Literacy Instruction

Scale-up Program in Jodhpur District of Rajasthan, India

Program Impact Evaluation (2023-25) – 2025 Endline Evaluation Report

Room to Read®

Prepared by:
Pinaki Joddar, Japheth Ogol and Rebecca Westbrook
Global Literacy Research, Monitoring and Evaluation Team

With support from:
Room to Read India

May 2025

Acronyms

CLPM	Correct letters per minute
CNWPM	Correct nonwords per minute
CWPM	Correct words per minute
DQA	Data Quality Assessments
EGLSA	Early Grade Literacy Skills Assessment
FLN	Foundational Literacy and Numeracy
IRR	Inter-rater reliability assessment
IRT	Independent Reading Time
MT	Master Trainers
OLS	Ordinary Least Squares
ORF	Oral Reading Fluency
PEEO	Panchayat Elementary Education officer
PFC	Program Field Coordinators
PTC	Parent-Teacher Committee
SD	Standard Deviation
SY	School Year
TLM	Teaching and Learning Materials

Table of Contents

List of Tables

List of Figures

Executive Summary

Since 2003, Room to Read has partnered with the Government of Rajasthan in India to improve children's literacy skills in the foundational phase and to develop reading habits for all primary grade children in government primary schools. In 2022, Room to Read initiated an intervention, Literacy Instruction Scale-up Program (henceforth referred to as the program), to advance these goals in Jodhpur district between 2022-25 with funding support from the Indira Foundation. The program aims to improve literacy skills of the children in grades 1-3 in Hindi and promote reading habits across primary grades in approximately 2,600 schools.

The program's core components include: 1) the provision of teaching and learning materials (TLMs) for teachers' and children's use during Hindi language classes in grades 1-3; 2) capacity strengthening training for system-based actors via a training-of-trainers (TOT) model; 3) the establishment of cluster libraries and classroom reading corners; and 4) on-site school-based coaching and mentoring support for the teachers.

In 2023, Room to Read initiated a 2-year (2023-25) evaluation of the program to assess the program's effectiveness at achieving key inputs, intermediary outputs and improved children's literacy skills. The evaluation used a repeated cross-sectional design to assess and compare the changes in end-of-Garde 3 children's literacy skills during 2023-25 in a sample of intervention schools in Jodhpur (defined as program group) and a sample non-intervention schools in Barmer (defined as comparison group). Here, the pre-intervention baseline and the post-intervention endline literacy skills assessment data from two different cohorts of end-of-Grade 3 children in program group was collected in May 2023 and March 2025, respectively. The baseline and endline literacy skills assessment data from the children in comparison group schools in Barmer was also collected at the same two points in time. Children were assessed using a Grade 3-levelled Early Grade Literacy Skills Assessment (EGLSA) Tool comprised of five Hindi reading and writing tasks: letter naming, nonword reading, oral passage reading, reading comprehension and sentence dictation.

Evaluation Results – Children's Early Grade Literacy Skills

Children in program schools achieved significantly greater gains in mean scores and larger reductions in zero scores across all five literacy skill tasks than their peers in comparison schools. Key findings from the oral reading fluency and reading comprehension tasks are included in Table 1 and can be summarized as follows:

- **Oral Reading Fluency (ORF)** mean scores grew by 24.5 correct words per minute (cwpm) in program schools (from 18.6 to 43.1 cwpm) as compared to 6.1 cwpm in comparison schools (from 25.1 to 31.2 cwpm) during 2023-25 ($p < 0.001$). At the same time, the proportion of children scoring zero on the same task declined from 35 to 10 percent in program schools versus 27 to 23 percent in comparison schools ($p < 0.001$).
- **Reading comprehension scores** (out of 6) also improved by 1.1 points (from 0.6 to 1.7) in program schools versus 0.5 points in comparison school (from 1.1 to 1.6) during 2023-25 ($p < 0.01$). Moreover, the prevalence of zero scores on the reading comprehension task fell from 66 to 30 percent in program schools versus 53 to 38 percent for comparison schools ($p < 0.01$). Despite the statistically significant higher gains in average score and sharp decline in

the zero scores during 2023-25, comprehension skills of the children in program schools, in general, was found to be low during the 2025 endline evaluation.

Table 1. Early Grade Literacy Skill Assessment Results at Baseline and Endline

Indicator	Group	Baseline (2023)		Endline (2025)		Est. diff. b/w the groups [Std Error]
		Count	Mean (SD) % Share	Count	Mean (SD) % Share	
Mean Scores						
Oral reading fluency (cwpm)	Program	820	18.6 (24.1)	695	43.1 (30.1)	16.38*** [3.26]
	Comparison	770	25.1 (29.1)	717	31.2 (26.9)	
Reading comprehension: Score (out of 6)	Program	820	0.6 (1.1)	702	1.7 (1.6)	0.49** [0.19]
	Comparison	770	1.1 (1.5)	724	1.6 (1.6)	
Zero Scores (% share)						
Oral passage Reading	Program	820	35%	702	10%	-1.38*** [0.36]
	Comparison	770	27%	724	23%	
Reading Comprehension	Program	820	66%	702	30%	-0.85** [0.26]
	Comparison	770	53%	724	38%	
Reading Benchmarks (% share)						
Oral reading fluency: 35+ cwpm	Project	820	23%	702	62%	1.28*** [0.26]
	Comparison	770	32%	724	44%	
Oral reading fluency: 45+ cwpm	Project	820	15%	702	45%	1.17*** [0.29]
	Comparison	770	21%	724	27%	
Reading comprehension: 80%+ score	Project	820	1%	702	5%	1.37* [0.64]
	Comparison	770	5%	724	6%	

*Legend of statistical significance of differences between project and comparison schools: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. SD – Standard Deviation.*

Children in program schools were 3-4 times more likely to achieve both Room to Read's and the Government of Rajasthan's oral reading fluency and reading comprehension benchmarks. As shown in Table 1, achievement on the 35+ cwpm ORF benchmark rose from 23 to 62 percent in program schools (a 39-percentage point gain), versus an increase from 32 to 44 percent (a 12-percentage point gain) in comparison schools ($p < 0.001$). Similarly, for reading comprehension score, the proportion of children in program schools achieving the benchmark of 80+% rose from 1 to 5 percent, while for comparison schools it only improved by one percentage point ($p < 0.05$).

Lastly, the program's effect sizes were large, indicating that the program was particularly effective at improving children's foundational reading and writing skills: letter naming fluency (2.3), nonword reading fluency (1.9), oral reading fluency (1.7), reading comprehension (0.9), and sentence dictation (1.3).

Predictors of Children's Early Grade literacy skills at endline

Multivariate regression analyses identified compelling evidence[1] for several predictors of children's literacy skills scores at endline:

- **Children in monograde classrooms outperformed their peers in multigrade classrooms** on the oral reading fluency and reading comprehension tasks ($p < 0.05$ and $p < 0.01$, respectively). They also were less likely to score zero on three of the five assessment tasks.
- **Children in classes taught by male teachers scored higher** on all five assessment tasks than those taught by female teachers. Differences were significant, either at $p < 0.01$ or $p < 0.05$.
- **Boys outperformed girls** in letter naming, nonword reading, oral reading fluency, and reading comprehension tasks ($p < 0.05$ or $p < 0.001$). They were also less likely to score zero on three of five assessment tasks and had a higher chance of achieving all three reading benchmarks[2]. **However, there was only convincing evidence for the influence of this factor in comparison schools as in program schools the difference was only statistically significant for one task**.
- **Children speaking Hindi at home scored lower that other children** on four tasks: nonword reading ($p < 0.05$), oral passage reading ($p < 0.01$), reading comprehension ($p < 0.001$) and sentence dictation ($p < 0.01$). There was compelling evidence for the influence of this factor in program schools, and only weak evidence in comparison schools.
- **Children from homes with a higher household standard of living score** had higher mean scores, fewer zero scores and were more likely to achieve reading benchmarks (the significance varied by task from $p < 0.05$, $p < 0.01$, or $p < 0.001$).
- **The presence of literate adults at home was associated with higher mean scores:** letter naming ($p < 0.01$), nonword reading, reading comprehension and sentence dictation ($p < 0.05$). Children living with literate adults were also more likely to achieve both oral reading fluency benchmarks ($p < 0.05$ for both 35+ and 45+ cwpm benchmarks), but there was no evidence of an association with the likelihood of scoring zero.
- **Children with access to books at home** had higher scores on four out of five assessment tasks ($p < 0.05$ or $p < 0.01$) and fewer zeros across all tasks ($p < 0.001$).
- **Children with strong reading habits,** measured as reading for pleasure at least three times in the last week, were strongly associated with higher mean scores and fewer zero scores ($p < 0.001$ for all assessment tasks). They were also more likely to achieve both oral reading fluency and reading comprehension benchmarks ($p < 0.001$ or $p < 0.01$). Combined, this suggests that this factor has a strong influence on children's literacy skills outcomes.

[1] Strong evidence is defined as a factor having a statistically significant association with at least three out of five EGLSA tasks and there is consistency in the sign and significance across regression models.
[2] Three reading benchmarks are used throughout the report: 1) Government of Rajasthan's ORF benchmark of 35+ cwpm; 2) Room to Read's global ORF benchmark of 45+ cwpm; and 3) Room to Read's global reading comprehension benchmark of 80%+ correct answers or 5 out of the 6 questions answered correctly for the Grade 3 EGLSA tool used in this evaluation.

Conclusion and Recommendations

Room to Read's Literacy Instruction Scale-up Program achieved substantial and equitable improvements in foundational Hindi literacy skills among children in Grade 3, significantly outperforming comparison schools on all literacy skills assessed, while reducing non-reader rates. Implementation fidelity was strong for materials distribution, library establishment, and training; however, frequency of teacher coaching support – particularly by system-based actors like Panchayat Elementary Education officer (PEEOs) – and multigrade classroom support warrant attention.

Based on these findings, recommendations to sustain and amplify the gains, include:

- **Enhance Fidelity of Coaching Support**: clarify expectations and strengthen accountability mechanisms to improve the frequency and quality of PEEO-led coaching support.
- **Targeted Support for Multigrade Classrooms:** develop specialized teaching modules, coaching support and materials tailored to multigrade classrooms, coupled with differentiated-instruction training.
- **Empower Female Teachers**: offer mentorship and peer-learning opportunities to bolster female teachers' instructional practices and support systems.
- **Improve Attendance and Grade Progression**: implement monitoring mechanisms to identify at-risk children, deploying targeted remediation and community engagement.
- **Sustain Reading Habits:** expand adoption of book-checkout processes to improve children's reading habits at home and integrate "reading for fun" routines in school.
- **Strengthen Community Engagement:** organize parent-teacher forums and home-visit initiatives to reinforce literacy support within households.
- **Investigate Home Language Findings**: conduct additional investigation to understand the underperformance of the Hindi-speaking children and design responsive interventions.

Introduction

Founded in 2000 on the belief that World Change Starts with Educated Children®, Room to Read is creating a world free from illiteracy and gender inequality through education. Room to Read is achieving this goal by helping children in historically low-income communities develop literacy skills and a habit of reading, and by supporting girls as they build life skills to succeed in school and negotiate key life decisions. We collaborate with governments and other partner organizations to deliver positive outcomes for children at scale. To date, Room to Read has benefited more than 45 million children in more than 213,000 communities across 24 countries.

Since 2003, Room to Read has partnered with the Government of Rajasthan in India to improve children's literacy skills in the foundation phase and develop reading habits for all primary grade children in government primary schools. In 2022, Room to Read initiated an intervention, Literacy Instruction Scale-up Program (henceforth referred to as the program), to advance these goals in Jodhpur district between 2022 to 2025 with funding support from the Indira Foundation. The program aims to improve Hindi literacy skills in Grades 1–3 and promote reading habits across primary grades in approximately 2,600 schools.

In 2023 Room to Read also initiated a 2-year (2023-25) evaluation of the program to assess the program's effectiveness at achieving key inputs, intermediary outputs and improved children's literacy skills outcomes. Moreover, the results will provide vital information to inform both internal and external audiences engaged with policymaking, designing, implementing, and advocating for this program and other similar programs worldwide.

This report begins with a summary of the program design and implementation over the three-year period, followed by the evaluation design, key findings on children's literacy outcomes, and concludes with recommendations for Room to Read and its partners.

Literacy Instruction Scale-up Program in Jodhpur

Room to Read's comprehensive literacy approach focuses on eight key components of foundational literacy, namely: oral language development, phonological awareness, phonics, fluency, vocabulary, comprehension, writing and Independent Reading Time (IRT). To support schools and teachers to implement this approach, the program included activities focused around four elements: 1) the provision of teaching and learning materials (TLMs); 2) capacity strengthening training for system-based actors; 3) the establishment of cluster-based libraries and classroom reading corners; and 4) on-site coaching and mentoring support for schools and teachers. Room to Read believes that this comprehensive literacy approach supported by activities around the four elements and a robust program implementation monitoring , acts as a catalyst in strengthening children's literacy skills and reading habits.

A summary of the implementation of different elements of the program is provided below. This summary focuses primarily on the program data for the School Year (SY) June 2023 – April 20224 and SY June 2024 – April 2025. However, in some places data from the SY 2022-23 i.e. first SY of implementation is also provided. Where data is available, the achievements of the program's intermediary outputs are also discussed. The results presented in this section pertain to all program schools.

Materials Distribution

The program provides TLMs for teachers and children to support Hindi language instruction for grades 1-3. Specifically, Teacher Guides and Student Workbooks were distributed. The distribution of materials happened in a phased manner. In SY 2023-24, the first batch of materials for grades 1 and 2 was delivered to program schools between October-November 2023. Then in SY 2024-25, Room to Read distributed materials covering grades 1-3 to program schools between November-December 2024. Table 2 below highlights the total number of TLMs distributed to program schools during both SYs.

Table 2. TLM Distribution to Program Schools

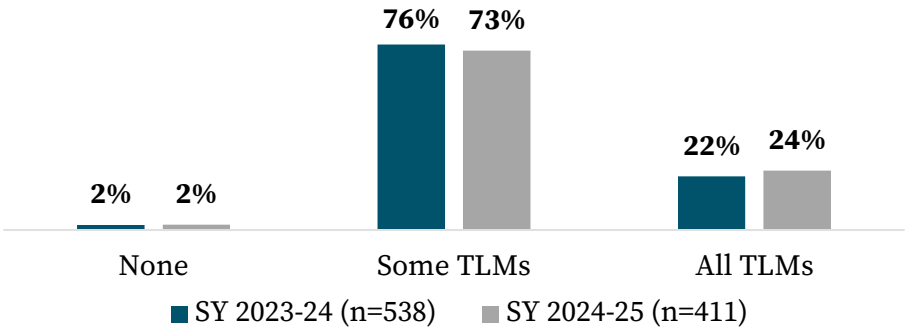
Indicators - Distribution of TLM	SY 2023-24	SY 2024-25
Room to Read Teacher Guides	2,641	2,595
Grade 1 Student Workbooks	13,155	10,780
Grade 2 Student Workbooks	22,987	17,239
Grade 3 Student Workbooks	n/a[3]	18,831

Source: program monitoring data

To assess teachers' uptake and use of Room to Read TLMs in program schools, Program Field Coordinators (PFCs) recorded their usage — none, some, or all — during classroom observations conducted during school visits. Among all observations conducted in both SY 2023 and SY 2024, PFCs reported moderate uptake of the TLMs. Figure 1 shows that approximately 22-24 percent observed teachers using all the materials properly, while an additional 73-76 percent used some of the TLMs.

[3] Implementation of the program in Grade 3 happened only in SY 2024-25.

