early reading skills gains attributable to the program.

Instrument Development

The EGRA used in the baseline assessment was adapted to Marathi. The initial Marathi EGRA adaptation workshop and assessor training occurred on October 6–10, 2015, and was held with ACR GCD grantee Beneficent Technologies (Benetech), which is also implementing a program in Maharashtra. Members of the SWI and Benetech teams, STS, a Marathi language expert, and two instructional design experts from the Maharashtra State Council for Educational Research and Training (MSCERT) participated in the adaptation workshop (see Annex A for full workshop schedule).

As a result of the pre-test and an initial pilot test, the following changes were made to the EGRA subtasks:

- The syllable identification subtask was added to allow for differentiation between the simple names of the letters and the many letter combinations found in Marathi. In Marathi, there are 49 letter names—36 consonants and 13 vowels. Each vowel and consonant combination creates a different syllable sound that is written in a unique way to indicate that the letters have been combined. In total, there are 408 letter combinations.
- Items on both the letter name identification and syllable identification subtasks were arranged randomly on the grid. Initially, for the syllable identification subtask, two of the least frequently used combinations were placed lower on the grid because children are not taught those combinations until Grade 3. During the baseline in Beed, it was observed that the children found letter name identification and syllable identification tasks very simple. As a result, 20 percent of the most frequent letter names and syllables were replaced by the least frequent letter names and syllables as identified from Grade 1 and 2 Marathi books.

- The nonword subtask was replaced by the familiar word reading subtask. Due to the nature of the Marathi language, reading of nonwords assesses the same decoding skills as the syllable identification task. For this reason, the nonword reading subtask was removed and replaced with familiar word reading.
- Stories were revised with simpler, more familiar words and simpler sentence formations. Listening comprehension questions were revised to form more direct and simpler questions than in the earlier versions.

The final EGRA assessment included the following six subtasks for the baseline instrument:

1. Letter name identification
2. Syllable identification
3. Familiar word reading
4. Oral reading fluency (ORF)
5. Reading comprehension
6. Listening comprehension

The full EGRA instrument can be found in Annex C, and the same EGRA instrument will be used at endline.

In addition to the EGRA, a demography questionnaire was administered to parents of the students participating in the study. The questionnaire captured details such as education level of parents, occupation, religion, caste, assets at home, smartphone ownership, and access to reading materials at home.

Sample

The Play.Connect.Learn program will reach 12,000 Grade 1 and 2 students, while the research study will only provide results on a subset of the total program population. The Play.Connect.Learn app can only be used on smartphones, and, as noted earlier, exploratory research conducted by SWI indicated that approximately only eight percent of families in the original target intervention areas owned smartphones. Further, research indicated that the students in families who own a smartphone have characteristics that are not necessarily representative of the average student population throughout India, such as higher socio-economic status than their peers and higher baseline early grade reading skills. The purpose of the research study is to determine if exposure to the app results in improvements in early grade reading skills for the students in the research sample.[4] To align with the research intent, SWI randomly selected participants into the

[4] The research sample is only representative of students in Maharashtra with similar characteristics as those included in the research sample. Research findings are not generalizable to a wider population throughout India.

program and will provide smartphones to any families who were selected that did not already own a smartphone. By randomly selecting students for the study rather than only allowing students with preexisting access to a smartphone, SWI removed the risk that the research findings would not represent the average student population in Maharashtra. This means, however, that the students in the research study are not representative of the other approximate 11,200 students who will be receiving the Play.Connect.Learn program. As a result, the findings from the research study should not be generalized to the entire student population receiving the Play.Connect.Learn program.

SWI determined that, in order to detect a significant improvement between the intervention and comparison groups, a minimum sample size of 325 in each group would be necessary.[5] The program is community-based and implemented through SHGs, and, because of limited family migration, low levels of attrition are anticipated. It is possible, however, that program participant attrition may occur because of other factors, such as smartphones getting damaged. SWI decided that an additional 20 percent should be added to each group to account for attrition throughout the implementation period.

Table 3: Research Study Sample Size Per Group

Target District	Proportion of Total Student Population	Research Sample per Group
Amravati	19.7%	79
Sangli	23.6%	94
Kohlapur	7.9%	31
Buldana	27.6%	110
Yavatmal	15.7%	63
Chandrapur	5.5%	23
Total	100%	400

SWI and Disha Research Group, the organization hired to collect baseline data, determined research sample sizes per district. First, they calculated the number of students that will be reached per district out of the total student population of 12,000. Next, they determined the proportion of the total student population that will be reached per district. Finally, they multiplied the total desired sample size per intervention group (400) by the proportion of the total student population for each district to get the research sample size per intervention group as presented in Table 3.

