Haryana Early Literacy Development Impact Bond

Impact Evaluation Report

IndusInd Bank | CENTRAL SQUARE FOUNDATION | SBI FOUNDATION Service Beyond Banking | SBI Capital Markets Limited | LLF Language and Learning foundation Strong Foundation, Stronger Future

CONTENTS

1. Executive summary

The Early Literacy intervention in Haryana is the first-ever Development Impact Bond (DIB) project in India funded by CSR partners – IndusInd Bank and SBI Capital Markets Limited through SBI Foundation, with Central Square Foundation being the guarantor, and Social Finance India as DIB convener, based on an outcome-based finance model. The program exclusively focuses on early literacy and is being implemented by Language and Learning Foundation (LLF).

Educational Initiatives (Ei) was selected as a third-party independent outcome evaluator in order to:

(i) Set targets for the project, and

(ii) Assess the impact of the project.

The project spans from the start of Grade 1 in 2020 to the end of Grade 2 in 2022 and is being implemented in 7 districts in Haryana.

In August 2021, Ei conducted the baseline assessment with students at the start of Grade 1. Given that baseline could not be conducted in 2020 due to COVID, incoming students in Grade 1 in 2021 were tested to establish baseline instead, with the assumption that the variation between incoming grade 1 student performance between one cohort and the next would be low. Control districts were identified, where no intervention was taking place, to serve as a comparison group. In April 2022, Ei conducted the Endline assessment with students (who started school in 2020) at the end of Grade 2.

The results of the Early Literacy DIB program are highly encouraging. The intervention went live just as the pandemic hit and the schools were shut down, so the partners had to quickly pivot to a new implementation model in order to ensure students were still learning at home. The learning gains that have been achieved in the program are even more impressive considering that the cohort of students had less than one year of in-school instruction and must have had significant learning losses as has been documented by many studies.

The summary of findings of the impact assessment is given below -

- The intervention was able to achieve learning gains of an additional 1.61 Equivalent Years of Schooling over business-as-usual schooling (using oral reading fluency as the indicator for acquisition of foundational literacy skills).
- In terms of effect size, the intervention achieved learning gains ranging from 0.65 SD to 1.07 SD over different sub-tasks related to literacy skills (excluding pre-school skills ceiling effects).
- The effects of the program were compared to similar EGR programs worldwide, as they are presented in the World Bank Education Global Practice Group's 2018 working paper "How Effective Are EGR Interventions? A Review of the Evidence". The program impact is larger than all other listed programs in terms of letter fluency, while is the top 3 programs in terms of ORF and RC.
- The results were also compared with Indian Early Grade Reading improvement programs as presented in USAID's 2019 report "Analysis of Early Grade Reading Assessment in India" and used a similar approach to measure efficacy of their literacy intervention. The DIB program has performed significantly better than the rest of the programs in India in terms of achieving gains in Oral Reading Fluency and Reading Comprehension.
- When compared with the Global Minimum Proficiency (GMP) equivalent standard of Oral Reading Fluency for Hindi (as defined in NCERT's Foundational Learning Study 2022), it is seen that 65 percent of students in the intervention group are at Meets or Exceeds GMP standard of being able to read more than 35 words per minute at the end of Grade 2.
- As expected, tasks like oral vocabulary have ceiling effects which shows children's familiarity with their environment.
- Struggling learners, captured by the percentage of students scoring zero has decreased considerably over the duration of the program with absolute decline of 56 to 95% and difference-in-difference decline of 11% to 26% over control, across subtasks.
- Overall, the girls are performing better than the boys in both intervention and control schools in 12 out of 14 sub-tasks.
- It was noted that in Baseline, less than 25% of children were able to read words and non-words, likely owing to exposure and home reading environment. In the Endline, more than 80% of the students were able to read words, non-words and sentences.

- As expected, word accuracy is lower[1] than letter accuracy. This can be because of low letter fluency scores. Impairment in letter fluency causes struggle in decoding skills as children do not have freed up working memory to use to decode words.
- As expected, word fluency is much lower than letter fluency. The rise is exponential though from BL to EL. Understandably[2], when children are unable to identify letters fluently, reading words fluently will be a challenge.

Table 1: Overall performance in baseline and Endline

Sub-task #	Sub-tasks	Total	Baseline	Baseline	Endline	Endline	DiD	Effect Size	EYOS	Goal	DiD as Multiple of Goal
			Control	Intervention	Control	Intervention					
1	Listening comprehension	4	2.9	2.9	3.0	3.3	0.4	0.35	8.46	1	-
2	Oral vocabulary	10	8.8	9.1	9.5	9.9	0.1	0.08	1.32	0	-
3	Initial sound identification	10	2.2	2.6	6.7	9.2	2.0	0.62	1.59	2	1
4a	Letter accuracy	15	4.1	3.3	11.0	13.4	3.2	0.77	1.67	1	3
4b	Letter fluency (cpm)	100	11.2	8.7	46.2	66.3	22.5	1.07	1.83	7	3
5a	Word accuracy	15	2.4	2.2	9.4	12.8	3.6	0.81	1.75	1	4
5b	Word fluency (cpm)	50	3.3	3.1	20.8	28.4	7.9	0.68	1.56	4	2
6	Non-word fluency (cpm)	50	3.3	2.6	20.9	27.3	7.1	0.64	1.56	5	1
7	Oral reading fluency (cpm)	54	3.7	2.9	30.3	42.4	12.9	0.67	1.61	8	2
8	Reading comprehension	43	5.0	4.6	25.2	36.7	11.9	0.88	1.82	1	12
10a	Letter writing	10	0.0	0.0	6.3	8.0	1.7	0.77	1.50	1	2
10b	Word writing	10	0.0	0.0	4.9	7.1	2.3	0.93	1.67	1	2
10c	Sentence Writing	5	0.0	0.0	1.6	2.6	1.0	0.68	1.61	2	-

[1] **LiRIL**: Over 62% of teachers say their end-of-Grade 1 expectations from students are the knowledge of moolaksharas, with or without swarchinhas. By the end of Grade 2, teachers expect their students to know the sound-symbol correspondence of all the aksharas and swarachinhas in the barakhadi and to be able to read with a sufficient degree of accuracy.