Figure 1. Observed Proper Use of TLMs by Teacher: PFC Individual Monitoring Visits



Capacity Strengthening Training

The Literacy Program provided capacity strengthening training to Government Master Trainers (MTs) and grades 1–3 teachers on Hindi language instruction, reading activities during library periods, and cluster library management. Training was delivered through a cascaded model whereby Room to Read trained MTs, who then trained language teachers at program schools. Additionally, Room to Read trained Panchayat Elementary Education officers (PEEO) on how to deliver high-quality supportive supervision (coaching and mentoring support) to schools and teachers.

Government MTs and PEEOs were trained twice during the first and second year of implementation (SY 2022-23 and SY 2023-24), while teachers were trained annually for all three years of implementation. However, in SY 2023-24 and SY 2024-25, the teacher training was conducted as part of the Government's Foundational Literacy and Numeracy (FLN) training. Room to Read staff served as facilitators for sessions on foundational literacy and library topics during the FLN training in Jodhpur. Given the scale of the teachers being trained, the training was organized into 157 batches covering 22 blocks in Jodhpur. Table 3 below highlights the total number of system-based actors trained under the program and the Room to Read supported FLN Training.

Table 3. Participation in the Program's Capacity Strengthening Training

Indicators - Training	SY 2022-23	SY 2023-24	SY 2024-25
By Room to Read			
Government MTs	67	68	--
PEEOs	491	574	--
By Government MTs			
Teachers (grades 1-3)	2,306	2,039[4]	4,952[5]

[4] A total of 7,142 participants were trained in the Government's FLN training, including 2,039 grade 1-3 teachers from program schools.
[5] In total 7,000 teachers participated in the FLN training in Jodhpur including sessions delivered by Room to Read staff. Of these 7,000, approximately 4,952 were from program schools.

To assess the knowledge gained on foundational literacy topics, pre- and post-training assessments were conducted with participants. Table 4 shows the pre and post-test mean scores and percent change for each year and by participant type. Across all training workshops, mean scores increased from pre- to post-test, however the percent change varied from a modest 7.7 percent to 40.7 percent depending on the year and target participant. This indicates the cascaded model was moderately successful at increasing the knowledge of participants.

Table 4. Pre- and Post-training Knowledge Assessment Results by Participant Type and Year

Outputs - Training	Pre-test Mean Score	Post-test Mean Score	Percent Change
Government MTs			
SY 2022-23 (n=58)	13.0	14.0	7.7%
SY 2023-24 (n=64)	8.6	12.1	40.7%
PEEOs[6]			
SY 2023 (n=338)	9.0	12.0	33.3%
Teachers (Grades 1-3)			
SY 2022-23 (n=1,620)	7.0	8.2	17.1%
SY 2023-24 (n=2,039)	10.0	11.0	10.0%
SY 2024-25 (n=4,952)	12.9	14.9	15.5%

The pre- and post-test for Government MTs included 18 questions in SY 2022-23 and 15 questions in SY 2023-24. The pre and post-test for PEEOs and teachers included 15 questions each. Mean scores were calculated by dividing the total number of correct responses by the total number of questions in each assessment.

Cluster Libraries and Classroom Reading Corners

The program helped establish cluster-level libraries and classroom reading corners in Jodhpur to increase children's access to quality reading materials and to foster reading habits. A cluster library is set up in a school with sufficient space for a library room. Then, books are rotated monthly to classroom reading corners set up in other close-by program schools. Each cluster library received storybooks, and schools also received professional development training on how to effectively manage them.

In SY 2023-24, 488 cluster libraries were established under the program, with Room to Read providing 377 storybooks to each of them (See Table 5). In addition to storybooks, each school received two jute display units to showcase storybooks and one cloth bag to use for the book rotations between the school and cluster library. In the beginning of the SY 2024-25, Room to Read distributed an additional 600 storybooks of varying levels and genres across 649 clusters.

Table 5. Implementation of the Cluster Library Activities

Indicators – Cluster Libraries	SY 2023-24	SY 2024-25
Number of Cluster Libraries set-up	488	649
Number of storybooks distributed per cluster	377	600

[6] While training was conducted with PEEOs in SY 2022, pre- and post-training knowledge assessment data was not available for analysis.

Indicators – Cluster Libraries	SY 2023-24	SY 2024-25
Number of books distributed (approximate)[7]	181,714	389,400
Jute display units distributed	5,282	--
Book bags	2,641	--

To measure the uptake of library practices, the program tracked three key output indicators: 1) the percent of visits where schools reported that the book rotation had occurred in the last month; 2) the percent of visits where teachers were observed conducting reading activities during the library period; and 3) the average number of books checked-out per child per grade.[8] Figure 2 and 3 present the results for all three for both SY 2023-24 and SY 2024-25.

During school monitoring visits, PFC and PEEOs observed strong implementation fidelity of the monthly book rotation processes across both SYs. Overall, about 89 percent of PFCs reported that the book rotation occurred in the previous month, while slightly more than 90 percent of PEEOs did the same. There was moderate to high implementation fidelity of reading activities during the library period. PEEOs reported a higher proportion of teachers conducting them than PFCs, although PFCs reported an increase in implementation fidelity from 55 percent in SY 2023-24 to 71 percent in SY 2024-25. In contrast, PEEOs saw a slight decrease from 91 to 85 percent.

Figure 2. Adoption of Library Practices Observed during Monitoring Visits (% Observations)

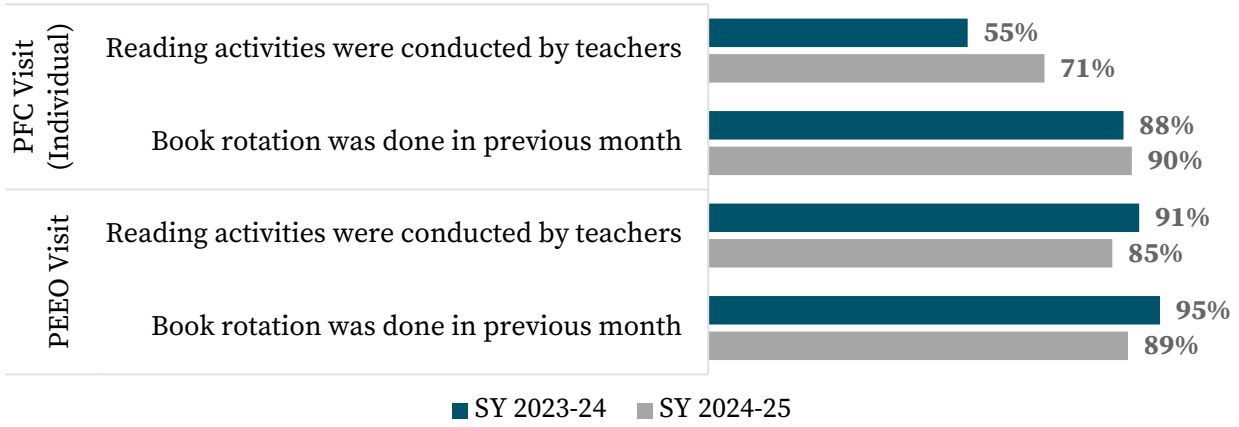


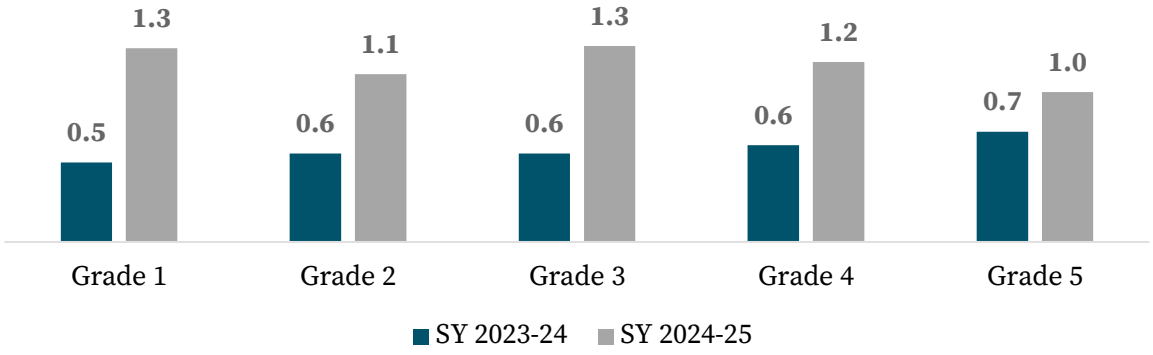
Figure 3 shows the average monthly storybook checkouts per child from cluster libraries and classroom reading corners, based on school monitoring visits. Between SY 2023-24 to SY 2024-25, the average per month checkout rate approximately doubled for Grades 1-4. This suggests

[7] The total number of books was calculated by multiplying the total number of books distributed per cluster by the total number of cluster libraries set-up. This provides a close approximation of the total number of books distributed, but there may be slight variations in some cluster libraries.

[8] Book check-out data was collected during school monitoring visits conducted by PEEOs and PFCs, and as such represents the proportion of visits where the behaviour was reported. It represents the proportion of program schools among those visited and may or may not be representative of all program schools.

an increase in the adoption of book check-out practices by schools and teachers, and improved reading habits among children.

Figure 3. Average Number of Storybooks Checked out per Child (per month)



Teacher Coaching and Mentoring Support

The Literacy Program aims to provide regular on-site teacher coaching and mentoring support to encourage schools and teachers to put new practices into use. Given the large-scale nature of the program, this support was provided by both Room to Read PFCs and local government officials (PEEOs) through school visits. PFCs were expected to conduct eight joint visits and four individual visits per month, while the target for PEEOs was for 75 percent of them to conduct at least two visits per month. The program monitoring forms for conducting coaching visits to schools were rolled out in October 2023. As such, school visit monitoring data was expected for five months during SY 2023-24 and ten months in SY 2024-25, considering holidays. Table 6 presents the total number of visits achieved against the target based on data submitted by PFCs and PEEOs.

Room to Read PFCs exceeded their target for the total number of individual coaching visits in both SY 2023-24 and SY 2024-25. However, the total number of joint visits achieved with PEEOs was approximately 51 and 64 percent of the target in SY 2023-24 and SY 2024-25, respectively[9]. Additionally, PEEOs achieved about 16 and 23 percent of their targets for individual visits across both years as well. This suggests low implementation fidelity of this aspect of the program.

Table 6. Frequency of Coaching and Mentoring Support

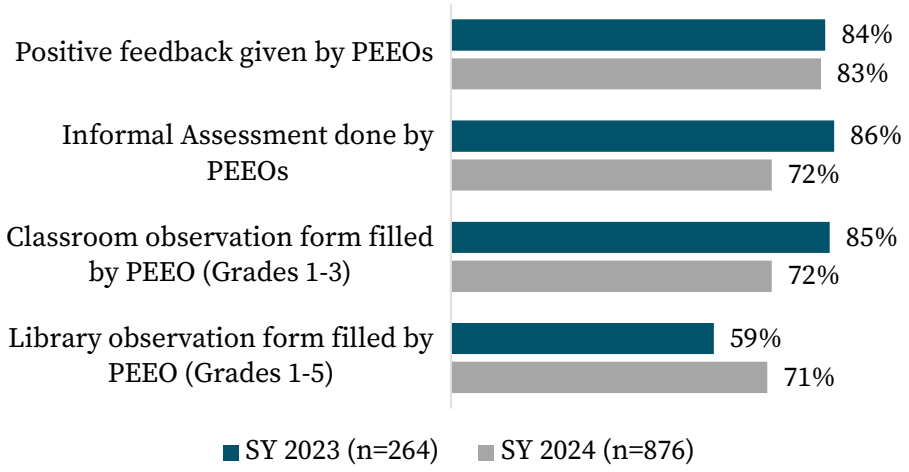
Indicators – Coaching Support	SY 2023-24		SY 2024-25	
	Target	*Achieved*	*Target*	*Achieved*
Room to Read PFCs – Individual visits	240	507	680	757
Room to Read PFCs – Joint visits	680	349	1,360	876
PEEOs – individual visits	4,500	709	9,000	2,051

[9] The percentage of the target achieved was calculated by dividing the actual number of visits by the target number of visits.

(i) For PFCs – the target for individual and joint visits was 4 and 8 per PFC per month, for a total of 5 months in SY 2023 and 10 months in SY 2024. This resulted in the calculation of 17 PFCs × 4 individual visits per month × 5 or 10 months. For joint visits, it was 17 PFCs × 8 joint visits × 5 or 10 months. (ii) For PEEOs – the target for individual visits was for 75 percent of the 600 PEEOs to complete 2 per month. This resulted in the calculation of (600 × .75) × 2 × 5 or 10 months depending on the SY.

The program aimed to strengthen PEEOs' capacity to provide quality coaching support to schools and teachers. To assess this, Room to Read PFCs observed PEEOs during joint visits, recording their use of Room to Read coaching practices as an indicator of support quality. Figure 4 shows the proportion of PEEOs implementing Room to Read's coaching practices during joint visits, with consistently high fidelity across both SYs. For instance, 83–84 percent provided positive feedback to teachers. However, implementation of informal assessments with a sample of children and classroom observation forms dropped from 85 to 72 percent, while the completion of library observation forms increased.

Figure 4. Select Indicators – Quality of Coaching Support provided by PEEOs (% of Joint Visits)



Evaluation Design

Room to Read initiated a two-year impact evaluation following a repeated cross-sectional design to measure the results of the program. In May 2023, a pre-intervention baseline assessed the early grade literacy skills of end-of-Grade 3 children in program schools not yet benefiting from the intervention in Jodhpur district (defined as program group) and children in a sample of schools not included in the intervention in the nearby Barmer district (defined as comparison group). Two academic years later, a post-intervention endline has been conducted with a new cohort of end-of-Grade 3 children in the same program and comparison schools where the 2023 baseline was conducted. At endline, children in program schools had received two academic years of support under the program. The evaluation had the following objectives:

- To evaluate the impact of Room to Read's literacy program on early grade children's reading and writing skills in contrast to children in comparison schools.
- To assess whether the program's implementation supports the acquisition of early literacy skills at a pace that ensures children read fluently with comprehension and demonstrate grade-appropriate writing skills by the end of Grade 3.
- To explore what school, teacher and child-level contextual factors were influencing children's literacy skills in both program and comparison schools at endline.

Data Collection Instruments

To achieve these evaluative objectives, quantitative information was collected using two structured data collection tools: 1) a school and teacher background tool; and 2) Grade-3 levelled Early Grade Literacy Skills Assessment (EGLSA) with relevant child background and home literacy environment questions.

The **school and teacher background tool** gathered information on the school environment, enrolment and attendance in the target grade, classroom amenities and features of the school library and classroom reading corner, if present, among other factors. Additionally, it gathered information about the Grade 3 teacher's years of experience, education levels, home language and other relevant demographic data at endline.

The **Grade 3 EGLSA** assessed children's literacy skills via five tasks: 1) letter name identification; 2) nonword reading; 3) oral passage reading; 4) reading comprehension; and 5) sentence dictation. Table 7 below presents information about each task, and the results of tool reliability analyses are included in Annex A.

Table 7. EGLSA Tasks

Assessment Task	Description
Letter Name Identification	Children receive a total of 100 different letters and have 60 seconds to read as many letters as possible. Child's letter naming fluency is calculated and reported as the number of correct letters per minute (clpm).
Nonword Reading	Children have 60 seconds to read 50 nonwords. Nonword reading fluency is calculated as the number of correct nonwords read per minute (cnwpm).

Assessment Task	Description
Oral Passage Reading	Children have 180 seconds to read a 140-word passage of Grade-3-level of difficulty. The data is analysed to calculate oral reading fluency which is reported as the number of correct words read per minute (cwpm)
Reading Comprehension	Measures children's ability to answer up to six questions based on the contents of the oral reading passage. The number of questions asked depends on the number of words the child read in the passage. Questions are either fact-based or inferential in nature.
Sentence Dictation	Children write a six-word long sentence that the assessor read orally to them. The sentence is scored on a 10-point scale based on correct spelling, including a space between each word, punctuation, and sequence of words.