In order to reach the research sample size in the intervention and comparison groups, Disha Research Group randomly selected clusters of villages from each target district and then randomly selected families with Grade 1 and 2 students to be incorporated into the

[5] Sample sizes were determined using a Wilcoxon-Mann-Whitney test with one tail, a normal parent distribution, an effect size of 0.2, and a α error probability of 0.05.

sample. This process was repeated until the required sample sizes were met in each location. The sampling process is detailed below:

1. **Step 1:** SWI requested all implementing NGO partners to combine villages in target districts into clusters. Each cluster included villages that are situated geographically close to each other.
2. **Step 2:** For each district, SWI randomly assigned clusters into either intervention or comparison clusters. This ensured that the intervention and comparison villages were not too close to each other geographically, thus minimizing the risk of contamination.
3. **Step 3:** The clusters selected into the research study were randomly assigned a number from 1 to 10 for both the intervention and comparison groups. The clusters were visited in numerical order and families with Grade 1 or 2 students were assessed until the total sample size per district was reached. This ensured that there was no geographical bias in the data collection process.
4. **Step 4:** The assessors were requested to engage NGO partner volunteers in each village to help coordinate the data collection process. Volunteers were contacted two days prior to the assessors' arrival in the village in order to mobilize families that met the research criteria in advance and arrange for a time to conduct the assessment.

Volunteers from the partner NGOs helped identify families with Grade 1 or 2 students that met the following selection criteria for the research study:

1. Families with a child in either Grade 1 or 2 that agree to participate in the study (only one child per household was included in the sample)
2. Families with a monthly household income of INR 7,000[6] or less

Families did not need to have a smartphone to be eligible to participate in the research sample. However, they were asked if they had a smartphone at home to determine how many units would need to be procured for the intervention group. Further, a family did not need to be a part of an SHG, although they were requested to join in order to facilitate the logistics of distributing the app.

Table 4 details the total number of students assessed by intervention group and gender.

[6] 7,000 Rupees monthly is approximately $102 USD.

Table 4: Total Number of Students Assessed by Group and Gender

Group	Boys	Girls	Total
Intervention	200	202	402
Comparison	197	203	400
Total: All students	397	405	802

Results from the demography questionnaire administered at baseline indicated that 60percent of families in the research sample had an income of less than INR 5,000 per month, which is under the selection criteria of INR 7,000 or less per month (see Figure 2). About 40 percent had an income of between INR 5,000 and 10,000 monthly, and the remaining one percent of households had an income of between INR 10,001 and 30,000. No households in the research sample reported earning more than INR 50,000 monthly.

Figure 2: Family Household Income (INR) (%)

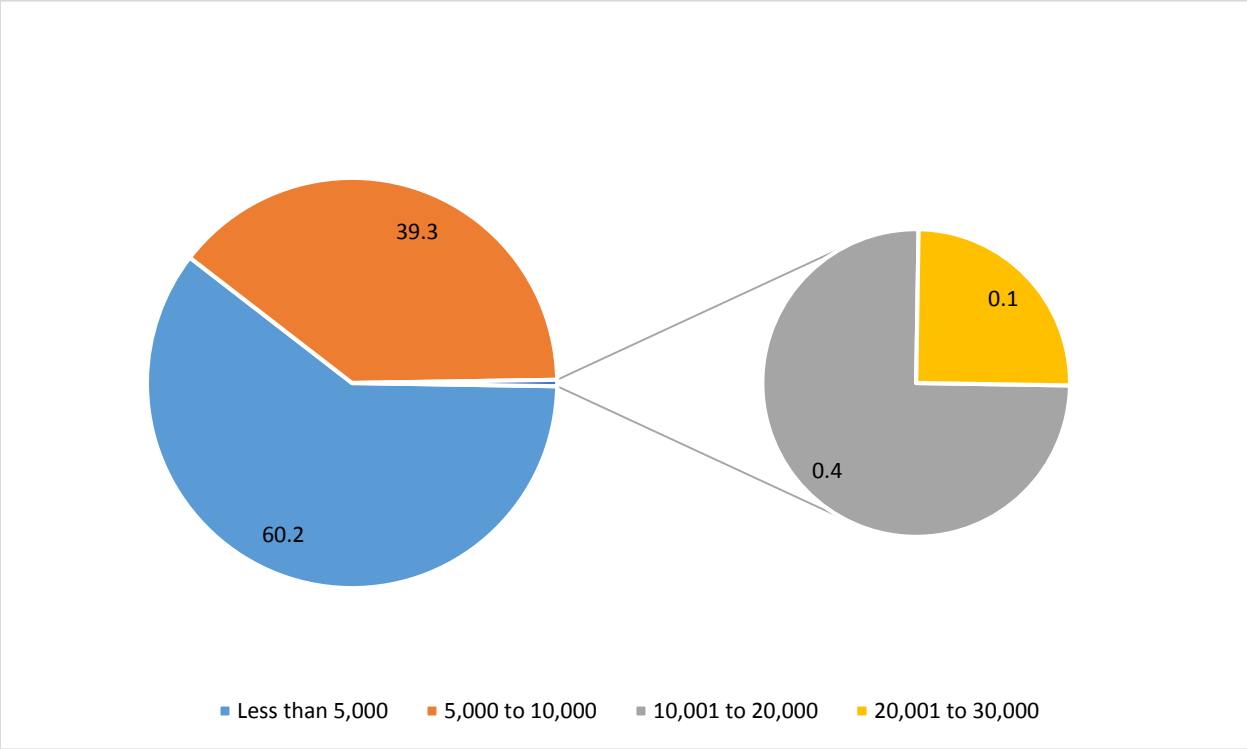
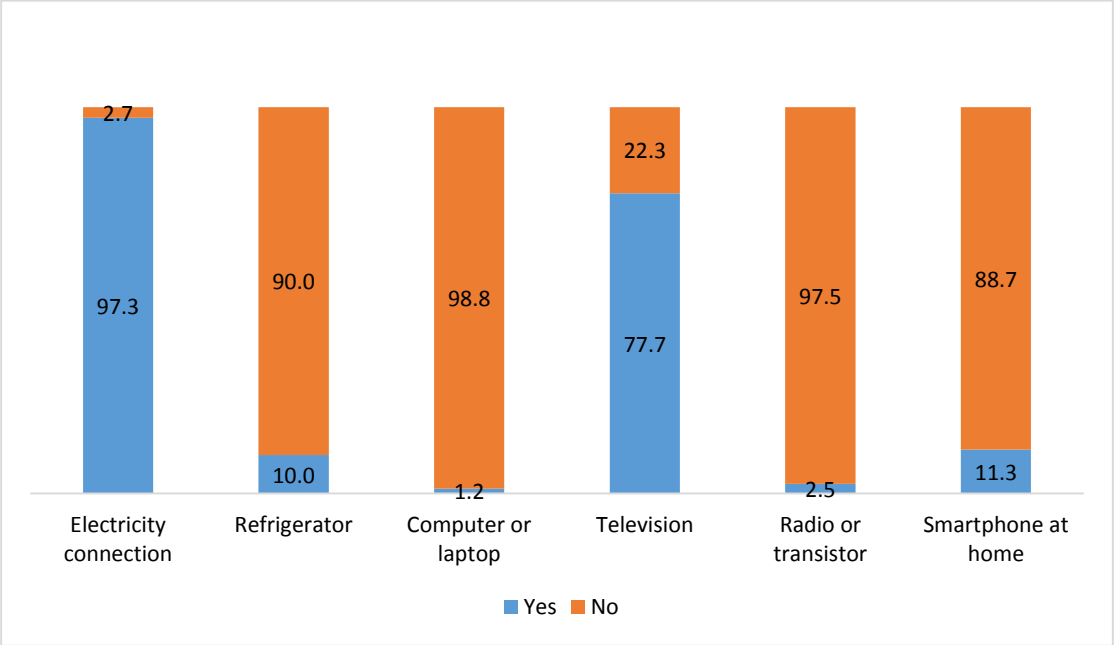