[2] **LiRIL**: Even by the end of Grade 3, and despite considerable instructional time spent on these, many students had not acquired fluency with the scripts. Swarachinhas and jodaksharas, in particular, presented significant challenges. Our findings echo those of Nag (2007) who has flagged the same concerns.

2. Introduction and Background

The Early Literacy intervention in Haryana is the first-ever Development Impact Bond (DIB) project in India funded by CSR partners – IndusInd Bank and SBI Capital Markets Limited through SBI Foundation, with Central Square Foundation being the guarantor, and Social Finance India as DIB convener, based on an outcome-based finance model. The program exclusively focusses on early literacy. This DIB scaled up the existing program of Language and Learning Foundation (LLF) in the state of Haryana. Educational Initiatives (Ei) played the role of an outcome evaluator in the DIB. The key features of the DIB and the entire program are summarized in the Appendix 1.

LLF is working with various Government schools in the state of Haryana starting from April 2020. The student learning outcomes in this intervention were measured at the beginning (start of Grade 1) and then at the end of the second year (end of Grade 2). The pay-outs to the risk investors would be based on the net learning gains achieved by intervention schools over control group schools after a two-year intervention.

2.1. Objectives

The objective of the outcome evaluation agency (Educational Initiatives) was to (i) set targets for learning gains, and (ii) measure the change in learning levels from baseline to Endline, for both the program implemented by LLF in Haryana and a comparison group (which here represents the status quo in schools in Haryana).

2.2. Methodology

To measure the impact of the project, baseline and Endline assessment data needs to be gathered for intervention and control districts. Baseline was conducted in 2021 instead of 2020 (when the program started) due to COVID restrictions, making data collection on the ground impossible. Instead, it was decided that the incoming students in Grade 1 from 2021-23 cohort will be used to set up baseline, with the assumption that there wouldn't be much variation in the existing learning levels of students entering grade 1 across a year. Rationale for cohort selection for baseline and Endline is provided in the table below.

Table 2: Cohort selection and rationale

Type of Assessment	Grade	Cohort	Rationale
Baseline	Grade 1 (start)	2021-2023	Students entering Grade 1 in Aug-2021 were considered for baseline assessment. This cohort was selected over previous year's cohort due to COVID-related travel restrictions in early 2020.
Endline	Grade 2 (end)	2020-2022	Students who have been part of LLF's intervention since April-2020 were evaluated in April-2022 (while they are at the end of Grade 2) to assess impact of the 2-year intervention.

Student enrolment for Grade 1 students in Haryana was completed between July-August-2021 and therefore, Ei conducted the baseline assessment in September-2021. By April 2022, students had already started enrolling in Grade 3, therefore Ei conducted the Endline assessment in the first 3 weeks of April 2022. Preparation for the Endline study included the following:

a) Tool finalisation: The Foundational Literacy Assessment Tool covers 9 skills (through 14 sub-tasks) that cover all four cognitive domains of listening, speaking, reading and writing (LSRW). The tools were designed by Ei for Grades 1 & 2 and have around 30% overlap in items between grades. Some of the sub-tasks are designed to test accuracy (the correctness with which students can answer irrespective of the time taken) while some are designed to test fluency (correct answers per minute).

Based on the findings and observations of the baseline assessment, minor changes were made to the assessment tool and one more reading comprehension task was added.

List of all sub-tasks are provided in the table below. The sub-tasks related to pay-outs are highlighted in green.

Table 3: List of sub-tasks

S.No	Sub-task	Metric
1	Listening comprehension - accuracy	Questions answered correctly
2	Oral Vocabulary and Picture Description- Accuracy	Pictures identified correctly
3	Initial Sound Identification- Accuracy	Sounds identified correctly
4a	Letter Naming - Accuracy	Letters read correctly
4b	Letter Naming - Fluency	Letters read correctly per minute
5a	Familiar Word Reading- Accuracy	Words read correctly
5b	Familiar Word Reading- Fluency	Words read correctly per minute
6	Non- word Reading- Fluency	Words read correctly per minute
7	Oral Reading- Fluency	Words read correctly in a minute
8	Reading Comprehension Accuracy	Words read correctly + questions answered correctly
9	Reading Comprehension Accuracy	Words read correctly + questions answered correctly
10a	Dictation: Letters, (Accuracy)	Letters written correctly
10b	Dictation: Words	Words written correctly
10c	Dictation: Sentences	Sentences written correctly