During EGLSA, children were asked a series of questions to understand their background characteristics and the home literacy environment. For example, each child was asked about their age, prior pre-school participation, school attendance in the week preceding the assessment, home language and household standard of living (measured through the presence of various amenities at home). To measure the home literacy environment, children were asked about the literacy status of the people who live with them, the type of support they provide them, their access to reading materials at home, and about their own reading habits for fun.

Sampling

The impact evaluation included a sample of 1,600 Grade 3 children drawn from 160 government-run primary schools, divided evenly from Jodhpur (program schools - 80) and Barmer (the comparison district – 80). Barmer was selected as comparison district based on the similarity of its characteristics relative to the program district. The sample size required was determined based on the following criteria: 80% power, 5% margin of error and 0.35 minimum detectable effect size.

Before the baseline, a three-stage sampling approach was employed to randomly select the sampled schools for inclusion in the two-year impact evaluation to ensure consistency and comparability across both time points.

1. **First stage:** PEEOs were selected from each block of both program and comparison districts. The targeted 80 PEEOs (40 from each district) were drawn from across all blocks in proportion to the number of PEEOs within each block.
2. **Second stage:** Schools with a minimum enrolment of at least 15 children in Grade 3 were included in the sampling frame and organized by PEEO. Then, two schools under each PEEO were randomly selected from the sampling frame. Thus, a total of 80 schools each were sampled from both program and comparison districts.

3. **Third stage:** 10 end-of-Grade 3 children who were present on the day of the assessment were randomly selected from each sample school for the EGLS assessment at both 2023 baseline and 2025 endline.[10]

Table 8 presents the target and actual achieved sample size for both the baseline and endline assessments.

Table 8. Target and Actual Sample Size – Baseline and Endline Assessments

Type	Target Sample		Actual Sample - Baseline (2023)		Actual Sample - Endline ()2025)	
	Schools	Children	School	Children	School	Children
Program	80	800	80	820	80	702
Comparison	80	800	80	770	79[11]	724
Total	160	1600	160	1590	159	1426

Assessor Training and Data Collection

Outline India, an independent survey firm, was hired by Room to Read, to support the recruitment and hiring of supervisors and assessors, training and data collection during the 2025 endline. A summary of the training and data collection conducted is outlined below.

1. **Assessor training**: 3-days of desk-based training and 1-day of school-based practice was organized for assessors and supervisors from March 2-5, 2025. The training covered familiarization with both the school background and child assessment tools; protocols for conducting data collection with schools and children; usage of the SurveyCTO mobile data collection app for both tools; data quality assurance procedures; random selection processes for selecting sampled children; logistical planning. An additional day of training was added after field practice to provide assessors with more time to practice administering the EGLSA tasks, resulting in a total of 5-days of training. As part of the training, four inter-rater reliability assessments were conducted to measure the percent agreement between assessors on the various tasks included in the EGLSA tool. The results of these assessments can be found in Annex B. These assessments help to ensure that all assessors administer the EGLSA tool in a reliable and consistent manner and that they meet global standards[12] before being deployed for data collection.

2. **Endline data collection:** Endline data collection took place between March 7 – 26, 2025, which coincided with the end of academic year in Rajasthan. A total of 13 assessors and 8 supervisors were deployed to conduct data collection at sampled schools. At each school, assessors conducted the EGLSA with 10 end-of-Grade 3 children. The school and teacher

[10] Children with visible physical, sensory and/or significant cognitive disabilities were excluded from the sample selection.
[11] At endline, one comparison school did not have any child enrolled in Grade 3. It was excluded from endline data collection.
[12] For EGLSA administration, the gold standard is set at 95% agreement across all tasks. All assessors and supervisors must score 95% or more on at least one IRR assessment to be allowed to conduct data collection.

background information were collected from the headteachers and Grade 3 teachers. Supervisors conducted both simultaneous and re-test (within three days of the original assessment) data quality assessments (DQA) for each assessor conducting the EGLSA. DQA assessments are used to measure data reliability throughout data collection, by examining the percent agreement between the assessors marking of each item against that of the supervisor. Additional details on the DQA procedures adhered to during endline data collection and full DQA results for each assessor can be found in Annex B.

Data Analysis

For this evaluation, descriptive analyses – such as frequencies, proportions, averages, and variations in distributions (e.g., standard deviation) – were calculated for several indicators. Comparative analyses between program and comparison schools were performed for all indicators using various statistical techniques. For instance, inferential tests such as chi-square and t-tests were conducted to determine the statistical significance of differences between program and comparison groups across school-, teacher-, and child-level features at baseline and endline.

Ordinary Least Squares (OLS) regression on the scores in each EGLS assessment task was performed for the same purpose. Logistic regression was conducted on the incidence of non-performance in each task (defined as a zero score) and on achievement in reading benchmarks. Each regression model, whether OLS or logistic, included a range of school-, teacher-, and child-level background features as covariates. Additionally, the regression models accounted for intra-school (i.e., cluster level) correlations in the dependent variable. Where necessary, the regression analyses were conducted separately for different groups (e.g., girls, boys, program and comparison schools). Detailed results from all the regression analyses are available in Annex C.

Effect size calculations for each of the five reading assessment tasks was also computed to quantify the magnitude of the impact of the Literacy Instruction Scale-up Program in Jodhpur after two years of implementation. Effect size calculations compute the mean difference in gains between program and comparison schools divided by the pooled standard deviation of scores in program and comparison schools.

Features of Sampled Schools, Teachers and Children at Endline

Schools

Across most background characteristics, program and comparison schools were similar at endline. For example, almost all were in rural areas, and average Grade 3 enrolment was 16-17 children. Comparison schools had a slightly higher average pupil-teacher ratio (5.2 pupils per teacher) as compared to program schools (4.4 pupils per teacher). However, the difference was not statistically significant. Attendance was also low across all schools, with an average of between 64-67 percent of enrolled children present on the day of the assessment.

Local government monitoring support was similar across both school types, with schools on average receiving three visits in the three months preceding data collection. All schools reported having a functioning parent-teacher committee (PTC), and across program and comparison schools they held a similar number of meetings in the last six months (about five,

on average). Lastly, program and comparison schools on average had a similar number of available facilities within the school environment. On average, program and comparison schools each scored slightly higher than eight on the 10-point school environment score scale[13]. Moreover, a similar proportion had school libraries (81 and 75 percent in program and comparison schools, respectively).

In contrast, there were some However, statistically significant difference between program and comparison school was found at endline on the following characteristics:

- There was a larger share of multigrade Grade 3 classrooms in comparison schools (61 percent) than in program schools (45 percent) ($p < 0.05$).
- PTC in comparison schools on average held more outreach events in the last six months (2.9) in contrast to program schools (2.3) ($p < 0.05$).
- Despite having similar school environment scores, program and comparison schools had statistically significant differences in the availability of various amenities within the Grade 3 classroom. On average, program schools scored 7.0 on the 8-point classroom environment scale, while comparison schools scored 5.7 on average[14] (p <0.001).

Table 9 presents details on the background characteristics of program and comparison schools at endline.

Table 9. Background Characteristics of Program and Comparison Schools at Endline

Indicator	Program	Comparison
	% Share/ Mean (SD)	% Share/ Mean (SD)
Number of schools	***80***	***79***
Location (rural)	94%	95%
Enrolment in Grade 3	15.5 (8.4)	17.0 (8.6)
Gender parity ratio in Grade 3	1.4 (0.9)	1.3 (1.7)
Pupil teacher ratio in Grade 3	4.4 (2.4)	5.2 (3.0)
Grade 3 Attendance on the day of the assessment (% share)	67.0 (19.9)	64.2 (23.1)
Type of classroom for Grade 3: multigrade*	**45%**	**61%**
School has a library	81%	75%
Reading corner in Grade 3 classroom***	**95%**	**1%**
Number of school visits by local government (in the last 3 months)	3.3 (1.4)	2.9 (1.5)

[13] The school environment score included the following facilities: power/electricity, potable drinking water, separate toilet of boys and girl, clean toilet, fence or boundary around the school, playground, computer/IT equipment available for student use, school library, school feeding program and a paved road leading to the school.

[14] The classroom environment score included the following classroom amenities: blackboard/ whiteboard, visual aids on the wall, teachers guide, student textbooks/workbooks for most of the students, adequate seating for all children, intact roof without any visible signs of damage, intact walls without any visible structural damage and a reading corner setup.

Indicator	Program	Comparison
	% Share/ Mean (SD)	% Share/ Mean (SD)
Number of PTC meetings (in the last 6 months)	4.8 (2.0)	4.9 (1.8)
Number of outreach events organized by PTC (in the last 6 months)*	**2.3 (1.5)**	**2.9 (1.7)**
Received support from any agency (other than RTR)	14%	6%
School environment: Score (10-point scale)	8.5 (1.0)	8.3 (1.1)
Classroom environment: Score (8-point scale)***	**7.0 (1.2)**	**5.7 (1.4)**
Reading corner environment: Score (6-point scale)	5.6 (0.8)	3.0 (-)

*The difference between program and comparison groups was statistically significant, either at * $p < 0.05$ or ** $p < 0.01$ or *** $p < 0.001$.*

Grade 3 Teachers

Across most background characteristics of Grade 3 teachers, program and comparison schools were similar (See Table 10). For example, Grade 3 teachers on average were in their late thirties, and slightly more than 40 percent reported speaking Hindi at home. Almost all had completed graduate level education or above, and they had an average of about 9-10 years of experience teaching. There was a slightly higher proportion of male Grade 3 teachers in comparison schools (54 percent), as compared to 50 percent in program schools. However, these differences were not statistically significant.

The only statistically significant difference was found in the proportion of Grade 3 teachers with prior experience in teaching outside their current school. While 44 percent of program teachers reported "yes," only 15 percent of Grade 3 teachers in comparison schools did (p <0.001). Given the overall years of experience were similar across both types of schools, this suggests that Grade 3 teachers in program schools have changed schools more often.

Table 10. Background characteristics of Grade 3 Teachers

Indicator	Program	Comparison
	% Share/ Mean (SD)	% Share/ Mean (SD)
Number of schools	***80***	***79***
Grade 3 Teacher Gender		
Male	50%	54%
Female	50%	46%
Age (years)	37.4 (10.2)	37.4 (10.2)
Teacher speaks Hindi at home	44%	43%
Teacher's education: Graduate or above	95%	91%
Had prior experience in teaching***	**44%**	**15%**
Teaching in the current school for 5+ years	66%	70%
Teaching experience (years)	10.0 (8.4)	8.5 (7.1)

Indicator	Program	Comparison
	% Share/ Mean (SD)	% Share/ Mean (SD)
Employment contract with the school: Regular/ Permanent	98%	96%

*The difference between program and comparison groups was statistically significant, either at * $p < 0.05$ or ** $p < 0.01$ or *** $p < 0.001$.*

Grade 3 Children

There were similarities in some of the background characteristics of Grade 3 children among program and comparison schools. For example, there were slightly more girls assessed than boys, which likely reflects gender differences in enrolment. Additionally, slightly more than 70 percent across all schools reported living at home with both parents and that they had four meals in the last 24 hours (breakfast, lunch, dinner and snacks). Children also had similar household sizes, with an average of six members living at home with them. Table 11 presents the full results.

In contrast, there were notable statistically significant differences in the background characteristics of sampled Grade 3 children. Most of these differences were also observed at baseline with the original cohort of Grade 3 children that were assessed. Observed differences include:

- On average Grade 3 children in program schools were slightly older (9.0 versus 8.8 years old, respectively) (p <0.01). At baseline, Grade 3 children in that cohort also averaged 9.0 and 8.8 years, respectively for program and comparison schools.
- Almost double the proportion of Grade 3 children in program schools (24 percent) reportedly speak Hindi at home in contrast to 12 percent from children in comparison schools ($p < 0.001$). Similar differences were also found at baseline (37 percent and 17 percent, respectively for program and comparison schools).
- Similar to the Baseline findings in 2023, a larger proportion of sampled children in program schools (44 percent) reported attending pre-school in contrast to 29 percent among children in comparison schools ($p < 0.001$).
- Double the proportion of children in program schools (seven percent) were repeating Grade 3, while only three percent of children in comparison schools were ($p < 0.01$).
- Higher absenteeism was reported among children in program schools in the week preceding the assessment (32 versus 21 percent for program and comparison schools, respectively) ($p < 0.001$)
- Lastly, children in program schools on average reported a higher number of amenities in their homes than children in comparison schools, resulting in statistically significant differences in their household standard of living scores (7.5 versus 6.7) ($p < 0.001$).

Table 11. Background Characteristics of Grade 3 Children at Endline

Indicator	Program	Comparison
	% Share/ Mean (SD)	% Share/ Mean (SD)
Number of Grade 3 children assessed	***702***	***724***
Age (years)**	**9.0 (1.0)**	**8.8 (0.9)**
Gender		
Boy	49%	47%
Girl	51%	53%
Language child speaks at home***		
Hindi	24%	12%
Marwari	74%	87%
Other	1%	1%
Attend pre-school***	**44%**	**29%**
Repeating the current grade**	**7%**	**3%**
Lives with both parents	74%	72%
Had 4 meals in the last 24 hours	72%	73%
Absent from the school at least for a day (in the last week)***	**32%**	**21%**
Household size	6.4 (1.7)	6.4 (1.7)
Household standard of living: Score (on 10-point scale)***	**7.5 (1.8)**	**6.7 (2.0)**

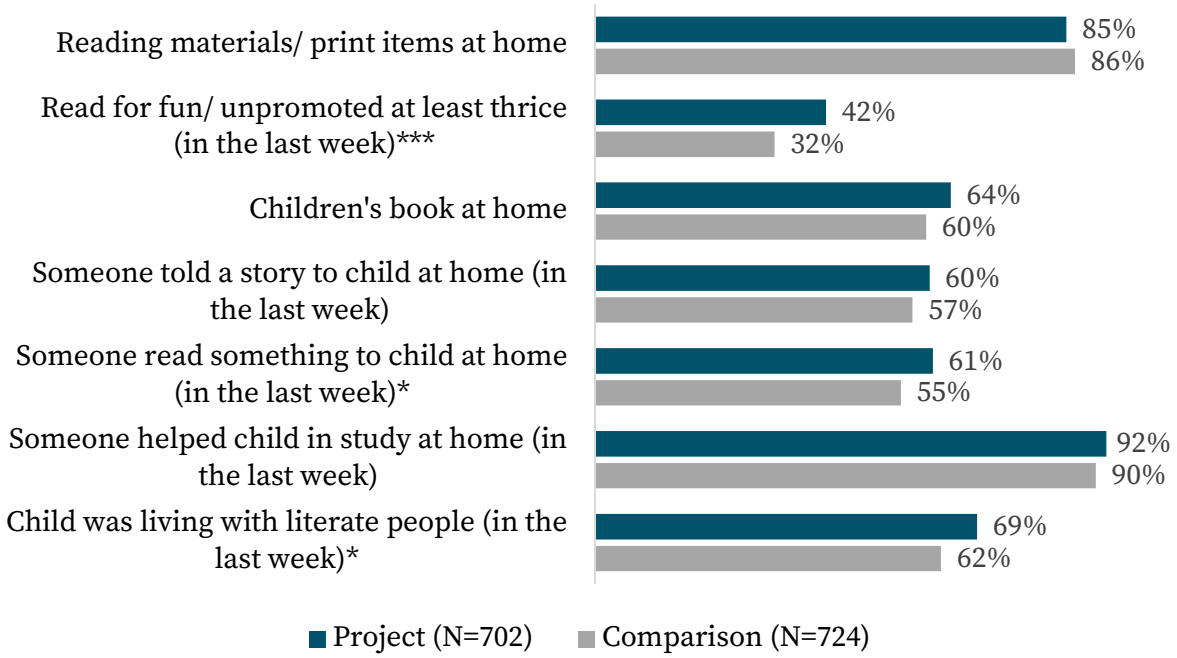
*The difference between program and comparison groups was statistically significant, either at * $p < 0.05$ or ** $p < 0.01$ or *** $p < 0.001$.*

Across most features of children's home literacy environment, children from program and comparison schools were similar. For example, about 85-86 percent reported having access to reading materials or print items at home, and about 60 percent also reported specifically having children's books at home. Family engagement was also similar, with about 60 percent reporting that someone had told them a story in the last week, and about 90 percent had someone help them to study as well. Figure 5 presents the full home literacy environment results.