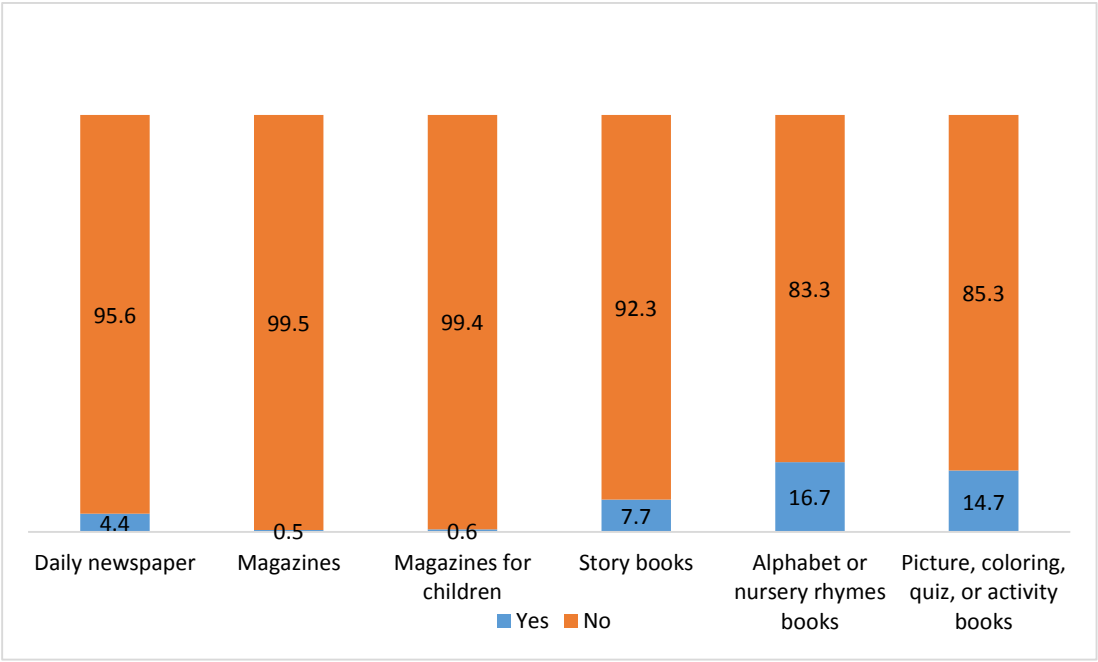
Figure 3 provides a summary of assets for the households in the research sample. Nearly all households had electricity connection, while only one percent of households had a desktop or laptop computer. About 11 percent of households had a smartphone.