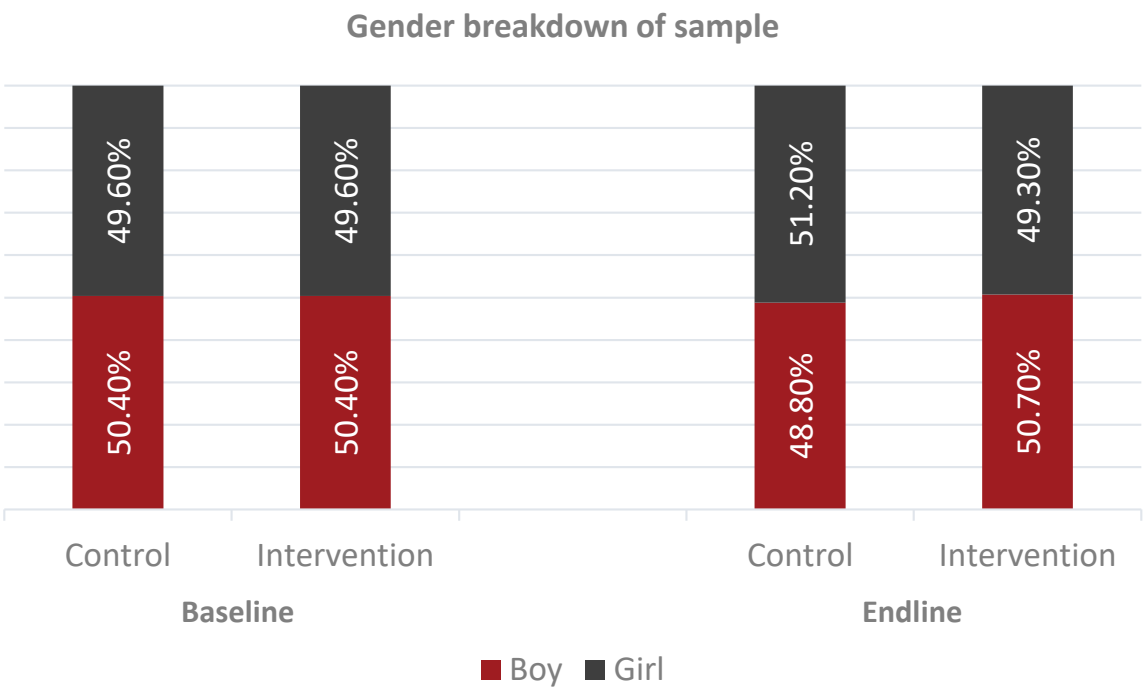
b) Government permission: Ei with support from LLF received government permission and student list of Grade 1 students in 7 intervention districts and 4 control[3] districts. For this project, control districts were selected as districts with similar composite scores (HDI, gender index, education index and ASER-2018 results) to intervention districts.

c) Sampling: Ei used school-level sampling and assessed 1531 students for intervention districts and 983 students for control districts. PPS (probability proportional to size) was used to account for variations in number of students per district. Additional buffer was added to account for absenteeism in the endline. Table summarising total number of students assessed is provided below. Minimum numbers required for the endline assessment were met across districts. index and ASER-2018 results) to intervention districts.

Table 4: Total number of students assessed

Sample per district		Baseline		Endline	
		# of students	# of students	# of students	# of students
Control	Bhiwani	331	893	395	983
	Kaithal	326		321	
	Panipat	236		267	
Intervention	Ambala	142	1413	176	1531
	Fatehabad	203		190	
	Hisar	237		298	
	Jind	193		190	
	Kurukshetra	158		224	
	Sirsa	282		219	
	Yamunanagar	198		234	

[3] Though evaluations were conducted in Sonipat for both BL and EL, it was not considered for analysis due to being an outlier in Baseline performance.



d) Evaluator selection and training: Ei recruited a minimum of 3-4 evaluators per district (44 evaluators in total) for test conduction. Selection criteria for the evaluators were personnel with past experience in large-scale data collection and/or past experience in teaching lower grades. A 2-day training was conducted for the evaluators in April 2022.

Objectives of the training were:

i). Evaluators understand each test item

ii). Evaluators understand how to use tool and administer the test

iii). Evaluators test the tool (via role-plays with other evaluators and with students)

iv). Evaluators understand how to find students and schools selected for the study

e) Field audits: Ei team conducted in-person audit in 2 districts and the state coordinator conducted in-person audit in 4 districts. Ei team observed how the evaluators were interacting with school principals, parents and students, and how they were scoring the students. Observations were shared as additional SOPs with the other evaluators. Additionally, one evaluator per district was selected as the coordinator and was trained on conducting audits. The coordinators selected random evaluators on a weekly basis and conducted quality checks.

f) Data cleaning: Once the data collection was completed, the quality check was conducted on the data. Only 1 student data was removed from the data set due to missing information or illogical information.

3. Impact Evaluation

This evaluation examined the program's effectiveness in terms of improving learning outcomes. It includes an analysis of impact on EGRA scores at sub-task level, an analysis of zero scores across sub-tasks to find critical areas of improvement, a comparison with India-specific and international studies to act as benchmark and to understand relative standing, and comparison across similar schools grouped based on their initial achievement.

3.1. Sub-Task wise Impact Calculation

The intervention schools showed significant gains across time and with respect to control group in almost all-subtasks (Learning Comprehension and Oral Vocabulary showed low gain due to ceiling effect, as children had already been achieving close to perfect scores). All six sub-tasks linked to payout (highlighted in green), met and exceeded their goals by a wide margin.

The program was able to achieve learning gains to the tune of 1.61 Equivalent years of Schooling (EYOS) over business as usual, based on Oral Reading Fluency, and 1.81 EYOS with respect to Reading Comprehension.

In terms of effect size, the difference-in-difference gains across sub-tasks ranged from 0.64 to 1.07 SD (excluding ceiling sub-tasks). How these effect sizes compare with other Indian and International studies is shown in the subsequent section.

How to read Table

- Table 4 shows the difference-in-difference score of the intervention Endline and the goal set for each sub-task
- Pay-out related sub-tasks are highlighted in green.
- Listening comprehension scores saw a ceiling effect at a total score of 4, hence it is one of the sub-tasks where the goal is not met.