There were a few notable differences as well. For example, a higher proportion of Grade 3 children in program schools reported that they lived with literate family members (69 versus 62 percent for program and comparison schools, respectively)($p < 0.05$). Moreover, while 61 percent of children from program schools reported someone had read to them in the last week, only 55 percent from comparison schools did ($p < 0.05$). Lastly, children from program schools reported stronger reading habits at home, with 42 percent reporting that they read for fun at least three times in the last week compared to 32 percent among comparison schools ($p <0.001$).

It is important to note that for the latter two items, the differences could reflect aspects of the program intervention. For example, children in program schools may have stronger reading habits due to the ability to check out books from the classroom reading corner set-up as part of the intervention.

Figure 5. Features of Children's Home Literacy Environment



Evaluation Results – Children's Early Grade Literacy Skills

The impact evaluation results of the EGLSA conducted with end-of-Grade 3 children during the 2023 baseline and 2025 endline assessments are presented here.

Mean Scores

Literacy skill scores of the children in program schools have significantly improved on all assessment tasks as compared to their peers in comparison schools from baseline to endline. Moreover, children from program schools significantly outperformed children from comparison schools on all assessment tasks at endline. For letter naming fluency, children from program schools could name an average of 55.2 correct letters per minute, while children from comparison schools managed only 43.4 letters per minute ($p < 0.001$). Similarly, in nonword reading fluency, children from program schools nearly doubled their scores (10.8 to 19.6 cnwpm), outperforming comparison school children (14.4 to 16.0 cnwpm) ($p < 0.001$).

Children in program schools made a large statistically significant improvement in oral reading fluency from baseline to endline. At baseline, they read an average of 18.6 cwpm, while at endline this improved to an average of 43.1 cwpm, in contrast, children in comparison schools improved from 25.1 to 31.2 over the same period ($p < 0.001$). In reading comprehension, children from program schools correctly answered about two questions out of six (1.7), slightly more than the comparison group (1.6) ($p < 0.01$). This was despite the fact the mean score was lower at baseline for program schools (0.6) than comparison schools (1.1).

Lastly, children from program schools significantly improved their writing skills, as measured by the sentence dictation task, from baseline to endline. While children from program schools improved from an average of 3.4 to 5.2 (out of 10), comparison children moved from 4.2 to 4.6 ($p < 0.05$). **Furthermore, across all five assessment tasks, children in program schools made two-year gains in their reading and writing skills that were approximately 1.1 to 1.4 times greater than those seen in comparison schools.**

Table 12. Early Grade Literacy Skill Assessment Results at Baseline and Endline

Score (unit)	Group	Baseline (2023)		Endline (2025)		Est. diff. b/w the groups [Std Error]
		Count	Mean (SD)	Count	Mean (SD)	
Letter naming fluency (clpm)	Program	820	33.2 (23.7)	695	55.2 (25.9)	20.00*** [2.93]
	Comparison	770	42.2 (27.9)	718	43.4 (25.6)	
Nonword reading fluency (cnwpm)	Program	820	10.8 (11.4)	695	19.6 (12.4)	6.55*** [1.48]
	Comparison	770	14.4 (14.3)	717	16.0 (11.9)	
Oral reading fluency (cwpm)	Program	820	18.6 (24.1)	695	43.1 (30.1)	16.38*** [3.26]
	Comparison	770	25.1 (29.1)	717	31.2 (26.9)	
Reading comprehension: Score (out of 6)	Program	820	0.6 (1.1)	702	1.7 (1.6)	0.49** [0.19]
	Comparison	770	1.1 (1.5)	724	1.6 (1.6)	

Score (unit)	Group	Baseline (2023)		Endline (2025)		Est. diff. b/w the groups [Std Error]
		Count	Mean (SD)	Count	Mean (SD)	
Sentence dictation: Score (out of 10)	Program	820	3.4 (3.4)	702	5.2 (3.3)	1.03* [0.41]
	Comparison	770	4.2 (3.3)	724	4.6 (3.5)	

*Legend of statistical significance of differences between project and comparison schools: * p < 0.05 ** p < 0.01 *** p < 0.001. SD – Standard Deviation.*

Zero Scores

Zero scores represent the percentage of children who did not read a single item or answer a single question correctly on an assessment task. Zero scores help identify non-readers or the most struggling children. Table 13 presents the proportion of children that scored zero on each assessment task for both the baseline and endline assessments, along with the estimated differences between groups and odds ratios after controlling for various school, teacher and child-level factors.

Table 13. Zero Scores in Literacy Skills Assessment Tasks

Assessment Task	Group	Baseline (2023)		Endline (2025)		Est. Diff. [Std. Error]	Odds Ratio
		N	Zero score % Share	N	Zero score % Share		
Letter naming	Program	820	10%	702	4%	-1.18** [0.41]	0.31
	Comparison	770	8%	724	9%		
Nonword reading	Program	820	25%	702	8%	-1.05** [0.32]	0.35
	Comparison	770	22%	724	17%		
Oral passage reading	Program	820	35%	702	10%	-1.38*** [0.36]	0.25
	Comparison	770	27%	724	23%		
Reading comprehension	Program	820	66%	702	30%	-0.85** [0.26]	0.43
	Comparison	770	53%	724	38%		
Sentence dictation	Program	820	31%	702	18%	-0.89** [0.30]	0.41
	Comparison	770	26%	724	30%		

*Legend of statistical significance of differences between project and comparison schools: * p < 0.05 ** p < 0.01 *** p < 0.001.*

There was a large and statistically significant reduction in the prevalence of zero scores for program schools across all tasks as compared to comparison schools. For example, the proportion of children scoring zero declined from 10 percent at baseline to 4 percent at endline in program schools, while comparison schools increased from 8 to 9 percent over the same period ($p < 0.01$). In nonword reading, program schools reduce the percentage of zero scores from 25 percent at baseline to 8 percent at endline. Comparison schools reduced theirs by five percentage points (22 percent at baseline to 17 percent at endline). The differences between

program and comparison schools in the changes in likelihood of scoring zero from baseline to endline on this task were statistically significant ($p < 0.01$)

For oral passage reading, the prevalence of zero scores in program schools reduced by 25 percentage points (from 35 to 10 percent), while in comparison schools it reduced by four percentage points from 27 to 23 percent ($p < 0.001$). In reading comprehension, at baseline about two-thirds of children scored zero in program schools. In contrast, at endline, approximately 30 percent did. For comparison schools, the reduction was not as large, from 53 to 38 percent from baseline to endline ($p < 0.01$). Lastly, in sentence dictation, the proportion of children scoring zero reduced from 31 to 18 percent, while in comparison schools it increased from 26 to 30 percent over the same period ($p < 0.01$).

Logistic regression analyses were used to estimate the effect of a one-unit increase in the interaction between the type of school (program and comparison) and assessment wave (baseline and endline) on the likelihood of scoring zero on each assessment task (producing odds ratios). For all assessment tasks, the odds ratios were statistically significant at $p <0.01$ or $p <0.001$. Specific results include:

- **Letter naming:** a 1-unit change resulted in a 69 percent reduction in the odds of scoring zero for program schools as compared to the comparison group from baseline to endline ($p < 0.01$).
- **Nonword reading:** a 1-unit change indicated a 65 percent reduction in odds of scoring zero ($p < 0.01$).
- **Oral reading fluency:** the strongest effect was found for this assessment task, where a 1-unit increase resulted in a 75 percent reduction in the odds of scoring zero ($p < 0.001$).
- **Reading comprehension**: a 1-unit increase resulted in a 57 percent reduction in odds of scoring zero for program schools as compared to comparison schools from baseline to endline ($p < 0.01$).
- **Sentence dictation**: a 1-unit increase reflected a 59 percent reduction in odds of scoring zero for program schools as compared to the comparison group from baseline to endline ($p < 0.01$)

Literacy Benchmarks

According to the Foundational Learning Study (FLS, 2022[15]), the oral reading fluency (ORF) benchmark for Grade 3 children in Hindi is **at least 35 correct words per minute (cwpm)**. Room to Read, however, uses its own benchmark of **45 or more cwpm** across the languages in which it implements its Literacy Program globally. Recently, many countries have developed their own language-specific ORF benchmarks. When such benchmarks are available, Room to Read reports achievement using both the global and local standards. Additionally, Room to Read's reading comprehension benchmark is set at 80 percent or correctly answering about five

[15] National Council of Educational Research and Training. (2022). *National report on benchmarking for oral reading fluency with comprehension and numeracy*. Ministry of Education, Government of India. https://nipunbharat.education.gov.in/fls/file/Benchmarking_for_ORF_and_Numeracy.pdf

questions out of the six included. Table 14 presents the proportion of children in program and comparison schools that met each of these benchmarks at baseline and endline.

Children from program schools made statistically significant gains in achieving both Room to Read and government benchmarks for oral reading fluency and reading comprehension. At baseline, 23 percent of children in program schools achieved the ORF benchmark of 35+ cwpm, and by endline, it had increased to 62 percent (39 percentage point increase). For comparison schools, the proportion of children achieving this benchmark increased from 32 to 44 percent (12 percentage point increase). For the ORF benchmark set at 45+ cwpm, the proportion of children achieving it increased from 15 to 45 percent, representing a 30-percentage point increase. In contrast, it rose 21 to 27 percent in the comparison group, reflecting a 6-percentage point increase. Lastly, in reading comprehension, the proportion of children scoring 80 percent or higher increased from one to five percent in program schools, while the comparison group showed a marginal change from five to six percent.

Table 14. Achievement of Oral Reading Fluency and Reading Comprehension Benchmarks

Benchmarks	Group	Baseline (2023)		Endline (2025)		Est. Diff. [Std. Error]	Odds Ratio
		Count	% Share	Count	% Share		
Oral reading fluency: 35+ cwpm	Project	820	23%	702	62%	1.28*** [0.26]	3.60
	Comparison	770	32%	724	44%		
Oral reading fluency: 45+ cwpm	Project	820	15%	702	45%	1.17*** [0.29]	3.22
	Comparison	770	21%	724	27%		
Reading comprehension: 80%+ score	Project	820	1%	702	5%	1.37* [0.64]	3.94
	Comparison	770	5%	724	6%		

*Legend of statistical significance of differences between project and comparison schools: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.*

The differences between groups were statistically significant at $p < 0.001$ for both ORF benchmarks and $p < 0.05$ for the reading comprehension benchmark. The most pronounced improvement was observed for ORF benchmark of 35+ cwpm, where children in program schools were 3.6 times more likely than those in comparison schools to reach this benchmark from baseline to endline (odds ratio of 3.6). This was followed by gains in the ORF benchmark of 45+ cwpm, where children in program schools were 3.2 times more likely to achieve it (odds ratio of 3.2). For the reading comprehension benchmark (80%+ score), children in program schools were nearly four times more likely to achieve this benchmark as compared to their peers in comparison schools (odds ratio of 3.94).

Results Disaggregated by Gender

Girls in program schools exhibited significantly greater increases in literacy skills than their peers in comparison schools across all five assessment tasks. As shown in Table 15, the average letter naming fluency score for program girls increased from 34.7 clpm at baseline to

55.1 clpm at endline, while girls from comparison schools showed minimal change (40.2 to 41.3 clpm). The differences between groups from baseline to endline were statistically significant ($p < 0.001$). Similarly, in nonword reading fluency, girls in program schools nearly doubled their scores (10.7 to 19.4 cnwpm), outperforming comparison girls (13.4 to 15.2 cnwpm) ($p < 0.001$).

Girls in comparison schools increased their average oral reading fluency score from 23.4 to 29.1 cwpm, while girls in program schools increased theirs from 18.4 to 42.0 cwpm over the same period ($p < 0.001$). However, program girls' reading comprehension score increased from 0.6 to 1.6 from baseline to endline, which was only marginally higher than that of comparison girls (0.9 to 1.4). Nonetheless, the differences were statistically significant ($p < 0.05$). Lastly, girls in program schools demonstrated a statistically significant gain ($p < 0.05$) in sentence dictation, as they increased their score from 3.5 to 5.3, whereas girls from comparison schools increased their average scores from 3.8 to 4.3 over the same period. **Furthermore, across all five assessment subtasks, girls in project schools made two-year gains in their reading and writing skills that were approximately 1.1 to 1.4 times greater than those seen by girls in comparison schools.**

Table 15. Early Grade Literacy Skills Assessment Results for Girls

Assessment Task	Group	Baseline (2023)		Endline (2025)		Est. Diff. [Std. Error]
		Count	Mean (SD)	Count	Mean (SD)	
Letter naming fluency (clpm)	Project	387	34.7 (23.7)	355	55.1 (26.2)	18.21*** [3.68]
	Comparison	354	40.2 (28.1)	380	41.3 (25.6)	
Nonword reading fluency (cnwpm)	Project	387	10.7 (11.3)	355	19.4 (12.8)	6.84*** [1.81]
	Comparison	354	13.4 (14.5)	379	15.2 (11.7)	
Oral reading fluency (cwpm)	Project	387	18.4 (24.3)	354	42.0 (30.9)	16.69*** [4.15]
	Comparison	354	23.4 (30.9)	379	29.1 (26.8)	
Reading comprehension: Score (out of 6)	Project	387	0.6 (1.0)	360	1.6 (1.5)	0.45* [0.21]
	Comparison	354	0.9 (1.3)	382	1.4 (1.5)	
Sentence dictation: Score (out of 10)	Project	387	3.5 (3.1)	360	5.3 (3.4)	1.00* [0.49]
	Comparison	354	3.8 (3.3)	382	4.3 (3.5)	

*Legend of statistical significance of differences between project and comparison schools: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. SD – Standard Deviation.*

Table 16 indicates that program boys also demonstrated significantly greater increase in literacy skills as compared to boys in comparison schools across all five assessment tasks. The average letter naming fluency score for program boys increased from 31.9 at baseline to 55.4 clpm at the endline, while for comparison boys it improved only slightly from 43.8 to 45.7 clpm ($p < .001$). For nonword reading, program boys almost doubled their mean score from 11.0 to 19.8 cnwpm, whereas comparison boys experienced a smaller increase from 15.2 to 17.0 cnwpm. These differences were statistically significant at $p < 0.001$. Program boys also experienced significant improvements in oral reading fluency, increasing the number of correct words read per minute from 18.7 to 44.2 from baseline to endline. In contrast, the average ORF score of the boys in comparison schools improved from 26.5 to 33.5 cwpm. The

differences between groups form baseline to endline were also statistically significant for this assessment task ($p < 0.001$).