Figure 3: Household Assets (%)



Household access to reading materials is summarized in Figure 4. Households had very few reading materials: eight percent reported having children's story books, 17 percent reported having alphabet or nursery rhyme books, and 15 percent reported having picture, coloring, quiz, or activity books.

Figure 4: Household Reading Materials (%)



V. Fieldwork Preparation and Data Collection

Assessor Training

After the Marathi EGRA tool was finalized, baseline assessors from Disha Research Group attended training from February 1–5, 2016, in advance of the first baseline data collection. The training consisted of the following activities:

- Review the EGRA principles and gain a comprehensive understanding of the EGRA instrument components
- Practice EGRA administration and scoring procedures
- Practice conducting the EGRA assessment on tablets
- Become familiar with the roles and responsibilities of both supervisors and assessors in the field
- Undergo interrater reliability (IRR) administration and scoring evaluation

Following the training, assessors pilot tested the Marathi EGRA tool and demographic questionnaire on 80 students in five *Zilla Parishad* schools in Beed district. Feedback from the pilot test was incorporated, and all data collection instruments were finalized.

Following the program redesign, an assessor refresher training was held on June 25–26, 2016, prior to the second operational baseline data collection.

Interrater Reliability (IRR) Test

IRR is a measure of reliability used to assess the degree to which different assessors agree in their assessment decisions. IRR tests ensure that the different assessors interpret answers in the same way. At least 90 percent consistency is considered the minimum requirement—meaning that at least 90 of assessors' ratings are consistent with the list of acceptable responses.

IRR tests for the assessors were conducted during the initial and refresher trainings prior to the operational baseline data collections. All assessors met the 90 percent threshold.

Institutional Review Board (IRB) for Human Participants[7]

The Institutional Review Board (IRB) is responsible for ascertaining the acceptability of proposed research in terms of institutional commitments and regulations, applicable laws, standards of professional conduct and practice, and ethical and societal norms. The IRB examines subject recruitment procedures, proposed remuneration, and the informed

[7] Following The Protection of Human Subjects in Research Supported by USAID, all ACR GCD projects sought human subjects approval through a local Institutional Review Board (IRB) to ensure there was minimal risk to the students participating in the interventions and associated assessments.

consent process. The IRB also evaluates the potential risks and benefits to participants outlined in each protocol.

SWI did not request IRB approval for the Play.Connect.Learn program. Permission to conduct research was granted by the Beed District Magistrate for the initial baseline EGRA assessment; however, permission from the District Magistrates of the final implementation districts was not requested due to SWI oversight. SWI conducted all research in compliance with publicly available information on guidelines for conducting educational research in India, and SWI is currently seeking additional IRB approval based on the final research design.

Data Collection

Tablets with the Tangerine[8] software were used to administer the EGRA assessment and to capture results. Tangerine is an open source electronic data collection software designed for mobile devices. All assessors were provided with tablets for use during data collection.

The operational baseline data collection was conducted from June 26–August 3, 2016. Nine assessors from Disha Research Group conducted the assessment. Assessors were supported by four team leaders who coordinated with NGO volunteers to identify families and students and arrange for assessment time slots. The team leaders also provided additional field and technical support to the assessors, such as providing additional information to concerned families or providing charged tablets when batteries ran out.

Data Analysis

Baseline data were analyzed using Stata and IBM SPSS Statistics softwares, as well as Microsoft Excel. Differences between the results of the intervention and comparison groups were tested for significance; where found, these differences were noted. Mean scores of multiple groups were compared using ANOVA, or Analysis of Variance, a statistical strategy that is used to analyze the differences between group means. Differences in the proportion of students who scored zero (or non-readers) were compared using the chi-square test for significance.

[8]Tangerine® is an electronic data collection software designed for use on mobile computers, including netbooks, tablet computers and smartphones. Its primary use is to enable recording of children's responses in oral early grade reading and mathematics skills assessments, specifically Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA), and interview responses from children, teachers and principals on home and school context information http://tangerinecentral.org/.