Table 5: Goal settings for endline

Sub-task #	Sub-tasks	Total	Baseline	Baseline	Endline	Endline	DiD	Effect Size[4]	EYOS[5]	Goal	DiD as Multiple of Goal
			Control	Intervention	Control	Intervention					
1	Listening comprehension	4	2.9	2.9	3.0	3.3	0.4	0.35	8.46	1	-
2	Oral vocabulary	10	8.8	9.1	9.5	9.9	0.1	0.08	1.32	0	-
3	Initial sound identification	10	2.2	2.6	6.7	9.2	2.0	0.62	1.59	2	1
4a	Letter accuracy	15	4.1	3.3	11.0	13.4	3.2	0.77	1.67	1	3
4b	Letter fluency (cpm)	100	11.2	8.7	46.2	66.3	22.5	1.07	1.83	7	3
5a	Word accuracy	15	2.4	2.2	9.4	12.8	3.6	0.81	1.75	1	4
5b	Word fluency (cpm)	50	3.3	3.1	20.8	28.4	7.9	0.68	1.56	4	2
6	Non-word fluency (cpm)	50	3.3	2.6	20.9	27.3	7.1	0.64	1.56	5	1
7	Oral reading fluency (cpm)	54	3.7	2.9	30.3	42.4	12.9	0.67	1.61	8	2
8	Reading comprehension	43	5.0	4.6	25.2	36.7	11.9	0.88	1.82	1	12
10a	Letter writing	10	0.0	0.0	6.3	8.0	1.7	0.77	1.50	1	2
10b	Word writing	10	0.0	0.0	4.9	7.1	2.3	0.93	1.67	1	2
10c	Sentence Writing	5	0.0	0.0	1.6	2.6	1.0	0.68	1.61	2	-

3.2. Comparison with Indian and International Interventions

Comparison with International Studies

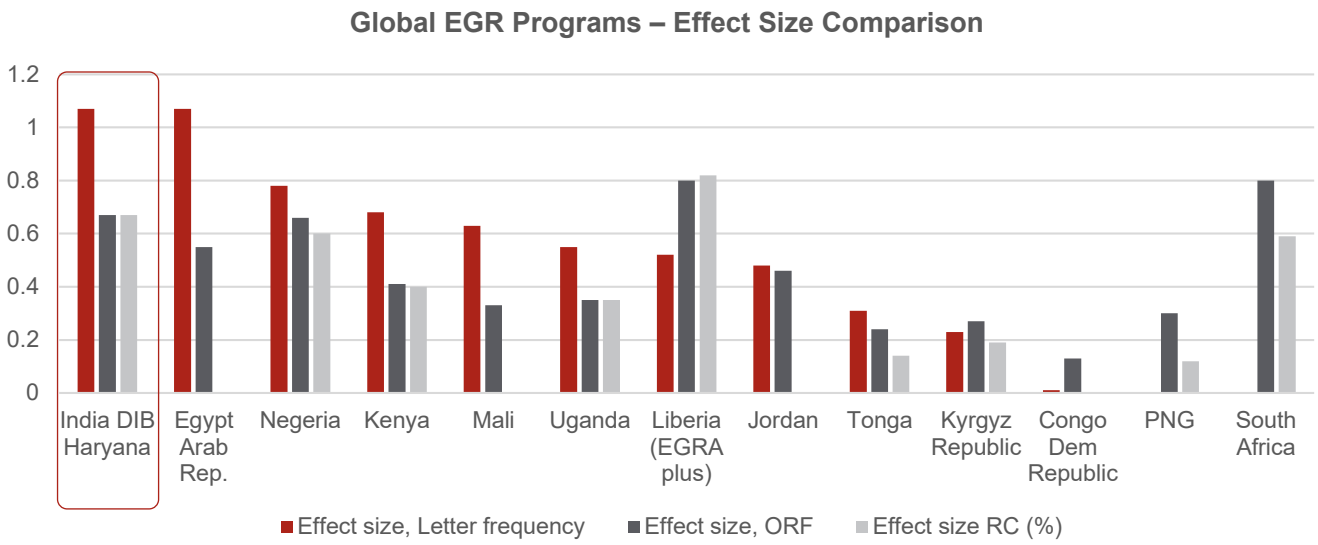
The effects of the program were compared to similar EGR programs worldwide, as they are presented in the World Bank Education Global Practice Group's 2018 working paper "How Effective Are EGR Interventions? A Review of the Evidence"[6] (data table in Appendix).

Figure below presents the endline program effects, benchmarked against various EGR programs included in the World Bank publication. The program impact is larger than all other listed programs in terms of letter fluency, while is the top 3 programs in terms of ORF and RC.

[4] Effect size is a standard metric of expressing the difference in performance between two groups. It indicates the standardised difference between means of two different groups. It is used internationally, and is well accepted in research and literature.

[5] Equivalent Years of Schooling (EYOS) has been calculated by taking the effect size gain of the intervention group between baseline and endline and dividing it by the control group's effect size gain in the same period. The assumption is that the control group represents business-as-usual learning gains (Evans and Yuan (2019), "Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms"