In reading comprehension, on average, program boys were able to answer an additional reading comprehension question correctly out of the six. Boys in comparison schools scored similarly, increasing from an average of 1.2 to 1.8 correct answers over the same period ($p < 0.05$). Similar to program girls, program boys increased their average sentence dictation score from 3.4 to 5.2, while for comparison boys it increased marginally from 4.5 to 5.0 ($p < 0.05$). **Furthermore, across all five assessment tasks, boys in program schools made two-year gains in their reading and writing skills that were approximately 1.1 to 1.3 times greater than those seen by boys in comparison schools.**

Table 16. Early Grade Literacy Skills Assessment Results for Boys

Assessment Task	Group	Baseline (2023)		Endline (2025)		Est. Diff. [Std. Error]
		Count	Mean (SD)	Count	Mean (SD)	
Letter naming fluency (clpm)	Project	433	31.9 (23.7)	340	55.4 (25.6)	20.66*** [3.15]
	Comparison	416	43.8 (27.6)	338	45.7 (25.4)	
Nonword reading fluency (cnwpm)	Project	433	11.0 (11.4)	340	19.8 (11.9)	6.50*** [1.61]
	Comparison	416	15.2 (14.1)	338	17.0 (12.0)	
Oral reading fluency (cwpm)	Project	433	18.7 (23.9)	341	44.2 (29.3)	15.77*** [3.36]
	Comparison	416	26.5 (28.1)	338	33.5 (26.9)	
Reading comprehension : Score (out of 6)	Project	433	0.7 (1.1)	342	1.9 (1.6)	0.54* [0.22]
	Comparison	416	1.2 (1.6)	342	1.8 (1.7)	
Sentence dictation: Score (out of 10)	Project	433	3.4 (3.2)	342	5.2 (3.2)	1.10* [0.46]
	Comparison	416	4.5 (3.3)	342	5.0 (3.5)	

*Legend of statistical significance of differences between project and comparison schools: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. SD – Standard Deviation.*

Table 17 presents the mean scores for boys and girls at both baseline and endline for program schools for all assessment tasks. At baseline, girls slightly outperformed boys in letter naming fluency with an average score of 34.7 vs. 31.9 clpm, while boys had marginally higher scores in nonword reading fluency (11.0 vs. 10.7 cnwpm), oral reading fluency (18.7 vs. 18.4), and reading comprehension (0.7 vs. 0.6). In contrast, at endline, boys scored higher than girls in four out of five tasks: letter naming fluency, nonword reading fluency, oral reading fluency, and reading comprehension. While boys performed slightly better than girls at both baseline and endline on most literacy skill tasks, the differences were small and not statistically significant at $p < 0.05$, $p < 0.01$, or $p < 0.001$. This suggests that the program supported both boys and girls comparably to improve their foundational literacy skills.

Table 17. Early Grade Literacy Skill Assessment Results for Program Schools

Assessment Task	Gender	Baseline (2023)		Endline (2025)		Est. Diff. [Std. Error]
		Count	Mean (SD)	Count	Mean (SD)	
Letter naming fluency (clpm)	Boys	433	31.9 (23.7)	340	55.4 (25.6)	-3.16 [2.34]
	Girls	387	34.7 (23.7)	355	55.1 (26.2)	
Nonword reading fluency (cnwpm)	Boys	433	11.0 (11.4)	340	19.8 (11.9)	-0.23 [1.07]
	Girls	387	10.7 (11.3)	355	19.4 (12.8)	
Oral reading fluency (cwpm)	Boys	433	18.7 (23.9)	341	44.2 (29.3)	-1.91 [2.53]
	Girls	387	18.4 (24.3)	354	42.0 (30.9)	
Reading comprehension: Score (out of 6)	Boys	433	0.7 (1.1)	342	1.9 (1.6)	-0.24 [0.14]
	Girls	387	0.6 (1.0)	360	1.6 (1.5)	
Sentence dictation: Score (out of 10)	Boys	433	3.4 (3.2)	342	5.2 (3.2)	-0.11 [0.30]
	Girls	387	3.5 (3.1)	360	5.3 (3.4)	

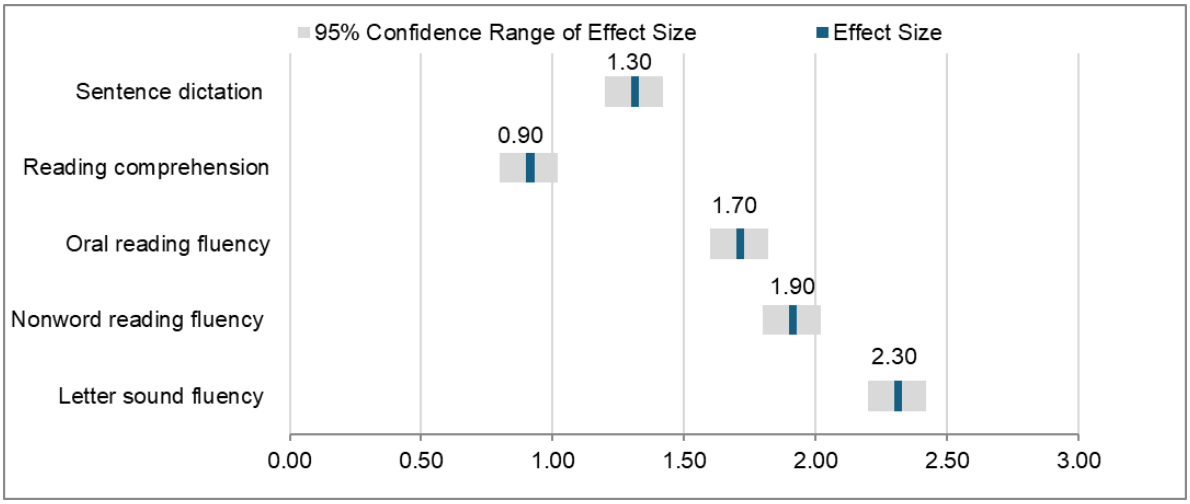
*Legend of statistical significance of differences between project and comparison schools: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. SD – Standard Deviation.*

Effect Sizes

Differences in children's gains from baseline to endline can be further examined through the calculated effect sizes across the assessment tasks (Figure 6). The effect size statistic is used to make comparisons across measures that use different scales or units[16]. For the purposes of this analysis, the standardized mean effect size statistic was used, through which an effect size of 0.8 or higher is considered large. The effect sizes for the Literacy Instruction Scale-up Program in Jodhpur were large across all five assessment subtasks. The largest effect size was found in the case of letter naming fluency (2.3), followed by non-word reading fluency (1.9), oral reading fluency (1.7), sentence dictation (1.3) and reading comprehension (0.9) respectively. These results suggest that the program in Jodhpur was particularly effective at building children's decoding, fluency, reading comprehension and writing skills.

[16] The effect size for each assessment task was calculated by determining the adjusted difference (i.e. after controlling for several school, teacher and child-level covariates) in two-year gains between end-of-Grade 3 children in program and comparison schools through OLS liner regression analysis. The resulting difference was then divided by the adjusted pooled standard deviation of the program and comparison groups.

Figure 6. Early Grade Literacy Skill Assessment Results – Program Effect Sizes



Predictors of Children's Literacy Skills at Endline

Multivariate regression analyses were performed on the 2025 endline evaluation data to identify factors, other than the program intervention, that could have influenced children's literacy skills. The regression models included various school-, teacher-, and child-specific indicators as covariates, in addition to the school type (program or comparison). These analyses helped identify the factors for which there was strong evidence of influence on children's performance on the endline assessment. Strong evidence is defined as a factor having a statistically significant association with at least three out of five literacy skill assessment tasks and there is consistency in the sign and significance across regression models. The full results from the regression analyses are available in Annex C , while the highlights from the analyses are summarized as follows:

1. **Classroom Type:** Children in multigrade classrooms (as compared to monograde classrooms) on average performed worse on the oral passage reading and reading comprehension tasks ($p < 0.05$ and $p < 0.01$ respectively). They also were more likely to score zero on the nonword ($p < 0.001$), oral passage reading ($p < 0.01$) and reading comprehension ($p < 0.01$) tasks. However, there were no statistically significant differences in the likelihood that they achieved oral reading and comprehension benchmarks.

2. **Grade 3 Teacher's Gender:** Children taught by male teachers, on average, scored higher across all assessment tasks in contrast to children taught by female teachers. These differences were statistically significant; letter naming, nonword reading and oral passage reading at $p < 0.01$, and reading comprehension and sentence dictation at $p < 0.05$. They were also less likely to score zero on the oral passage reading and sentence dictation tasks ($p < 0.05$ for both tasks, respectively). Lastly, children with male teachers were also more likely to achieve the oral reading fluency benchmarks ($p < 0.01$ for both the 35+ cwpm and 45+ cwpm benchmarks), but they were no more likely to achieve the reading comprehension benchmark (80% +).

3. **Children's Gender:** At endline, boys scored, on average, significantly higher than the girls on the letter naming ($p < 0.05$), nonword reading ($p < 0.05$), oral passage reading ($p < 0.05$) and reading comprehension assessment tasks ($p < 0.001$). They were also less likely to score zero on the oral passage reading ($p < 0.01$), reading comprehension ($p < 0.001$) and sentence dictation tasks ($p < 0.05$). Boys also had a higher chance of achieving the oral reading fluency and reading comprehension benchmarks ($p < 0.01$ for the 35+ cwpm ORF benchmark and $p < 0.05$ for the 45+ cwpm ORF and 80%+ reading comprehension benchmarks). Interestingly, the differences in performance by gender was more noticeable between girls and boys in comparison schools, with statistically significant differences on four out of the five assessment tasks. In contrast, there was only weak evidence on the difference between boys and girls at endline in program schools with statistically significant differences only on one task.

4. **Children's home language:** Children who reported speaking Hindi primarily at home, on average, scored less than children who reported speaking other languages or dialects on four out of the five assessment tasks. This included the nonword reading ($p < 0.05$), oral passage reading ($p < 0.01$), reading comprehension ($p < 0.001$) and sentence dictation ($p < 0.01$) tasks. There was also stronger evidence for the influence of this factor in program schools than comparison schools, where it was only significant for one assessment task. Moreover, children speaking Hindi at home more likely to score zero on the reading comprehension task ($p < 0.001$). Similar to children's overall performance, children who speaking Hindi at home were less likely to achieve the set reading benchmarks: ORF benchmark of 35+ cwpm ($p < 0.05$), ORF benchmark of 45+ cwpm ($p < 0.01$), and reading comprehension ($p < 0.05$). These findings are surprising because Hindi was the language of instruction in the schools and assessments were conducted to evaluate children's Hindi literacy skills. This may reflect that Hindi speaking children did not benefit form the program much in a pre-dominantly non-Hindi speaking, teaching and learning environment.[17] Globally, there is strong evidence that children learn to read best in a language they know and understand, therefore more investigation is needed to further explore this finding.

5. **Household standard of living:** Children from more affluent homes (as measured by the number of amenities reported to be present in the home) on average performed better on four out of the five assessment tasks. This includes letter naming ($p < 0.01$), nonword reading ($p < 0.01$), reading comprehension ($p < 0.001$), and sentence dictation ($p < 0.01$). These children were also less likely to score zero on the oral passage ($p < 0.05$), reading comprehension ($p < 0.01$) and sentence dictation ($p < 0.01$), tasks. Moreover, they had significantly higher chance at achieving the ORF 45+ cwpm ($p < 0.05$) and reading comprehension benchmarks ($p < 0.001$). Given the consistency in significance across the various regressions, there is strong evidence for the influence of this factor on children's literacy skills at endline.

[17] Less than half of the teachers and 12 to 24 percent of the children across program and comparison group schools included in the evaluation sample were speaking in Hindi at home.

6. **Living in a literate household:** Children who reported living with literate family members on average performed better, with statistically significant differences on all assessment tasks except oral reading fluency ($p < 0.01$ for letter naming and $p < 0.05$ for nonword reading, reading comprehension and sentence dictation tasks). These children also had a significantly higher chance of achieving both oral reading fluency benchmarks ($p < 0.05$ for each benchmark). However, this factor was not associated with children's likelihood of scoring zero.

7. **Access to reading materials at home:** Children who reported that they had access to reading materials or print items at home on average scored significantly higher on the nonword reading ($p < 0.05$), oral reading fluency ($p < 0.01$), reading comprehension ($p < 0.01$), and sentence dictation tasks ($p < 0.01$). Children with access to reading materials at home were also significantly less likely to score zero on all assessment tasks ($p < 0.001$ for all tasks). There was stronger evidence for the influence of this factor in comparison schools where it was statistically significant for all tasks. In contrast, within program schools, there was weak evidence for its influence as it was not significant on any task. The presence of classroom reading corners and regular book checkout opportunities in program schools may have mitigated the influence of this factor on children's literacy skills in program schools.

8. **Habit of reading**: Children's self-reported data on reading voluntarily for fun at least three times in the last week preceding the assessment was used as a proxy indicator for children's habit of reading in the regression models. The analyses showed that the children who had done so performed significantly (at $p < 0.001$) better than other children across all five assessment tasks. The strong evidence of the influence of good reading habits was also detected in the sample of children from program and comparison schools. Moreover, children with such a habit were least likely to score zero on all the EGLS assessments tasks, and their chances of achieving both oral reading fluency and reading comprehension benchmarks were significantly higher ($p < 0.001$ for both ORF benchmarks and $p < 0.01$ for the reading comprehension benchmark).

For the other school-, teacher-, and child-specific factors, regression analyses did not find evidence of their strong influence on children's literacy skills.

Conclusions and Recommendations

The 2023-25 impact evaluation showed that Room to Read's Literacy Instruction Scale-up Program on Jodhpur significantly improved children's foundational Hindi literacy skills. Children in program schools outperformed their peers in comparison schools across all five assessment tasks, with particularly strong gains in letter naming fluency and oral reading fluency. The proportion of non-readers (measured through incidence of zero scores) also declined substantially in program schools, and children were 3-4 times more likely to meet reading fluency and comprehension benchmarks than those in comparison schools. Furthermore, across all five assessment tasks, children in program schools made two-year gains in their reading and writing skills that were approximately 1.1 to 1.4 times greater than those seen in comparison schools.

Key program components – such as teacher training, access to quality TLMs, establishment of cluster libraries and classroom reading corners – were implemented with moderate to high fidelity and likely contributed to these gains. However, the cascade model of training only produced modest gains in participant knowledge and teacher coaching and mentoring support was not implemented as frequently as desired. Additionally, high absenteeism rates and repetition rates among children were noted in program schools.

Importantly, the program appears to benefit both boys and girls comparably, with no statistically significant gender differences in endline performance, suggesting progress toward equitable outcomes.

Recommendations based on the findings from this evaluation and data from various program monitoring records includes:

1. **Enhance Fidelity of Coaching support**: Increase the frequency and quality of PEEO-led coaching visits by clarifying expectations and strengthening accountability mechanisms.
2. **Differentiated Support for Multigrade classrooms**: Provide tailored training and coaching support on effective strategies for managing large, multigrade classrooms to support more equitable learning outcomes across classroom types.
3. **Empower Female Teachers**: Given that children with male teachers performed better, gender-sensitive training and mentorship programs for female teachers should be explored to improve the support they receive.
4. **Improve Attendance and Grade Progression**: Address higher absenteeism and grade repetition in program schools through targeted support, such as early warning systems, home visits, or remedial programs.
5. **Support Gender Equity**: Though gains were similar, boys slightly outperformed girls. Tailored strategies –such as targeted classroom activities and community encouragement –can help ensure girls' progress is sustained.
6. **Sustain Reading Habits**: Positive and statistically significant association between children's reading habits and their literacy skills was found in this evaluation. As such, it

is important that schools maintain momentum by continuing book check-out systems and book rotations to encourage children to read frequently at home and by further integrating reading for pleasure into school routines.

7. **Strengthen Community Engagement**: Program schools should improve outreach practices to enhance parent involvement and reinforce reading habits at home.
8. **Investigate Home Language Findings**: Further exploration should be conducted to help understand why children speaking Hindi at home underperformed in Hindi assessment as compared children who speak other languages. Possible areas to probe are teacher language practices, home support structures and additional household demographic and economic data to devise language-sensitive strategies to address the inequity in learning outcomes.

Overall, the program demonstrates that with a structured, school-based approach, significant improvements in foundational literacy skills are possible at scale. Future program implementation should focus on further strengthening the capacity of system-based actors, addressing equity gaps, and expanding efforts to integrate effective program elements into the broader education system for sustained literacy improvements.