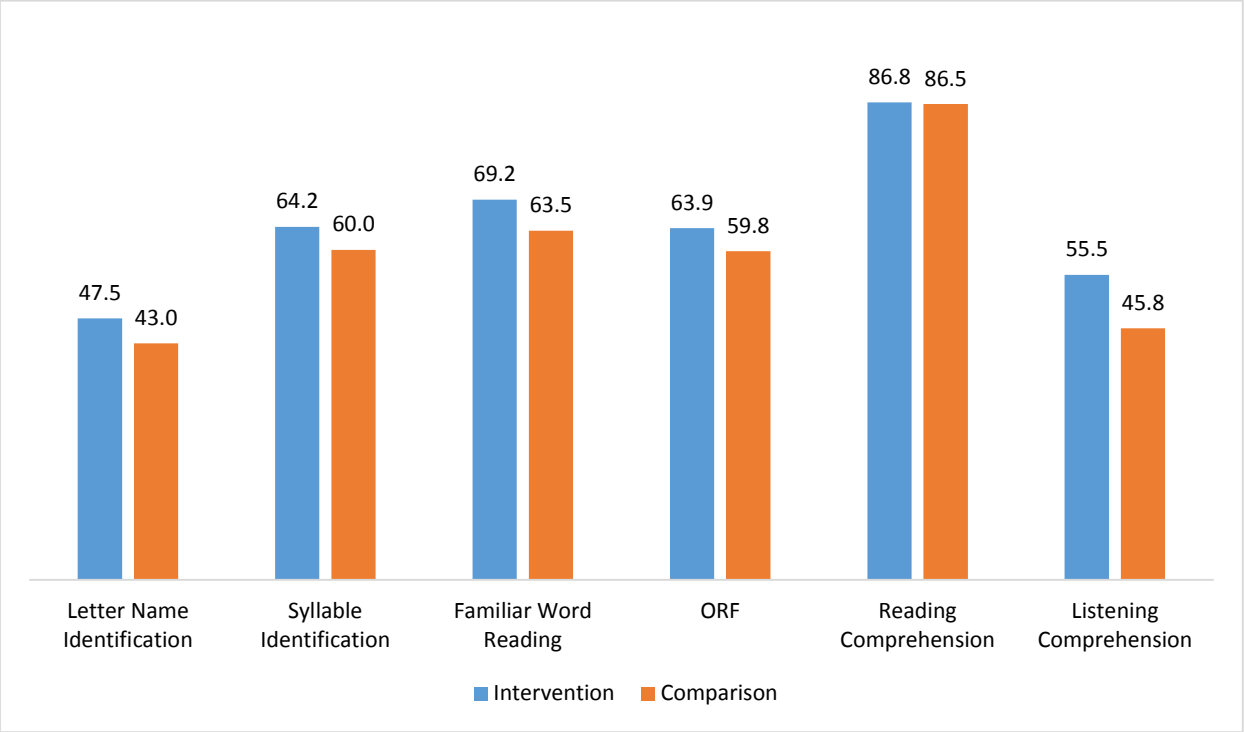
Table 5: Subtask and Data Analysis Methods

Subtask	Type	Analysis
Letter name identification	Timed	Letter name identification is measured as correct letters named per minute (CLNPM). Letter name identification is a measure of alphabet knowledge and is highly predictive of later reading achievement. Each student had one minute to name up to 100 letters.
Syllable identification	Timed	Syllable identification is measured as correct syllable sounds read per minute (CSSPM). Syllable identification is a measure of compound letter sound correspondence and is a more advanced predictor of decoding ability. Each student had one minute to read up to 100 syllable sounds.
Familiar word reading	Timed	Familiar word reading is measured as the number of correct familiar words read per minute (CFWPM). Familiar word reading measures word recognition and decoding. Each student had one minute to read up to 50 high-frequency words.
Oral reading fluency (ORF)	Timed	ORF is measured as correct words read per minute (CWPM). ORF is a decoding and reading fluency measure. Each student had one minute the opportunity to read 59 words. The ORF passage formed the textual basis for the reading comprehension subtask.
Reading comprehension	Untimed	Reading comprehension is measured as the number of correct answers verbally delivered to the assessor based on questions asked about the passage read as part of the ORF subtask. Each student had the opportunity to answer four factual questions and one inferential question.
Listening comprehension	Untimed	Listening comprehension is measured as the number of correct answers verbally delivered to the assessor. Listening comprehension is a measure of vocabulary. Each student had the opportunity to answer four questions based on a passage read to them by the assessor.

VI. Summary of Findings

Results in Figure 5 show the percent of zero scores by group and EGRA subtask. Students in the intervention and comparison groups performed similarly across subtasks. Out of all subtasks, the largest proportion of students received zero scores on reading comprehension, and the lowest proportion of students received zero scores on letter name identification.

Figure 5: Percent of Zero Scores by Group (%)



VII. Results by Group and Gender

The following section provides detailed baseline results for each EGRA subtask by intervention group and gender (see Table 5 for complete subtasks and analysis method). Each subsection contains a description of the subtask followed by the mean score (untimed subtasks) or mean fluency rate (timed subtasks), standard deviation (SD),[9] and number of zero scores.

Letter Name Identification

Letter name identification measures students' knowledge of the alphabet and is predictive of later reading success. For this subtask, students were presented with a stimulus of 100 letters and were asked to name as many as they could in one minute. The subtask was discontinued if a student was unable to correctly name any of the first ten letters of the stimulus.