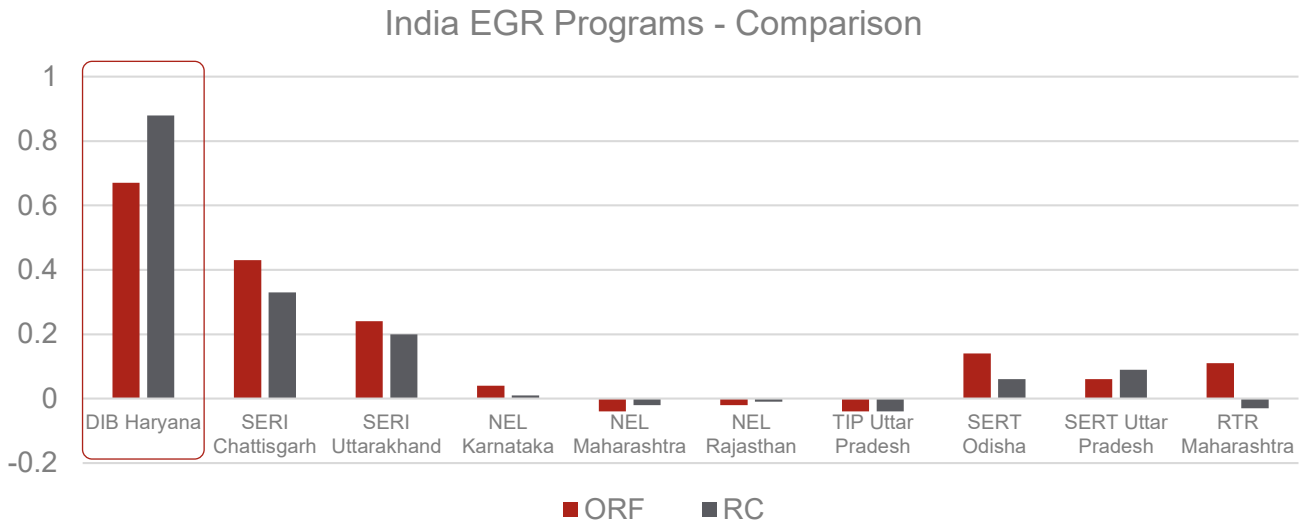
[6] Graham, Jimmy and Sean Kelly. January 2018. http://documents.worldbank.org/curated/en/289341514995676575/pdf/WPS8292.pdf



Comparison with Indian Programs

The results were also compared with the following programs which are India-specific (as they are presented in USAID's 2019 report "Analysis of Early Grade Reading Assessment in India"[7] and used a similar approach to measure efficacy of their literacy intervention (data table in Appendix).

The DIB program has performed significantly better than the rest of the programs in India in terms of achieving gains in Oral Reading Fluency and Reading Comprehension.



[7] RTI and Pratham, April 2019. https://earlygradereadingbarometer.org/pdf/India_2018_Impact_Assessement.pdf

Comparison with Global Minimum Proficiency Standards (as defined in NCERT's FLS)

When compared with the Global minimum proficiency equivalent standard of Oral Reading Fluency for Hindi (as defined in NCERT's Foundational Learning Study 2022[8]), it is seen that 65 percent of students in the intervention group are at Meets or Exceeds Global Minimum proficiency standards of being able to read more than 35 words per minute at the end of Grade 2. This translates to a difference-in-difference gain of 23% relative to control between baseline and endline. Per the FLS study, this is comparable to where students are at the end of Grade 3 in the rest of the state (65% students at end of Grade 3 met or exceeded the benchmark).

Table 6: Proficiency benchark of ORF

Oral Reading Fluency (count of words per minute)						
Global Minimum Proficiency		Baseline		Endline		Difference in difference
Standard	Range	Control	Intervention	Control	Intervention	
Below Partially Meets	0-16	91%	93%	38%	14%	-27%
Partially Meets	17-34	5%	4%	19%	20%	3%
Meets	35-54	4%	3%	32%	46%	15%
Exceeds	55-	0%	0%	11%	19%	8%

Please note that for all these comparisons, though EGRA tools have been used, there would be some variation as the specific passages in ORF and RC would vary from tool to tool.

3.3. Low Performance Analysis: Zero Scores

Zero scores are commonly used indicators for tracking the prevalence and progress of learners who are struggling and require remediation. The program has been able to significantly reduce the percentage of zero scores across subtasks and has been able to do it better than the control group by a margin of 10-26%.

Children have shown tremendous improvement in reading with meaning, with percentage of children not being able to read a single word going from 81% at baseline to 5% at endline (oral reading fluency). Only 7% children weren't able to answer a single question based on reading a passage compared to 78% at the start of grade 1.

[8] https://ncert.nic.in/pdf/FLS/fls_sr/haryana.pdf

Table 7: Percentage of students scoring zero per sub-task

% Zero Scorers						
#	Sub-tasks	Baseline	Baseline	Endline	Endline	
		Control	Intervention	Control	Intervention	Difference-in-Difference
1	Listening comprehension	7%	10%	8%	1%	10%
2	Oral vocabulary	3%	3%	1%	0%	1%
3	Initial sound identification	61%	60%	16%	4%	11%
4a	Letter accuracy	54%	62%	9%	2%	16%
4b	Letter fluency (cpm)	51%	56%	8%	1%	13%
5a	Word accuracy	69%	68%	19%	3%	15%
5b	Word fluency (cpm)	78%	77%	20%	3%	16%
6	Non-word fluency (cpm)	75%	76%	21%	4%	17%
7	Oral reading fluency (cpm)	80%	81%	24%	5%	20%
8	Reading comprehension	79%	78%	26%	7%	18%
10a	Letter writing	98%	99%	17%	4%	13%
10b	Word writing	100%	100%	31%	9%	22%
10c	Sentence Writing	100%	100%	53%	28%	26%

3.4. Composite Scores for School matching

In order to compare schools based on their initial level of performance and look at how the program affected incoming students with different levels of skills, we adopted the following approach. We bucketed the schools into low-to-high performing ones based on baseline data.