Annexes

A. EGLSA Tool Reliability Analyses

To examine the reliability and validity of the EGLSA tool and data, several statistics are routinely calculated as part of the analysis, including but limited to: 1) Cronbach Alpha – which measures the internal consistency within the tool; 2) Average inter-item correlation – which measures the consistency between the individual items within each task: and 3) Correlations among the different tasks for program, comparison and all schools at baseline and endline – which measures the construct validity of the assessment or the extent to which the varying tasks are related and likely to be measuring the same underlying skill. The results are presented in Table 18, 19 and 20 below.

The Cronbach alpha for each individual assessment task was quite high, indicating there was a high level of internal consistency in the tool. The average inter-item correlations for each task also indicated good to high consistency between items, indicating that they are measure the same foundational literacy skill.

Table 18. Results of EGLSA Tool Reliability Analyses

Statistics	Letter Naming	Nonword Reading	Oral Reading	Reading Comprehension	Dictation
Average inter-item correlation	0.41	0.35	0.61	0.31	0.50
Cronbach Alpha	0.99	0.96	0.99	0.73	0.86

Furthermore, both at baseline and endline across program, comparison and all schools, the correlations between the tasks are high and statistically significant ($p < 0.001$). This provides strong support for the validity of the assessment tool and that the tasks are indeed measuring the same underlying skill, children's foundational literacy skills.

Table 19. Pair-wise Correlation Coefficients Between EGLSA Task Scores - Baseline

	Letter naming	Nonword reading	Oral reading	Reading comprehension	Sentence dictation
Program Schools					
Letter naming	1.000				
Nonword reading	0.748	1.000			
Oral reading	0.754	0.856	1.000		
Reading comprehension	0.560	0.617	0.713	1.000	
Sentence dictation	0.732	0.697	0.762	0.622	1.000
Comparison Schools					
Letter naming	1.000				
Nonword reading	0.659	1.000			
Oral reading	0.586	0.773	1.000		

	Letter naming	Nonword reading	Oral reading	Reading comprehension	Sentence dictation
Reading comprehension	0.486	0.540	0.612	1.000	
Sentence dictation	0.620	0.616	0.668	0.617	1.000
All Sampled Schools					
Letter naming	1.000				
Nonword reading	0.703	1.000			
Oral reading	0.663	0.810	1.000		
Reading comprehension	0.527	0.578	0.656	1.000	
Sentence dictation	0.677	0.655	0.712	0.620	1.000
**All correlation coefficients were statistically significant at p < 0.001.*					

Table 20. Pair-wise Correlation Coefficients Between EGLSA Task Scores - Endline

	Letter naming	Nonword reading	Oral reading	Reading comprehension	Sentence dictation
Program Schools					
Letter naming	1.000				
Nonword reading	0.744	1.000			
Oral reading	0.669	0.814	1.000		
Reading comprehension	0.531	0.589	0.616	1.000	
Sentence dictation	0.637	0.635	0.633	0.600	1.000
Comparison Schools					
Letter naming	1.000				
Nonword reading	0.818	1.000			
Oral reading	0.761	0.888	1.000		
Reading comprehension	0.610	0.692	0.730	1.000	
Sentence dictation	0.698	0.744	0.724	0.679	1.000
All Sampled Schools					
Letter naming	1.000				
Nonword reading	0.786	1.000			
Oral reading	0.726	0.851	1.000		
Reading comprehension	0.568	0.640	0.666	1.000	
Sentence dictation	0.668	0.693	0.677	0.643	1.000
**All correlation coefficients were statistically significant at p < 0.001.*					

B. Data Quality Assessments – EGLSA

To ensure accuracy in scoring children's responses particularly for timed tasks, assessors undergo IRR assessments during training. These assessments involve scoring a mock EGLS session, typically based on a video of a child being assessed or a simulation by another individual. Each assessor's ratings are then compared to a set of correct responses, and their accuracy is measured as a percentage agreement. The sector standard for IRR is a minimum of 95 percent agreement, which all supervisors and assessors must meet to qualify for data collection.

During training, three initial IRR assessments were conducted using the EGLSA tool. Assessors who failed to meet the threshold set at 95 percent agreement on any of these three initial assessments, were provided with an additional day of training and a fourth IRR assessment to determine suitability for data collection. Table 21 presents the results for each assessor. As expected, scores generally improved from IRR1 to IRR4, reflecting increased familiarity and practice with the tool. By IRR4, all assessors had achieved an IRR score of 95 percent or higher, demonstrating notable progress from the initial assessment.

Table 21. IRR Results for EGLSA during Training

Assessor	IRR1	IRR2	IRR3	IRR4
Assessor 1	89%	80%	95%	
Assessor 2	92%	89%	93%	97%
Assessor 3	93%	84%	95%	
Assessor 4	94%	86%	92%	100%
Assessor 5	84%	96%	96%	
Assessor 6	88%	87%	98%	
Assessor 7	93%	85%	96%	
Assessor 8	99%	89%	96%	
Assessor 9	89%	81%	92%	100%
Assessor 10	91%	81%	96%	
Assessor 11	92%	74%	92%	97%
Assessor 12	98%	90%	97%	
Assessor 13	88%	95%	94%	
Assessor 14	81%	96%	95%	
Assessor 15	80%	91%	94%	98%
Assessor 16	83%	90%	92%	99%
Assessor 17	91%	78%	94%	100%
Assessor 18	98%	90%	91%	97%
Assessor 19	99%	89%	96%	
Assessor 20	85%	92%	94%	99%
Assessor 21	90%	91%	97%	
Assessor 22	79%	91%	89%	100%
Average	**90%**	**88%**	**94%**	**99%**

Ongoing monitoring of assessor reliability and consistency is essential throughout data collection. To support this, IRR assessments were conducted at each sampled school with a subset of children. These assessments were carried out using both simultaneous and retesting methods where a supervisor and an assessor either rated the same child at the same time (simultaneous)or on separate days (re-test), respectively.

According to best practices in the sector, a minimum of 150 IRR assessments should be conducted for all EGLS assessments. For the Endline evaluation of the Literacy Instruction Scale-up Program in Jodhpur, a total of 514 IRR assessments were analysed, significantly exceeding this benchmark. Approximately 60 percent were conducted simultaneously, while the remaining 40 percent were conducted as re-test assessments

The overall inter-rater reliability during data collection was high, averaging 96 percent agreement across all assessments. Moreover, among simultaneous assessments it was 99 percent, while for re-test assessments it was 95 percent. Task-specific reliability was particularly strong for letter naming, oral passage reading, and nonword reading. While reading comprehension and sentence dictation tasks showed slightly lower agreement levels, they still met the established threshold of 95 percent, indicating acceptable consistency. Table 22 presents the percent agreement by assessor, assessment task, and overall performance.

Table 22. IRR Results for EGLSA during Data Collection

Assessor	Letter naming	Nonword reading	Oral Reading	Reading comprehension	Sentence dictation	Mean
Assessor 1	99%	99%	99%	98%	98%	99%
Assessor 2	100%	98%	99%	95%	96%	98%
Assessor 3	98%	98%	99%	93%	100%	97%
Assessor 4	100%	99%	100%	99%	100%	99%
Assessor 5	100%	99%	99%	98%	99%	99%
Assessor 6	100%	100%	99%	96%	99%	99%
Assessor 7	98%	98%	98%	88%	99%	96%
Assessor 8	99%	97%	98%	88%	93%	95%
Assessor 9	98%	91%	93%	82%	85%	90%
Assessor 10	98%	97%	98%	94%	94%	96%
Assessor 11	99%	98%	99%	99%	96%	98%
Assessor 12	99%	98%	99%	93%	96%	97%
Assessor 13	95%	87%	91%	92%	86%	90%
Average	**99%**	**97%**	**98%**	**93%**	**95%**	**96%**

** Per standard practices in the sector, all EGLS data is anonymized before analysis to protect individual's identity. As such assessor numbers are generated randomly.*

C. Regression Results

Table 23. Difference-in-Difference in Children's Literacy Skills Scores – Results from OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Type	School Type [Program -1, Comparison – 0]	-9.16*** [2.41]	-3.15* [1.27]	-5.13* [2.56]	-0.35* [0.14]	-0.51 [0.32]
Wave	Wave [Endline – 1, Baseline -0]	0.51 [1.92]	1.33 [1.02]	5.88** [2.16]	0.48** [0.15]	0.45 [0.29]
Interact. b/w wave and type	Type * Wave [Program and Endline – 1, Otherwise – 0]	20.00*** [2.93]	6.55*** [1.48]	16.38*** [3.26]	0.49** [0.19]	1.03* [0.41]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.63 [1.03]	-0.90 [0.46]	-1.91* [0.94]	-0.29*** [0.05]	-0.24 [0.13]
	Age (in years)	2.68*** [0.45]	0.75*** [0.23]	1.90*** [0.49]	0.09*** [0.03]	0.24*** [0.06]
	Child attended pre-school/ECD [Yes – 1, No – 0]	0.76 [1.30]	-0.14 [0.56]	0.19 [1.26]	0.02 [0.07]	0.02 [0.16]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.91 [1.36]	-1.59* [0.72]	-5.96*** [1.35]	-0.35*** [0.06]	-0.54** [0.17]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	3.47 [1.82]	1.63 [0.96]	3.90* [1.97]	0.32** [0.10]	0.48 [0.27]
	Grade 3 enrolment	-0.11 [0.13]	-0.03 [0.06]	-0.00 [0.12]	-0.00 [0.01]	-0.03 [0.02]
	Attendance (%) on the day of the visit	0.07 [0.04]	0.02 [0.02]	0.07 [0.05]	0.01 [0.00]	-0.00 [0.01]
Constant		14.74** [5.45]	6.64* [2.75]	3.60 [5.74]	-0.08 [0.34]	2.50*** [0.76]
Observations		2993	2992	2992	3006	3006
R-squared – Within		0.11	0.07	0.11	0.12	0.05
R-squared – Between		0.06	0.08	0.14	0.14	0.10
R-squared – Overall		0.10	0.07	0.12	0.12	0.06

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 24. Difference-in-Difference in Children's Literacy Skills scores – Boys – Results from OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores - Boys				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Type	School Type [Program - 1, Comparison – 0]	-12.05*** [2.61]	-3.83** [1.40]	-6.01* [2.55]	-0.44** [0.16]	-0.91* [0.36]
Wave	Wave [Endline – 1, Baseline -0]	0.94 [2.02]	1.19 [1.14]	6.30** [2.22]	0.50** [0.18]	0.33 [0.34]
Interact. b/w wave and type	Type * Wave [Program and Endline – 1, Otherwise – 0]	20.66*** [3.15]	6.50*** [1.61]	15.77*** [3.36]	0.54* [0.22]	1.10* [0.46]
Children's demographic and economic features	Age (in years)	2.61*** [0.54]	0.60* [0.30]	1.50* [0.62]	0.10* [0.04]	0.22** [0.07]
	Child attended pre-school/ECD [Yes – 1, No – 0]	3.21* [1.43]	1.04 [0.66]	3.52* [1.46]	0.14 [0.09]	0.29 [0.19]
	Child speaks Hindi at home [Yes – 1, No – 0]	-2.81 [1.77]	-2.65* [1.03]	-9.20*** [1.68]	-0.53*** [0.09]	-0.96*** [0.24]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	2.02 [2.19]	1.45 [1.11]	3.23 [2.15]	0.34** [0.12]	0.46 [0.30]
	Grade 3 enrolment	-0.03 [0.12]	0.02 [0.06]	0.09 [0.12]	0.00 [0.01]	-0.02 [0.02]
	Attendance (%) on the day of the visit	0.04 [0.05]	0.02 [0.02]	0.06 [0.05]	0.01 [0.00]	0.00 [0.01]
Constant		17.33** [6.62]	7.77* [3.60]	6.61 [7.49]	-0.13 [0.44]	2.65** [0.89]
Observations		1521	1521	1522	1527	1527
R-squared – Within		0.12	0.08	0.14	0.14	0.06
R-squared – Between		0.06	0.06	0.13	0.15	0.10
R-squared – Overall		0.11	0.08	0.14	0.14	0.08

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 25. Difference-in-Difference in Children's Literacy Skills scores – Girls – Results from OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores - Girls				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Type	School Type [Program - 1, Comparison – 0]	-5.98* [2.91]	-2.78 [1.47]	-4.24 [3.09]	-0.27 [0.15]	-0.20 [0.37]
Wave	Wave [Endline – 1, Baseline -0]	0.53 [2.60]	1.26 [1.35]	5.26 [2.97]	0.45** [0.17]	0.49 [0.35]
Interact. b/w wave and type	Type * Wave [Program and Endline – 1, Otherwise – 0]	18.21*** [3.68]	6.84*** [1.81]	16.69*** [4.15]	0.45* [0.21]	1.00* [0.49]
Children's demographic and economic features	Age (in years)	2.62*** [0.70]	0.87** [0.32]	2.13** [0.75]	0.06* [0.03]	0.20* [0.08]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-2.81 [1.82]	-1.55 [0.82]	-3.64 [1.95]	-0.13 [0.09]	-0.32 [0.23]
	Child speaks Hindi at home [Yes – 1, No – 0]	0.85 [1.77]	-0.55 [0.84]	-3.08 [1.86]	-0.20* [0.09]	-0.14 [0.23]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	4.65* [1.99]	1.41 [0.96]	4.36* [2.20]	0.29* [0.11]	0.54* [0.27]
	Grade 3 enrolment	-0.21 [0.15]	-0.08 [0.07]	-0.11 [0.15]	-0.00 [0.01]	-0.04* [0.02]
	Attendance (%) on the day of the visit	0.10 [0.06]	0.03 [0.03]	0.08 [0.07]	0.01* [0.00]	0.00 [0.01]
Constant		13.58 [7.88]	5.43 [3.78]	0.84 [8.61]	-0.17 [0.41]	2.54* [1.01]
Observations		1472	1471	1470	1479	1479
R-squared – Within		0.10	0.07	0.10	0.10	0.05
R-squared – Between		0.11	0.08	0.13	0.13	0.11
R-squared – Overall		0.11	0.08	0.11	0.11	0.06

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 26. Difference-in-Difference in Children's Literacy Skills scores – Child's Gender in Program Schools – Results from OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores – Program Schools				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Child Gender	Child's Gender [Girls - 1, Boys – 0]	3.32 [1.72]	-0.16 [0.82]	0.37 [1.63]	-0.11 [0.08]	0.19 [0.23]
Wave	Wave [Endline – 1, Baseline - 0]	22.32*** [2.41]	8.13*** [1.14]	23.44*** [2.47]	1.09*** [0.12]	1.57*** [0.30]
Interact. b/w wave and type	Wave * Gender [Endline and a Girl – 1, Otherwise – 0]	-3.16 [2.34]	-0.23 [1.07]	-1.91 [2.53]	-0.24 [0.14]	-0.11 [0.30]
Children's demographic and economic features	Age (in years)	2.73*** [0.58]	0.74* [0.32]	2.41*** [0.66]	0.06* [0.03]	0.21** [0.07]
	Child attended pre-school/ECD [Yes – 1, No – 0]	0.37 [1.69]	-0.62 [0.75]	-0.53 [1.71]	-0.02 [0.09]	-0.16 [0.20]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.60 [1.74]	-1.84* [0.75]	-5.47** [1.76]	-0.26** [0.08]	-0.21 [0.20]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	3.00 [2.93]	1.11 [1.33]	3.56 [2.88]	0.32** [0.11]	0.48 [0.40]
	Grade 3 enrolment	-0.06 [0.20]	-0.02 [0.08]	-0.03 [0.16]	-0.00 [0.01]	-0.03 [0.03]
	Attendance (%) on the day of the visit	0.13 [0.07]	0.04 [0.03]	0.16* [0.07]	0.01* [0.00]	0.00 [0.01]
Constant		-1.54 [7.53]	2.37 [4.11]	-13.19 [7.49]	-0.31 [0.35]	1.77 [1.01]
Observations		1515	1515	1515	1522	1522
R-squared – Within		0.21	0.13	0.20	0.19	0.09
R-squared – Between		0.08	0.14	0.17	0.13	0.08
R-squared – Overall		0.19	0.14	0.20	0.18	0.09