The mean fluency rates, reported as correct letter names per minute (CLNPM), are presented in Table 6. **On average, students were able to name 15.2 letters in one minute.** Students in the intervention group named an average of 14.3 letters, while students in the

[9] The standard deviation (SD) of the measure of interest—here, mean fluency rates—describes the spread between scores. Smaller SD values indicate that the majority of values lie close to the mean; larger SD values indicate that mean fluency rates varied and were more spread out.

comparison group named an average of 16.2. The mean difference in fluency rates between intervention groups was not statistically significant, indicating that students in the intervention and comparison groups were not significantly different at baseline on this subtask.[10]

Overall, about 45 percent of students could not name a single letter at baseline. The proportion of students who were unable to name a single letter was slightly higher in the intervention group (48 percent) than in the comparison group (43 percent). In the intervention group, 48 percent of boys were unable to name any letters, and 47 percent of girls were unable to name any letters. In the comparison group, 47 percent of boys were unable to name any letters, and 39 percent of girls were unable to name any letters. The difference in the proportion of students receiving zero scores between intervention and comparison groups at baseline was not statistically significant.[11]

Table 6: Letter Name Identification Fluency by Group and Gender

Group	Gender	N	Mean Fluency (CLNPM)	SD	Zero Scores (n)
Intervention	Boys	200	13.0	18.1	96
	Girls	202	15.5	20.2	95
	Total	402	14.3	19.2	191
Comparison	Boys	197	15.3	26.1	92
	Girls	203	17.1	20.4	80
	Total	400	16.2	23.4	172
Total: All Students		802	15.2	21.4	363

Syllable Identification

The syllable identification subtask measures students' understanding of how letter combinations correspond to a specific sound. To demonstrate syllable identification, students must identify the appropriate sounds for each syllable. The ability to match letters with correct sounds is critical to reading fluency and comprehension. For this subtask, each student was presented with a stimulus of 100 syllables and asked to read as many of the sounds as they could in one minute. The subtask was discontinued if a student was unable to correctly identify any of the first 10 syllables on the stimulus. Results for the syllable identification subtask are presented in Table 7.

[10] p=0.19
[11] Chi2(1)=1.648, p=0.199

On this subtask, students were able to correctly identify an average of 6.9 syllable sounds in one minute. Students in the intervention group had a slightly lower average fluency rate than students in the comparison group, but the difference was not statistically significant.[12] Girls in both the intervention and the comparison groups were able to identify more syllable sounds on average than boys.

Table 7: Syllable Identification Fluency by Group and Grade

Group	Gender	N	Mean Fluency (CSSPM)	SD	Zero Scores (n)
Intervention	Boys	200	5.8	12.3	132
	Girls	202	7.8	13.9	126
	Total	402	6.8	13.2	258
Comparison	Boys	197	6.1	11.9	124
	Girls	203	7.9	13.4	116
	Total	400	7.0	12.7	240
Total: All Students		802	6.9	12.9	498

Students receiving zero scores on the syllable identification subtask were higher than on the letter name identification subtask. About 64 percent of students in the intervention group and 60 percent of students in the comparison group were unable to correctly identify a single syllable sound in one minute, and **overall, 498 students out of 802 could not identify a single syllable sound.** Girls received proportionally lower zero scores than boys in both intervention and comparison groups. The difference in the proportion of students with zero scores between intervention and comparison groups at baseline was not statistically significant.[13]

Familiar Word Reading

In the familiar word reading subtask, students were presented with 50 familiar words[14] and asked to read as many as they could within one minute. Knowledge of familiar words and the ability to read them quickly enables a child to read with automaticity—a skill critical to learning to read with fluency and comprehension. The subtask was discontinued if a child was unable to correctly name any of the first five familiar words.

Results for the familiar word reading subtask are presented in Table 8. **On average, students were able to correctly read 5.3 familiar words per minute.** Intervention students correctly read 4.5 words, while comparison students correctly read six words.

[12] p=0.81

[13] Chi2(1)=1.488, p=0.223

[14] The words in this subtask were derived from frequently used words for the age group.

The difference between the fluencies for intervention and comparison students was not statistically significant.[15] Girls read more words correctly than boys in both intervention groups.

Table 8: Familiar Word Reading Fluency by Group and Grade

Group	Gender	N	Mean Fluency (CFWPM)	SD	Zero Scores (n)
Intervention	Boys	200	3.4	7.4	147
	Girls	202	5.7	10.6	131
	Total	402	4.5	9.2	278
Comparison	Boys	197	5.6	24.9	131
	Girls	203	6.4	12.4	123
	Total	400	6.0	19.5	254
Total: All Students		802	5.3	15.3	532

In total, 532 students—or about 66 percent—were unable to identify a single familiar word. The percentage of girls with zero scores was lower than boys in both the intervention and comparison groups, and the percentage of students with zero scores was higher in the intervention group (about 69 percent) than in the comparison group (about 64 percent). The difference in the proportion of students receiving zero scores between intervention and comparison groups at baseline was not statistically significant.[16]

Oral Reading Fluency (ORF)

For the ORF subtask, the assessor provided each student with a story of 59 words to read in one minute. ORF is perhaps the strongest predictor of reading comprehension. Along with skills like decoding and vocabulary, ORF is a strong predictor of comprehension because to understand groups of words, a certain amount of automaticity is required so that the reader can store what is read in working memory. If a student reads too slowly, he or she may be unable to remember all the words in a sentence and thus not understand the story's meaning.

Results for the ORF subtask are presented in Table 9. **Students, on average, were able to correctly read 6.3 words on the ORF subtask.** Comparison students correctly read nearly seven words in one minute compared to the nearly six words correctly read by students in the intervention group, though the difference was not statistically significant.[17] As with

[15] p=0.17
[16] Chi2(1)=2.870, p=0.090
[17] p=0.38

the previous subtasks, girls in both the intervention and comparison groups had higher fluencies than boys.