We then created a composite score (purely for comparison) based on the following logic:

i. Oral vocabulary, listening comprehension tasks have been excluded from composite scores as these are pre-school skills and ceiling effect was noted in both control and intervention groups (as both groups had similar scores and high scores)

ii. Dictation tasks have been excluded from composite score as both groups scored close to 0 at BL.

iii. Only reading tasks have been included in composite scores

Table 8: Weightage of sub-taskd for composite score buckets

Composite Score Weightage	BL	EL
Letter accuracy	20%	20%
Letter fluency	80%	30%
Word accuracy	0%	15%
Word fluency	0%	15%
Non-word fluency	0%	10%
Oral reading fluency	0%	5%
Reading comprehension	0%	5%

iv. Letter reading (accuracy and fluency) were given higher weightage at baseline because:

   a. Letter reading ability at the start of Grade 1 is a strong influencer in reading abilities in later grades

   b. Letter reading (accuracy and fluency) had the most significant degree of difference between intervention and control groups.

Once sub-tasks for inclusion and exclusion were identified, then all subtasks which were to be included were normalised {(student score-min score)/ (max score - min score)}. Then composite score per student was calculated, post which average composite score per school was calculated.

Once average composite score per school was calculated, then schools with similar composite scores between control and intervention group were listed under various buckets. For ease of comparison, buckets were then combined into 5 buckets to reach the following result.

It is seen that the highest gain in composite score was obtained for schools which had the most number of struggling students as intake in grade 1, implying that the program was able to make significant gains where it was most needed.

Table 9: Performance by composite score buckets

Bucket	Schools Intervention	Schools Control	Intervention Baseline	Intervention Endline	Intervention Gain	Control Baseline	Control Endline	Control Gain	DiD (out of 10)
0-1	48	15	0.36	4.88	4.52	0.37	3.52	3.15	1.37
1-2	19	15	1.29	5.06	3.77	1.10	3.60	2.50	1.27
2-4	19	16	2.52	5.08	2.56	2.36	4.31	1.95	0.61
4-6	2	4	5.43	5.50	0.07	4.27	4.21	-0.06	0.14
6-8	2	1	7.77	5.25	-2.52	5.97	5.69	-0.27	-2.24

4. Detailed Results for each Sub-Task

4.1. Overall Performance

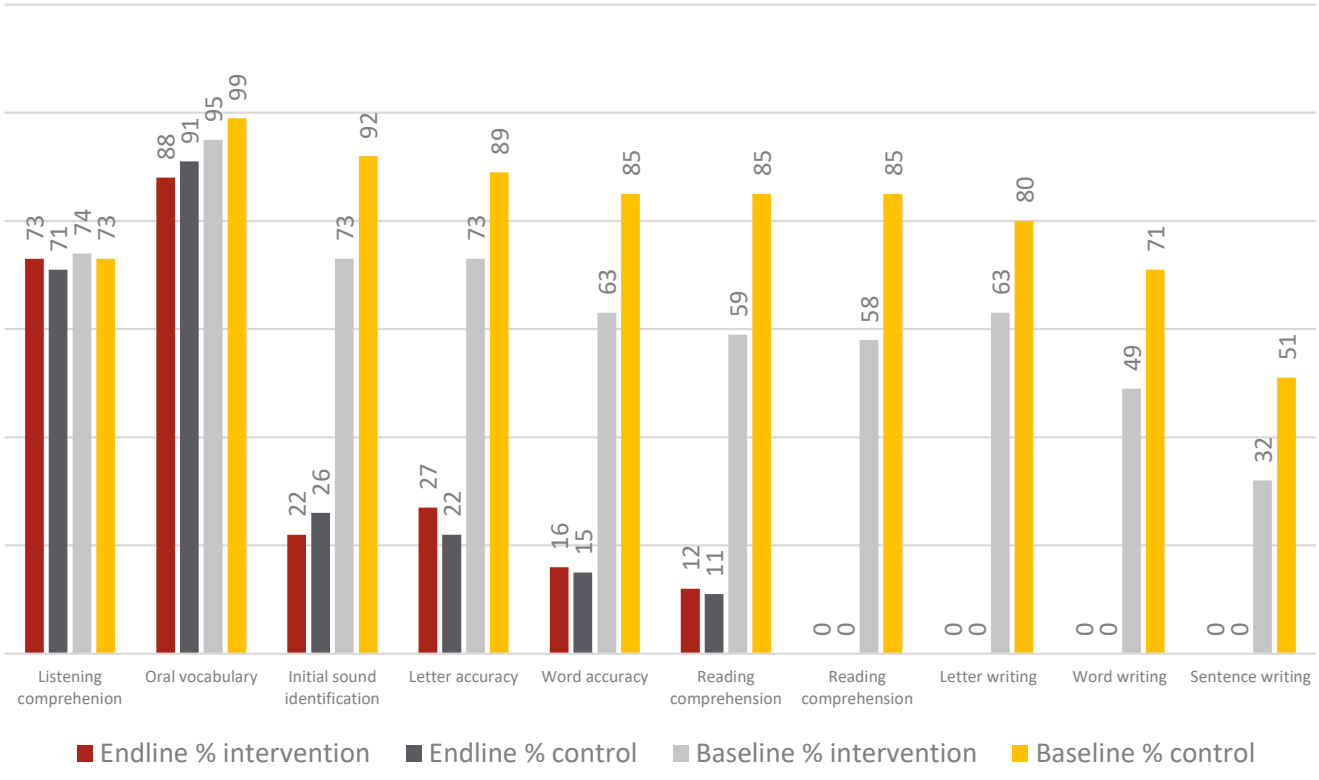
How to read table:

- This table includes **average raw** scores.
- Average score per sub-task in intervention and control group is provided in the table for both baseline and Endline. These can be compared with the total maximum score, which is the total number of questions asked per sub-task. For instance, in task 4a, letter accuracy, students in intervention group, on average, read 3.3 letters out of the 15 letters provided to them.
- Accuracy tasks had no timers, students were allowed to take as much time as required to read the letters.
- Fluency tasks were timed tests and did not have a maximum score as students were expected to read as many letters as possible within a minute.
- Time taken to read letters or words in fluency tasks were also noted. This was used to calculate count per minute. For instance, the ORF task (oral reading fluency) had a passage with 43 words. If children read 35 of 43 words correctly, in say, 40 seconds, then ORF was calculated as (35 words/40 seconds)*60 seconds, which gave us the count of words read correctly in a minute. All fluency tasks were calculated similarly.
- Reading comprehension task had two parts, first, number of words read correctly without a timer and number of questions answered correctly. The questions were based on the passage, and the passage was available to the child for reference.