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 27. Difference-in-Difference in Probability of Zero Scores – Results from Logistic regressions

Covariates		Zero Scores: Early Grade Literacy Skills Assessment				
		Zero Score: Letter naming	**Zero Score: Nonword reading**	**Zero Score: Oral passage reading**	**Zero Score: Reading comprehension**	**Zero Score: Sentence Dictation**
Type	School Type [Program - 1, Comparison – 0]	0.04 [0.34]	0.05 [0.26]	0.38 [0.24]	0.50* [0.22]	0.22 [0.23]
Wave	Wave [Endline – 1, Baseline -0]	0.19 [0.24]	-0.33 [0.17]	-0.22 [0.21]	-0.72*** [0.19]	0.28 [0.20]
Interact. b/w wave and type	Type * Wave [Program and Endline – 1, Otherwise – 0]	-1.18** [0.41]	-1.05** [0.32]	-1.38*** [0.36]	-0.85** [0.26]	-0.89** [0.30]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.06 [0.16]	0.07 [0.12]	0.24* [0.11]	0.40*** [0.10]	0.15 [0.10]
	Age (in years)	-0.30*** [0.09]	-0.27*** [0.06]	-0.26*** [0.05]	-0.20*** [0.04]	-0.28*** [0.05]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.14 [0.21]	0.07 [0.15]	-0.10 [0.14]	-0.11 [0.12]	-0.07 [0.13]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.32 [0.21]	-0.11 [0.13]	0.04 [0.13]	0.48*** [0.12]	0.07 [0.14]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	-0.48* [0.21]	-0.42* [0.18]	-0.60** [0.20]	-0.59*** [0.16]	-0.36 [0.19]
	Grade 3 enrolment	0.02 [0.01]	0.03* [0.01]	0.02 [0.01]	0.02 [0.01]	0.02 [0.01]
	Attendance (%) on the day of the visit	0.00 [0.01]	0.00 [0.00]	-0.00 [0.00]	-0.00 [0.00]	-0.00 [0.00]
Constant		-0.38 [0.89]	0.31 [0.69]	1.05 [0.61]	2.00*** [0.51]	1.05 [0.59]
$\ln \sigma^2_u$		0.33 [0.21]	0.25 [0.18]	-0.04 [0.21]	-0.07 [0.19]	-0.12 [0.19]
Observations		3006	3006	3006	3006	3006

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 28. Difference-in-Difference in Probability of Achieving Reading Benchmarks – Results from Logistic regressions

Covariates		Reading Benchmarks: Early Grade Literacy Skills Assessment		
		Oral Reading Fluency: 35+ cwpm	**Oral Reading Fluency: 45+ cwpm**	**Reading comprehension: Score 80%+**
Type	School Type [Program -1, Comparison – 0]	-0.42* [0.21]	-0.26 [0.24]	-1.60** [0.59]
Wave	Wave [Endline – 1, Baseline - 0]	0.56** [0.18]	0.42* [0.18]	0.38 [0.47]
Interact. b/w wave and type	Type * Wave [Program and Endline – 1, Otherwise – 0]	1.28*** [0.26]	1.17*** [0.29]	1.37* [0.64]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.26** [0.09]	-0.16 [0.10]	-0.57** [0.20]
	Age (in years)	0.09* [0.05]	0.12* [0.05]	0.25* [0.11]
	Child attended pre-school/ECD [Yes – 1, No – 0]	0.04 [0.11]	0.00 [0.12]	0.07 [0.26]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.45*** [0.13]	-0.79*** [0.16]	-1.00** [0.39]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	0.18 [0.15]	0.20 [0.17]	0.30 [0.32]
	Grade 3 enrolment	-0.01 [0.01]	0.01 [0.01]	0.01 [0.02]
	Attendance (%) on the day of the visit	0.01 [0.00]	0.01 [0.00]	0.02 [0.01]
Constant		-1.86*** [0.55]	-3.13*** [0.58]	-7.22*** [1.43]
ln σ^2_u		-0.25 [0.19]	-0.26 [0.18]	0.77** [0.28]
Observations		3006	3006	3006

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 29. Probability of Zero Scores – Baseline (2023) – Results from Logistic regressions

Covariates		Zero Scores: Early Grade Literacy Skills Assessment - Baseline				
		Zero Score: Letter naming	**Zero Score: Nonword reading**	**Zero Score: Oral passage reading**	**Zero Score: Reading comprehension**	**Zero Score: Sentence Dictation**
Type	School Type [Program -1, Comparison – 0]	0.29 [0.33]	0.21 [0.23]	0.38 [0.21]	0.44* [0.19]	0.29 [0.21]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.08 [0.20]	0.11 [0.13]	0.20 [0.12]	0.36** [0.11]	0.08 [0.12]
	Age (in years)	-1.10** [0.09]	-0.66*** [0.07]	-0.60*** [0.07]	-0.37** [0.06]	-0.62*** [0.07]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.15 [0.26]	0.07 [0.18]	-0.15 [0.16]	-0.03 [0.15]	-0.23 [0.16]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.19 [0.26]	0.12 [0.17]	0.16 [0.16]	0.54*** [0.14]	0.14 [0.17]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	-0.89 [0.27]	-0.17 [0.22]	-0.51* [0.19]	-0.44* [0.18]	-0.43 [0.20]
	Grade 3 enrolment	0.71 [0.02]	0.58* [0.02]	0.34 [0.02]	0.16 [0.01]	0.39 [0.02]
	Attendance (%) on the day of the visit	0.16 [0.01]	0.00 [0.01]	-0.11 [0.01]	-0.06 [0.00]	-0.15 [0.01]
Observations		1580	1580	1580	1580	1580
Pseudo R-squared		0.027	0.024	0.035	0.044	0.029

Standardized beta coefficients reported, Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 30. Probability of Zero Scores – Endline (2025) – Results from Logistic regressions

Covariates		Zero Scores: Early Grade Literacy Skills Assessment - Endline				
		Zero Score: Letter naming	**Zero Score: Nonword reading**	**Zero Score: Oral passage reading**	**Zero Score: Reading comprehension**	**Zero Score: Sentence Dictation**
Type	School Type [Program -1, Comparison – 0]	-1.42 [0.39]	-0.90* [0.27]	-1.13** [0.28]	-0.33 [0.19]	-0.73** [0.21]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	0.08 [0.21]	0.08 [0.17]	0.55** [0.15]	0.41** [0.12]	0.31* [0.13]
	Age (in years)	-1.23* [0.13]	-0.58* [0.08]	-0.43* [0.08]	-0.22 [0.06]	-0.39* [0.07]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.76 [0.30]	-0.18 [0.21]	0.02 [0.21]	-0.03 [0.15]	0.16 [0.17]
	Child speaks Hindi at home [Yes – 1, No – 0]	-1.65* [0.43]	-0.66 [0.33]	0.01 [0.25]	0.35* [0.18]	0.12 [0.21]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	-1.28* [0.32]	-0.97** [0.24]	-0.81** [0.22]	-0.56** [0.19]	-0.41 [0.20]
	Grade 3 enrolment	1.18 [0.02]	0.37 [0.01]	0.48 [0.01]	0.20 [0.01]	0.34 [0.01]
	Attendance (%) on the day of the visit	-0.50 [0.01]	-0.33 [0.01]	-0.61* [0.01]	-0.38* [0.00]	-0.34 [0.01]
Observations		1426	1426	1426	1426	1426

Pseudo R-squared	0.067	0.047	0.064	0.036	0.036

Standardized beta coefficients reported, Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 31. Probability of Achieving Reading Benchmarks – Baseline (2023) – Logistic regressions

Covariates		Reading Benchmarks: Early Grade Literacy Skills Assessment		
		Oral Reading Fluency: 35+ cwpm	**Oral Reading Fluency: 45+ cwpm**	**Reading comprehension: Score 80%+**
Type	School Type [Program -1, Comparison – 0]	-0.43* [0.18]	-0.32 [0.19]	-4.95** [0.58]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.22 [0.12]	0.01 [0.14]	-2.36* [0.33]
	Age (in years)	0.17 [0.07]	0.30 [0.07]	1.48 [0.13]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.03 [0.17]	-0.01 [0.20]	-0.60 [0.43]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.66*** [0.16]	-1.20*** [0.22]	-6.98* [1.03]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	0.14 [0.19]	0.07 [0.21]	0.38 [0.53]
	Grade 3 enrolment	-0.01 [0.01]	0.16 [0.01]	0.40 [0.03]
	Attendance (%) on the day of the visit	0.16 [0.01]	0.02 [0.01]	3.13* [0.01]
Observations		1580	1580	1580
Pseudo R-squared		0.025	0.033	0.137

Standardized beta coefficients; Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 32. Probability of Achieving Reading Benchmarks – Endline (2025) – Logistic regressions

Covariates		Reading Benchmarks: Early Grade Literacy Skills Assessment		
		Oral Reading Fluency: 35+ cwpm	**Oral Reading Fluency: 45+ cwpm**	**Reading comprehension: Score 80%+**
Type	School Type [Program -1, Comparison – 0]	0.74*** [0.18]	0.88*** [0.18]	-0.47 [0.35]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.32** [0.12]	-0.26* [0.12]	-1.00* [0.23]
	Age (in years)	0.07 [0.06]	0.14 [0.06]	0.57 [0.14]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.06 [0.14]	-0.16 [0.14]	0.08 [0.32]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.22 [0.18]	-0.45** [0.19]	-1.14 [0.39]
School characteristics	Type of classroom [monograde - 1, multigrade - 0]	0.33 [0.18]	0.30 [0.19]	0.83 [0.33]
	Grade 3 enrolment	-0.20 [0.01]	0.05 [0.01]	-0.19 [0.02]
	Attendance (%) on the day of the visit	0.35 [0.00]	0.39 [0.01]	0.01 [0.01]

Observations	1426	1426	1426	
Pseudo R-squared	0.042	0.047	0.021	

Standardized beta coefficients; Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 33. Factors Influencing Children's Literacy Skills – Endline (2025) –OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Type	School Type [Program -1, Comparison – 0]	10.01*** [2.36]	2.53* [1.10]	10.62*** [2.84]	-0.01 [0.15]	0.14 [0.34]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-2.96* [1.40]	-1.32* [0.64]	-3.64* [1.46]	-0.39*** [0.09]	-0.33 [0.18]
	Age (in years)	2.14**[0.66]	0.59 [0.31]	1.47 [0.79]	0.05 [0.04]	0.12 [0.09]
	Child speaks Hindi at home [Yes – 1, No – 0]	-3.35 [2.07]	-2.23* [1.00]	-7.78** [2.39]	-0.57*** [0.12]	-0.78** [0.30]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-3.66 [1.90]	-1.32 [0.83]	-2.24 [1.93]	-0.01 [0.11]	-0.49 [0.26]
	Child repeating the current grade [Yes – 1, No – 0]	-1.30 [2.34]	-0.00 [1.07]	1.75 [2.88]	-0.13 [0.14]	0.04 [0.34]
	Absent from school at least one day in the last week [Yes – 1, No – 0]	-3.33 [1.85]	-0.58 [0.83]	-1.97 [2.03]	0.05 [0.10]	-0.11 [0.27]
	Household standard of living: Score (on a 10-point scale)	1.18** [0.44]	0.63** [0.19]	0.86 [0.51]	0.11*** [0.03]	0.17** [0.06]
Home literacy environment	Child was living with literate people in the last week [Yes – 1, No – 0]	5.01** [1.77]	1.89* [0.82]	3.59 [2.09]	0.21* [0.11]	0.61* [0.24]
	Someone read something to child at home in the last week [Yes – 1, No – 0]	4.32* [1.95]	1.50 [0.84]	2.65 [2.17]	0.17 [0.10]	0.48 [0.26]
	Someone told a story to the child at home in the last week [Yes – 1, No – 0]	1.04 [1.54]	0.67 [0.69]	0.76 [1.79]	0.16 [0.09]	0.10 [0.21]
	Read for fun at least thrice in last week [Yes – 1, No – 0]	10.07*** [1.80]	4.11*** [0.82]	8.48*** [1.97]	0.61*** [0.12]	1.00*** [0.25]
	Reading materials/print items at home [Yes – 1, No – 0]	5.12 [2.93]	2.74* [1.24]	8.03** [2.92]	0.40** [0.13]	1.07** [0.40]
Teacher characteristics	Teacher's Gender [Female – 1, Male – 0]	-6.16** [2.08]	-3.01** [0.92]	-7.04** [2.34]	-0.33* [0.13]	-0.62* [0.31]
	Teacher's age (years)	-0.04 [0.12]	-0.01 [0.05]	0.02 [0.13]	0.00 [0.01]	-0.02 [0.02]
	Teacher speaks Hindi at home [Yes – 1, No – 0]	-0.08 [2.20]	-0.38 [0.98]	0.97 [2.38]	0.08 [0.13]	-0.35 [0.33]
	Prior experience in teaching [Yes – 1, No – 0]	0.43 [2.62]	0.11 [1.14]	0.79 [2.87]	-0.02 [0.14]	0.20 [0.37]
School characteristics	Class size: 15 or less [Yes – 1, No – 0]	1.10 [2.23]	0.04 [0.98]	-0.86 [2.42]	-0.03 [0.13]	0.36 [0.30]
	Attendance (%) on the visit day	0.08 [0.05]	0.04 [0.02]	0.11 [0.06]	0.00 [0.00]	0.01 [0.01]

	Multigrade classroom for Grade 3 [Yes – 1, No – 0]	-4.15 [2.29]	-1.57 [1.04]	-5.21* [2.62]	-0.39** [0.13]	-0.54 [0.33]
	Classroom environment: Score [on an 8-point scale]	-0.24 [0.91]	0.13 [0.44]	-0.18 [1.03]	0.01 [0.06]	0.17 [0.13]
Constant		8.65 [11.14]	1.07 [5.18]	1.29 [13.10]	-0.30 [0.66]	0.60 [1.59]
Observations		1403	1402	1402	1416	1416
Adjusted R-squared		0.172	0.125	0.132	0.159	0.121

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 34. Factors influencing children's literacy skills – Endline (2025) – Program Schools - OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-1.08 [1.81]	-0.64 [0.76]	-2.65 [1.97]	-0.39** [0.12]	-0.01 [0.20]
	Age (in years)	1.82 [0.92]	0.51 [0.43]	2.00 [1.14]	0.03 [0.05]	0.13 [0.11]
	Child speaks Hindi at home [Yes – 1, No – 0]	-3.66 [2.91]	-2.83* [1.32]	-9.00** [3.27]	-0.51** [0.15]	-0.69* [0.33]
	Child attended pre-school [Yes – 1, No – 0]	-4.07 [2.64]	-2.60* [1.23]	-5.29 [2.82]	-0.14 [0.15]	-0.98** [0.33]
	Child repeating the current grade [Yes – 1, No – 0]	-1.64 [2.96]	0.80 [1.37]	5.79 [3.46]	-0.10 [0.17]	0.28 [0.40]
	Absent from school at least one day in the last week [Yes – 1, No – 0]	-2.03 [2.60]	-0.33 [1.15]	-0.93 [2.91]	0.11 [0.13]	-0.04 [0.36]
	Household standard of living: Score (10-pt scale)	1.93** [0.68]	1.03*** [0.29]	1.82* [0.70]	0.16*** [0.04]	0.23** [0.08]
Home literacy environment	Child was living with literate people in the last week [Yes – 1, No – 0]	4.10 [2.41]	1.68 [1.16]	2.34 [3.29]	0.13 [0.15]	0.70* [0.32]
	Someone read something to child at home in the last week [Yes – 1, No – 0]	2.52 [2.57]	1.19 [1.17]	2.01 [3.61]	0.26 [0.15]	0.60 [0.38]
	Someone told a story to the child at home in the last week [Yes – 1, No – 0]	3.72 [2.22]	1.36 [0.96]	1.33 [2.43]	0.10 [0.11]	0.09 [0.23]
	Read for fun at least thrice in last week [Yes – 1, No – 0]	8.51** [2.60]	2.88* [1.13]	7.78** [2.51]	0.54*** [0.16]	0.89** [0.31]
	Reading materials at home [Yes – 1, No – 0]	0.94 [4.04]	0.70 [1.58]	6.01 [4.11]	0.23 [0.18]	0.53 [0.52]
Teacher characteristics	Teacher's Gender [Female – 1, Male – 0]	-6.86* [3.40]	-3.65* [1.44]	-10.29** [3.56]	-0.39 [0.20]	-0.52 [0.42]
	Teacher's age (years)	-0.17 [0.17]	-0.05 [0.07]	-0.04 [0.17]	0.00 [0.01]	-0.04* [0.02]
	Teacher speaks Hindi at home [Yes – 1, No – 0]	0.07 [3.50]	-0.95 [1.53]	-0.70 [3.52]	-0.14 [0.19]	-0.37 [0.45]
	Prior experience in teaching [Yes – 1, No – 0]	1.93 [3.95]	1.28 [1.75]	3.97 [4.16]	0.01 [0.20]	0.59 [0.48]
Scho ol char	Class size: 15 or less [Yes – 1, No – 0]	0.19 [3.75]	-1.11 [1.55]	-5.12 [3.71]	-0.01 [0.20]	0.55 [0.49]