Nearly 62 percent of students were unable to read a single correct word on the ORF subtask. The proportion of students receiving zero scores were highest in boys in the intervention group and lowest in girls in the comparison group. Overall, a higher proportion of students in the intervention group—nearly 64 percent—received zero scores than those in the comparison group. The difference in the proportion of students receiving zero scores between intervention and comparison groups at baseline was not statistically significant.[18]

Table 9: ORF by Group and Grade

Group	Gender	N	Mean Fluency (CWPM)	SD	Zero Scores (n)
Intervention	Boys	200	4.5	9.6	133
	Girls	202	7.3	13.7	124
	Total	402	5.9	11.9	257
Comparison	Boys	197	5.0	10.2	125
	Girls	203	8.3	15.9	114
	Total	400	6.7	13.5	239
Total: All Students		802	6.3	12.7	496

Reading Comprehension

For the reading comprehension subtask, the assessor removed the story from the ORF subtask, then asked each student five comprehension questions based on what they read. Comprehension is the purpose of reading. Once a child learns the sound-letter relationship (alphabetic principle) and becomes able to decode and read with automaticity, he or she becomes increasingly able to understand the meaning of a text. This subtask assesses that ability.

The number of questions students were asked was dependent on how many words they were able to read on the ORF subtask. For instance, if a student read the first ten words, he or she would be asked the first comprehension question. Similarly, if a student read all 59 words, he or she would be asked all five questions. Students who received zero scores on the ORF subtask received a zero score on the reading comprehension subtask.

[18] Chi2(1)=1.485, p=0.223

Reading comprehension results are presented in Table 10. **On average, students were able to correctly answer 0.3 reading comprehension questions out of a maximum of five.** Students in the intervention group were able to answer an average of 0.2 questions correctly, and students in the comparison group were able to answer an average of 0.3 questions correctly. This difference was not statistically significant.[19]

Table 10: Reading Comprehension Score by Group and Grade

Group	Gender	N	Mean Score (Number of Questions Correct)	SD	Zero Scores (n)
Treatment	Boys	200	0.2	0.6	180
	Girls	202	0.3	0.9	169
	Total	402	0.2	0.7	349
Control	Boys	197	0.2	0.7	176
	Girls	203	0.4	1.0	170
	Total	400	0.3	0.8	346
Total: All Students		802	0.3	0.8	695

Because of low ORF fluencies and, as a result, many students not being asked any comprehension questions, the number of students receiving zero scores on this subtask was the highest out of all the subtasks. **About 87 percent of students were not asked any reading comprehension questions or were unable to correctly answer a single reading comprehension question**. The difference in the proportion of zero scores between intervention and comparison groups at baseline was not statistically significant.[20]

Listening Comprehension

The listening comprehension subtask is an untimed assessment of children's abilities to comprehend the meaning of a story read to them orally. Students do not need to know how to read to answer listening comprehension questions. As a result, this subtask is an important measure of students' pre-reading abilities because it helps detect obstacles that prevent them from learning to read, such as limited language proficiency, auditory problems, attention deficit and other difficulties. In this subtask, the assessor read a short passage to the student and asked him or her to answer four comprehension questions about the passage.

Results for the listening comprehension subtask are presented in Table 11. **On average, students correctly answered nearly one question out of a maximum of four.** Students

[19] p=0.54
[20] Chi2(1)=0.017, p=0.895

in the comparison group correctly answered an average of one question, and students in the intervention group correctly answered an average of 0.8 questions. The difference in means between the intervention and comparison groups was statistically significant.[21]

Table 11: Listening Comprehension by Group and Grade

Group	Gender	N	Mean Score (Number of Questions Correct)	SD	Zero Scores (n)
Intervention	Boys	200	0.9	1.1	107
	Girls	202	0.8	1.1	116
	Total	402	0.8	1.1	223
Comparison	Boys	197	1.1	1.2	87
	Girls	203	1.0	1.2	96
	Total	400	1.0	1.2	183
Total: All Students		802	0.9	1.2	406

Over half of students—about 51 percent—were unable to correctly answer a single listening comprehension question. A larger proportion of students in the intervention group had zero scores than in the comparison group—about 57 percent versus about 46 percent, respectively. Further, girls had higher proportions of zero scores than boys in both the intervention and comparison groups on this subtask. The difference in the proportion of zero scores between intervention and comparison groups at baseline was statistically significant.[22]

VIII. Conclusions and Recommendations

The results presented in this baseline report suggest that the research study population needs targeted support to increase letter name identification and syllable reading skills, which may in turn increase reading fluency and comprehension. Over 66 percent of students could not read a single familiar word, and nearly 62 percent of students could not read a single word in the oral reading fluency subtask. While there are no known Marathi fluency benchmarks for letter naming, familiar word, and passage reading, it is clear that more needs to be done to support greater decoding ability.