एक चींटी और एक कबूतर दोस्त थे। एक दिन चींटी नदी के किनारे खेल रही थी। तभी वह फिसल कर नदी में गिर गई।कबूतर ने उसे बचाने के लिए एक पत्ती नदी में डाली। चींटी उस पत्ती को पकड़ कर बच गयी।

Table 10: Overall performance - raw scores

Sub Task #	Sub-tasks	Baseline		Endline		Total Score
		Control	Intervention	Control	Intervention	
1	Listening comprehension	2.9	2.9	3.0	3.3	4.0
2	Oral vocabulary	8.8	9.1	9.5	9.9	10.0
3	Initial sound identification	2.2	2.6	6.7	9.2	10.0
4a	Letter accuracy	4.1	3.3	11.0	13.4	15.0
4b	Letter fluency (cpm)	11.2	8.7	46.2	66.3	-
5a	Word accuracy	2.4	2.2	9.4	12.8	15.0
5b	Word fluency (cpm)	3.3	3.1	20.8	28.4	-
6	Non-word fluency (cpm)	3.3	2.6	20.9	27.3	-
7	Oral reading fluency (cpm)	3.7	2.9	30.3	42.4	-
8	Reading comprehension	5.0	4.6	25.2	36.7	43.0
9	Reading comprehension	0.0	0.0	40.2	59.0	69.0
10a	Letter writing	0.0	0.0	6.3	8.0	10.0
10b	Word writing	0.0	0.0	4.9	7.1	10.0
10c	Sentence writing	0.0	0.0	1.6	2.6	5.0



How to read Table

- This table shows the average %age on a sub-task compared to the total score.
- This table does not include performance on fluency tasks as these sub-tasks are calculated on the basis of letters/ words read per minute not on the basis of total score.
- Sentence writing was not tested in Baseline as students were not even able to write letters and words on dictation.

Table 11: Performance on non-fluency tasks

Sub-tasks	Baseline Average %age		Endline Average %age	
	Control	Intervention	Control	Intervention
Listening comprehension	73%	71%	74%	83%
Oral vocabulary	90%	91%	95%	99%
Initial sound identification	25%	26%	67%	92%
Letter accuracy	31%	22%	73%	89%
Word accuracy	19%	15%	63%	85%
Reading comprehension	15%	11%	59%	85%
Letter writing	0%	0%	63%	80%
Word writing	0%	0%	49%	71%
Sentence writing	-	-	32%	51%

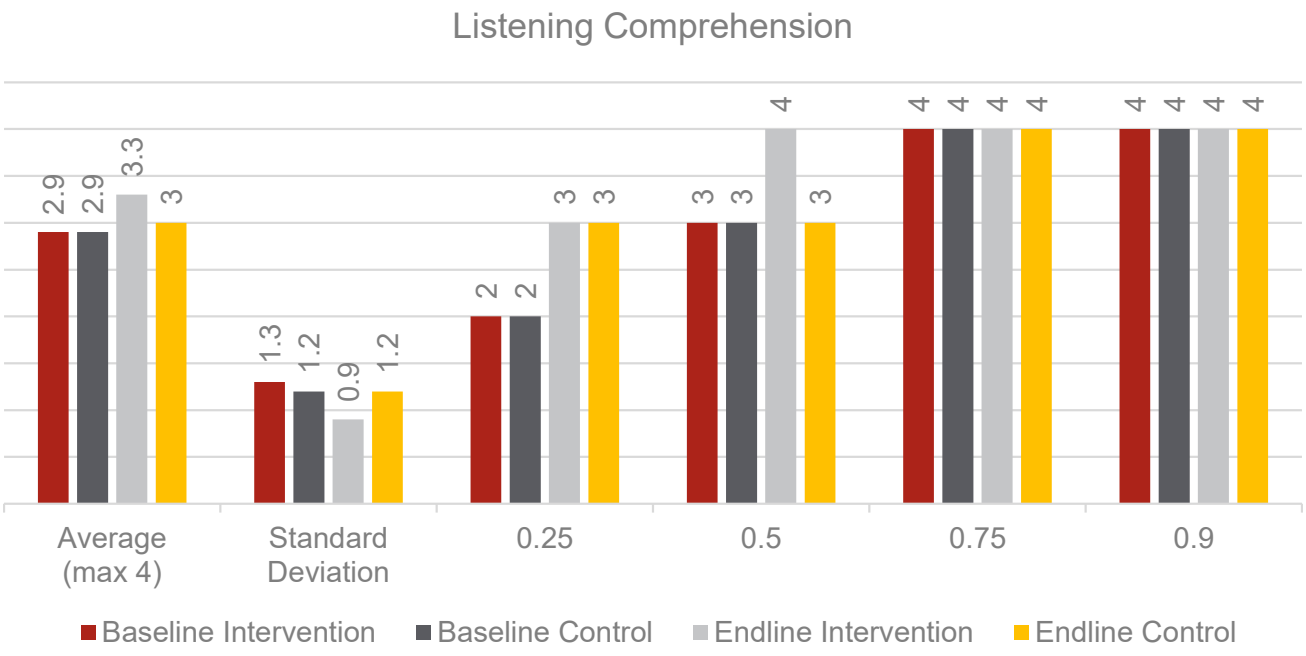
Findings:

- In the initial sub-tasks such as listening comprehension, oral vocabulary, initial sound identification, students in intervention group performed better than the control group.
- Students overall did quite well in listening comprehension and oral vocabulary both in baseline and Endline, while their performance had improved significantly in sound identification, letter, word accuracy and fluency, ORF and reading comprehension in the intervention group.
- Difference between letter and word fluency (from subtask 4 and 5) shows expected gap as well, given the increase in complexity and decoding approach to teaching reading skills usually followed in schools.