	Attendance (%) on the day of the visit	0.10 [0.10]	0.06 [0.04]	0.20* [0.10]	0.00 [0.01]	0.00 [0.01]
	Multigrade classroom for Grade 3 [Yes – 1, No – 0]	-5.40 [3.64]	-1.26 [1.67]	-4.77 [4.26]	-0.42* [0.18]	-0.55 [0.45]
	Classroom environment: Score [on an 8-point scale]	1.48 [1.54]	0.56 [0.72]	1.89 [1.67]	0.02 [0.11]	0.26 [0.22]
Constant		11.43 [16.79]	1.29 [7.98]	-12.79 [18.85]	-0.29 [0.89]	0.96 [2.28]
Observations		693	693	693	700	700
Adjusted R-squared		0.109	0.089	0.107	0.146	0.125

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 35. Factors influencing children's literacy skills – Endline (2025) – Comparison Schools - Results from OLS regressions

Covariates		Early Grade Literacy Skills Assessment Scores				
		Letter naming fluency (clpm)	**Nonword reading fluency (cnwpm)**	**Oral reading fluency (cwpm)**	**Reading comprehension (6-point scale)**	**Sentence Dictation (10-point scale)**
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-4.30* [2.05]	-1.83 [1.03]	-4.61* [2.14]	-0.38** [0.13]	-0.65* [0.28]
	Age (in years)	2.52** [0.90]	0.74 [0.40]	1.13 [0.90]	0.08 [0.07]	0.12 [0.16]
	Child speaks Hindi at home [Yes – 1, No – 0]	-2.88 [2.60]	-1.13 [1.36]	-5.72 [3.00]	-0.63*** [0.16]	-0.93 [0.54]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-2.58 [2.68]	0.46 [1.12]	0.74 [2.84]	0.18 [0.14]	0.28 [0.40]
	Child repeating the current grade [Yes – 1, No – 0]	1.19 [3.89]	-0.88 [2.07]	-4.01 [4.37]	-0.16 [0.24]	-0.20 [0.67]
	Absent at least once in last week [Yes – 1, No – 0]	-4.98* [2.50]	-0.81 [1.15]	-2.87 [2.61]	-0.03 [0.16]	-0.16 [0.41]
	Household standard of living: Score (on a 10-point scale)	0.79 [0.59]	0.31 [0.25]	0.34 [0.67]	0.07* [0.04]	0.12 [0.08]
Home literacy environment	Child was living with literate people in the last week [Yes – 1, No – 0]	5.70* [2.45]	2.07 [1.17]	5.42* [2.64]	0.27 [0.15]	0.43 [0.37]
	Someone read something to child at home in the last week [Yes – 1, No – 0]	5.96* [2.81]	1.76 [1.24]	2.65 [2.58]	0.09 [0.14]	0.44 [0.36]
	Someone told a story to the child at home in the last week [Yes – 1, No – 0]	-1.54 [2.14]	0.11 [0.95]	0.65 [2.46]	0.23 [0.14]	0.23 [0.32]
	Read for fun at least thrice in last week [Yes – 1, No – 0]	11.36*** [2.60]	5.48*** [1.14]	9.35** [2.89]	0.73*** [0.19]	0.99* [0.42]
	Reading materials/print items at home [Yes – 1, No – 0]	10.64** [3.53]	4.88** [1.59]	10.46** [3.75]	0.56** [0.19]	1.76** [0.54]
Teacher characteristics	Teacher's Gender [Female – 1, Male – 0]	-5.89* [2.69]	-2.71* [1.31]	-5.44 [3.30]	-0.33 [0.20]	-0.70 [0.50]
	Teacher's age (years)	0.24 [0.17]	0.11 [0.07]	0.22 [0.17]	0.01 [0.01]	0.02 [0.02]
	Teacher speaks Hindi at home [Yes – 1, No – 0]	0.57 [2.79]	0.79 [1.30]	3.31 [3.19]	0.34 [0.20]	-0.18 [0.48]

	Prior experience in teaching [Yes – 1, No – 0]	-3.83 [2.60]	-1.99 [1.31]	-6.77* [3.25]	-0.09 [0.22]	-0.41 [0.47]
School characteristics	Class size: 15 or less [Yes – 1, No – 0]	0.31 [2.78]	0.69 [1.29]	1.70 [3.15]	-0.02 [0.18]	0.01 [0.38]
	Attendance (%) on the day of the visit	0.08 [0.07]	0.03 [0.03]	0.08 [0.08]	0.00 [0.00]	0.02 [0.01]
	Multigrade classroom for Grade 3 [Yes – 1, No – 0]	-2.52 [2.86]	-1.45 [1.39]	-4.35 [3.41]	-0.36 [0.20]	-0.37 [0.48]
	Classroom environment: Score [on an 8-point scale]	-1.40 [1.09]	-0.17 [0.57]	-1.32 [1.25]	-0.00 [0.07]	0.07 [0.16]
Constant		0.19 [14.83]	-2.71 [7.42]	3.90 [18.33]	-0.75 [1.07]	-0.95 [2.53]
Observations		710	709	709	716	716
Adjusted R-squared		0.177	0.148	0.108	0.169	0.132
Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$						

Table 36. Factors influencing the probability of Zero Scores – Endline (2025) – Results from Logistic regressions

Covariates		Zero Scores: Early Grade Literacy Skills Assessment				
		Zero Score: Letter naming	**Zero Score: Nonword reading**	**Zero Score: Oral passage reading**	**Zero Score: reading comprehension**	**Zero Score Sentence Dictation**
Type	School Type [Program -1, Comparison – 0]	-0.50 [0.43]	-0.47 [0.29]	-0.65 [0.34]	-0.06 [0.20]	-0.52* [0.24]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	0.12 [0.21]	0.11 [0.18]	0.49** [0.15]	0.45*** [0.12]	0.31* [0.12]
	Age (in years)	-0.29* [0.12]	-0.20* [0.08]	-0.15 [0.10]	-0.09 [0.06]	-0.17* [0.08]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.87* [0.42]	-0.33 [0.30]	0.32 [0.24]	0.63*** [0.19]	0.30 [0.22]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.12 [0.34]	0.13 [0.22]	0.16 [0.23]	0.04 [0.15]	0.32 [0.18]
	Child repeating the current grade [Yes – 1, No – 0]	0.35 [0.36]	-0.01 [0.35]	-0.32 [0.33]	-0.04 [0.27]	-0.06 [0.25]
	Absent from school at least one day in the last week [Yes – 1, No – 0]	0.70* [0.28]	0.64** [0.22]	0.37 [0.23]	0.08 [0.18]	0.34 [0.19]
	Household standard of living: Score (on a 10-point scale)	-0.09 [0.07]	-0.04 [0.04]	-0.11* [0.04]	-0.12** [0.04]	-0.11** [0.04]
Home literacy environment	Child was living with literate people in the last week [Yes – 1, No – 0]	-0.53 [0.34]	-0.47 [0.26]	0.11 [0.20]	-0.14 [0.17]	-0.20 [0.19]
	Someone read something to child at home in the last week [Yes – 1, No – 0]	-0.45 [0.42]	-0.17 [0.29]	-0.40 [0.22]	-0.24 [0.16]	-0.41* [0.19]
	Someone told a story to the child at home in the last week [Yes – 1, No – 0]	-0.16 [0.28]	-0.41* [0.20]	-0.27 [0.19]	-0.14 [0.13]	-0.12 [0.17]
	Read for fun at least thrice in last week [Yes – 1, No – 0]	-1.45*** [0.44]	-1.18*** [0.30]	-0.72** [0.24]	-0.57*** [0.15]	-0.69*** [0.20]
	Reading materials/print items at home [Yes – 1, No – 0]	-1.03*** [0.24]	-1.16*** [0.21]	-0.98*** [0.25]	-0.75*** [0.20]	-0.93*** [0.22]

Teacher characteristics	Teacher's Gender [Female – 1, Male – 0]	0.23 [0.34]	0.20 [0.23]	0.56* [0.24]	0.37 [0.19]	0.49* [0.20]
	Teacher's age (years)	0.02 [0.02]	0.01 [0.01]	0.01 [0.02]	0.00 [0.01]	0.01 [0.01]
	Teacher speaks Hindi at home [Yes – 1, No – 0]	0.32 [0.31]	-0.01 [0.22]	0.04 [0.25]	-0.05 [0.20]	0.37 [0.22]
	Prior experience in teaching [Yes – 1, No – 0]	-0.20 [0.40]	0.15 [0.27]	0.03 [0.30]	-0.19 [0.24]	-0.06 [0.24]
School characteristics	Class size: 15 or less [Yes – 1, No – 0]	-0.50 [0.36]	-0.44 [0.25]	-0.49 [0.27]	-0.20 [0.21]	-0.10 [0.22]
	Attendance (%) on the day of the visit	-0.00 [0.01]	-0.00 [0.01]	-0.01 [0.01]	-0.00 [0.00]	-0.01 [0.01]
	Multigrade classroom for Grade 3 [Yes – 1, No – 0]	0.68 [0.37]	0.85*** [0.25]	0.73** [0.28]	0.57** [0.20]	0.27 [0.22]
	Classroom environment: Score [on an 8-point scale]	-0.07 [0.12]	-0.14 [0.09]	-0.15 [0.09]	-0.09 [0.08]	-0.09 [0.08]
Constant		1.92 [1.81]	1.95 [1.20]	2.20 [1.27]	2.36* [1.00]	2.35* [1.11]
Observations		1416	1416	1416	1416	1416
Adjusted R-squared		0.178	0.163	0.148	0.097	0.121

Standard errors in brackets. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 37. Factors influencing the probability of Achieving Reading Benchmarks – Endline (2025) – Results from Logistic regressions

Covariates		Reading Benchmarks: Early Grade Literacy Skills Assessment		
		Oral Reading Fluency: 35+ cwpm	**Oral Reading Fluency: 45+ cwpm**	**Reading comprehension: Score 80%+**
Type	School Type [Program -1, Comparison – 0]	0.75*** [0.21]	0.77*** [0.22]	-0.27 [0.44]
Children's demographic and economic features	Gender [Girl - 1, Boy - 0]	-0.37** [0.12]	-0.27* [0.12]	-0.55* [0.23]
	Age (in years)	0.02 [0.05]	0.07 [0.06]	0.17 [0.15]
	Child speaks Hindi at home [Yes – 1, No – 0]	-0.40* [0.19]	-0.64** [0.20]	-0.95* [0.40]
	Child attended pre-school/ECD [Yes – 1, No – 0]	-0.11 [0.14]	-0.18 [0.15]	0.22 [0.34]
	Child repeating the current grade [Yes – 1, No – 0]	0.35 [0.22]	0.04 [0.25]	-1.32* [0.64]
	Absent from school at least one day in the last week [Yes – 1, No – 0]	-0.13 [0.16]	0.11 [0.14]	0.12 [0.24]
	Household standard of living: Score (on a 10-point scale)	0.06 [0.04]	0.09* [0.05]	0.31*** [0.09]
Home literacy environment	Child was living with literate people in the last week [Yes – 1, No – 0]	0.37* [0.16]	0.37* [0.18]	0.26 [0.37]
	Someone read something to child at home in the last week [Yes – 1, No – 0]	0.31 [0.17]	0.08 [0.18]	0.41 [0.35]
	Someone told a story to the child at home in the last week [Yes – 1, No – 0]	-0.01 [0.13]	0.13 [0.15]	0.19 [0.28]
	Read for fun at least three times in last week [Yes – 1, No – 0]	0.63*** [0.16]	0.66*** [0.13]	0.99** [0.34]

	Reading materials/print items at home [Yes – 1, No – 0]	0.50* [0.23]	0.39 [0.23]	0.57 [0.45]
Teacher characteristics	Teacher's Gender [Female – 1, Male – 0]	-0.51** [0.17]	-0.52** [0.19]	-0.56 [0.36]
	Teacher's age (years)	0.00 [0.01]	-0.00 [0.01]	0.03* [0.01]
	Teacher speaks Hindi at home [Yes – 1, No – 0]	0.08 [0.18]	0.03 [0.18]	0.13 [0.30]
	Prior experience in teaching [Yes – 1, No – 0]	-0.01 [0.21]	0.12 [0.22]	-0.93* [0.38]
School characteristics	Class size: 15 or less [Yes – 1, No – 0]	0.03 [0.19]	-0.23 [0.20]	-0.05 [0.31]
	Attendance (%) on the day of the visit	0.01 [0.00]	0.01 [0.00]	-0.01 [0.01]
	Multigrade classroom for Grade 3 [Yes – 1, No – 0]	-0.33 [0.18]	-0.25 [0.19]	-0.42 [0.33]
	Classroom environment: Score [on an 8-point scale]	-0.04 [0.08]	-0.04 [0.08]	-0.07 [0.15]
Constant		-1.53 [0.95]	-2.65* [1.09]	-7.22*** [1.67]
Observations		1416	1416	1416
Adjusted R-squared		0.096	0.097	0.140

D. Evaluation Limitations

Alongside the results of the two-year impact evaluation, Room to Read wants to highlight the main limitations of the study.

1. **Compatibility**: A key limitation is that the cross-sectional design hinges on the assumption that program and comparison schools were equivalent. Although the comparison district and schools were carefully selected and some baseline covariates were checked, significant differences in literacy scores—and in some school, teacher, and child characteristics—emerged at both baseline and endline. Because some of these differences could not be controlled for in the impact regressions (they were omitted at baseline), unobserved group disparities may still bias the results.
2. **Children Sampling Bias**: The sample of children in each school was drawn from those present on the day of assessment rather than the entire enrolment list for Grade 3. This could result in a bias, assessing only children who regularly attend school and potentially overlooking those who are struggling or frequently absent.
3. **Informed Visit**: Authorities at selected schools were informed in advance about the visits, which might have led to improved performance on the assessment day due to heightened awareness. This advance notice could create a Hawthorne effect, where participants alter their behaviour in response to being observed or interviewed.
4. **Response Biases**: Data on the background features of the teachers and children were collected through one-to-one interviews and assessments. It is possible that self-reported data on these indicators may have suffered from response biases due to inaccurate recollection or intent to provide desired responses. However, the data analyses did not reveal any systematic response bias across the program and comparison groups that may have influenced the results.

www.roomtoreadindia.org

C-21, Qutub Institutional Area, Block C, New Delhi - 110016.