The following recommendations should be considered by SWI during the implementation of the Play.Connect.Learn program:

1. **Ensure stories and games reinforce and expand students' letter name and syllable identification skills.** Create practice activities that support repetition of

[21] p=0.01
[22] Chi2(1)=7.582, p=0.006

these foundational skills, which are needed to increase students' ability to decode more complex texts such as words and passages.

2. **Provide support for vocabulary development.** Students were only able to read about five familiar words on average in one minute. The stories in the Play.Connect.Learn app should highlight key familiar words that can be used as building blocks for increased word reading skills. The stories can cycle through the same words and gradually add more as a way to build both comprehension and sight recognition.
3. **Explore ways to support boys to further develop their reading skills.** The baseline results indicated that girls performed better than boys and received lower proportions of zero scores on all subtasks except for listening comprehension. The project should seek to better understand and address any underlying issues that might be causing this disparity.

IX. Annexes

Annex A: EGRA Adaptation Workshop Agenda

Adaptation Workshop Agenda
Tuesday, October 6
Daily objective: Understanding of EGRA purpose and content, and components for letter name identification, syllable awareness

9:30-10:30 a.m.	Welcome and introductions Kristina Solum, School to School International (STS) Madhu Ranjan, USAID India Jacob Devabhaktula, World Vision India Hanumanta Talar, Benetech Siddharth Pillai, Sesame Workshop India Trust
10:30-10:45 a.m.	Break
10:45-12:30 p.m.	Overview of EGRA and review of components: Kristina Solum, STS
12:30-1:30 p.m.	Lunch
1:30-3:00 p.m.	Prepare components: letter name identification and syllable identification
3:00-3:15 p.m.	Break
3:15-5:30 p.m.	Continued

Wednesday, October 7
Daily objective: Oral reading and listening comprehension components, non-word reading

9:30-11:00 a.m.	Prepare component: oral reading fluency (ORF) passages
11:00-11:15 a.m.	Break
11:15-12:30 p.m.	Continue ORF stories and develop questions
12:30-1:30 p.m.	Lunch
1:30-3:00 p.m.	Prepare component: listening comprehension passages
3:00-3:15 p.m.	Break
3:15-5:30 p.m.	Prepare component: Non-word reading

Thursday, October 8
Daily objective: Pupil questionnaire and preparation for pre-testing

9:30-11:00 a.m.	Review and update pupil questionnaire
11:00-11:15 a.m.	Break
11:15-12:30 p.m.	Review and finalization of all components
12:30-1:30 p.m.	Lunch
1:30-2:30 p.m.	Prepare for pre-testing
3:00-3:15 p.m.	Break
3:15-5:30 p.m.	Tangerine testing

Friday, October 9
Daily objective: Pre-test the tool Marathi and Braille tools

9:30-12:30 a.m.	Visit local schools to test EGRA tools in Marathi and in Braille

12:30-1:30 p.m.	Lunch
1:30-2:30 p.m.	School visit debrief
3:00-3:15 p.m.	Break
3:15-5:30 p.m.	Begin finalizing tools

Saturday, October 10

Daily objective: Finalize the tools

9:30-5:30	Finalize tools
	Breaks at 10:30 and 2:30

Annex B: Statistics by EGRA Subtask

Letter Name Identification – Zero Score

Group	No	Yes	Total
Comparison	228	172	400
	57.00%	43.00%	100.00 %
Intervention	211	191	402
	52.49%	47.51%	100.00 %
Total	439	363	802
	54.74%	45.26%	100.00 %

Chi2(1)=1.648, p=0.199

Syllable Identification – Zero Score

Group	No	Yes	Total
Comparison	160	240	400
	40.00%	60.00%	100.00 %
Intervention	144	258	402
	35.82%	64.18%	100.00 %
Total	304	498	802
	37.91%	62.09%	100.00 %

Chi2(1)=1.488, p=0.223

Familiar Word Reading – Zero Score

Group	No	Yes	Total
Comparison	146	254	400
	36.50%	63.50%	100.00%
Intervention	124	278	402
	30.85%	69.15%	100.00 %
Total	270	532	802
	33.67%	66.33%	100.00 %

Chi2(1)=2.870, p=0.090

ORF – Zero Score

Group	No	Yes	Total
Comparison	161	239	400
	40.25%	59.75%	100.00 %
Intervention	145	257	402
	36.07%	63.93%	100.00 %
Total	306	496	802
	38.15%	61.85%	100.00 %

Chi2(1)=1.485, p=0.223

Reading Comprehension – Zero Score

Group	No	Yes	Total
Comparison	54	346	400
	13.50%	86.50%	100.00 %
Intervention	53	349	402
	13.18%	86.82%	100.00 %
Total	107	695	802
	13.34%	86.66%	100.00 %

Chi2(1)=0.017, p=0.895

Listening Comprehension – Zero Score

Group	0	1	Total
Comparison	217	183	400
	54.25%	45.75%	100.00 %
Intervention	179	223	402
	44.53%	55.47%	100.00 %
Total	396	406	802
	49.38%	50.62%	100.00 %

Chi2(1)=7.582, p=0.006

Annex C: Baseline EGRA Instrument

Sesame Baseline Paper EGRA.pdf

Annex D: Demography Tool

ACR2_Demography questionnaire_revised