4.2. Performance by Sub-Task

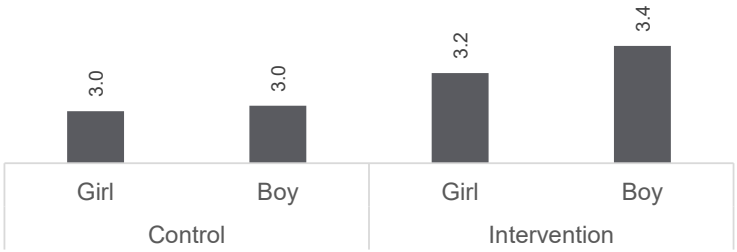
How to read the data - Standard deviation signifies the spread from the mean. Score at 25th percentile is the score of the student who is at the 25th percentile mark, which means 25 percent of the children who answered this question were below the given student in rank based on their response. For instance, in the table below, the student at 25th percentile in intervention group had a score of 2 out of 4 in listening comprehension sub-task.

4.2.1. Performance per sub-task: Listening comprehension

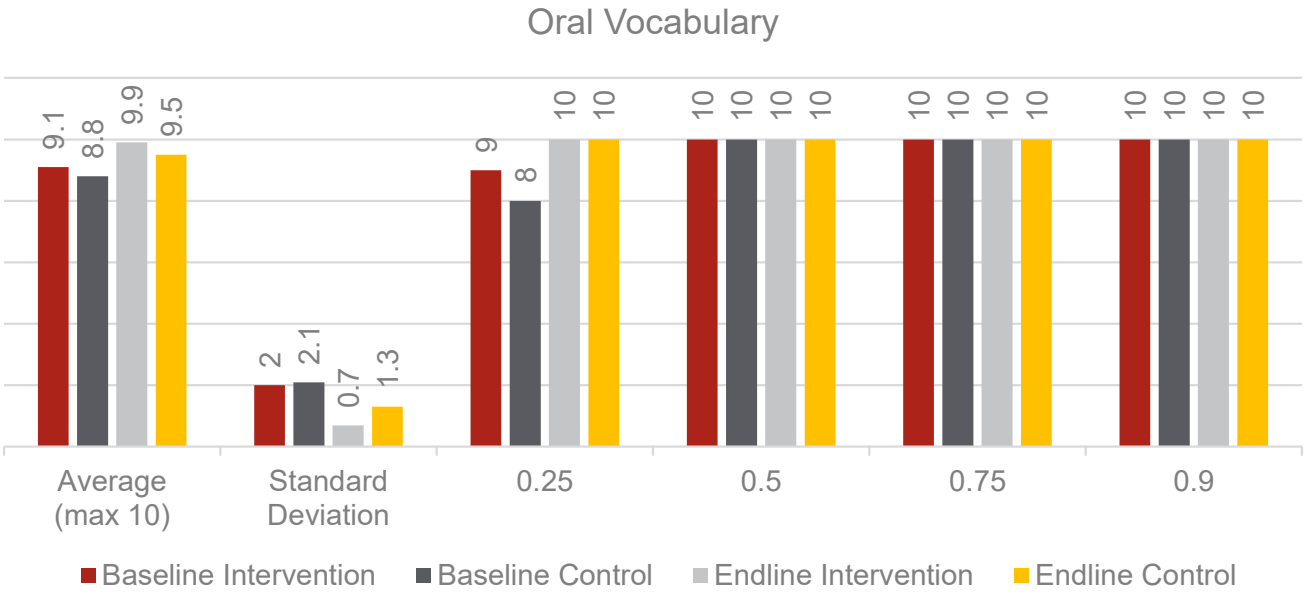


- The objective of this sub-task is to check for the student's ability to comprehend the story on the basis of vocabulary knowledge.
- Ei's past data has shown high performance among Grade 1 students in this task as it is a pre-school skill.
- Not a substantial growth can be seen in this task: 72% in baseline to 79% in Endline. Here students only had to listen to a 40 to 60 words passage, which could be heard twice. This task ideally should be showing ceiling effect, given India being an oral tradition.
- This also shows that children do not often engage with the text they are listening to. Another reason could also be students' inattentiveness while listening to the passage or evaluators taking a long pause between the passage and the questions.

- Gender-wise comparison on this sub-task shows boys performing better that girls in the intervention schools with only 0.2 difference in score. Boys and girls are performing similarly in the control schools.



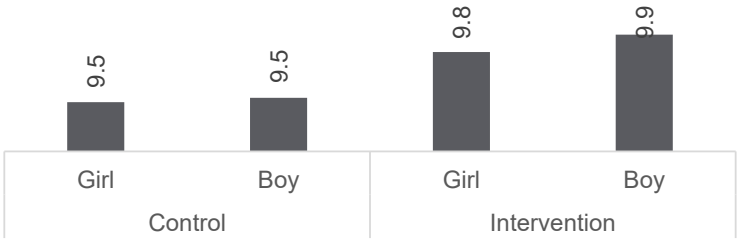
4.2.2. Performance per sub-task: Oral vocabulary



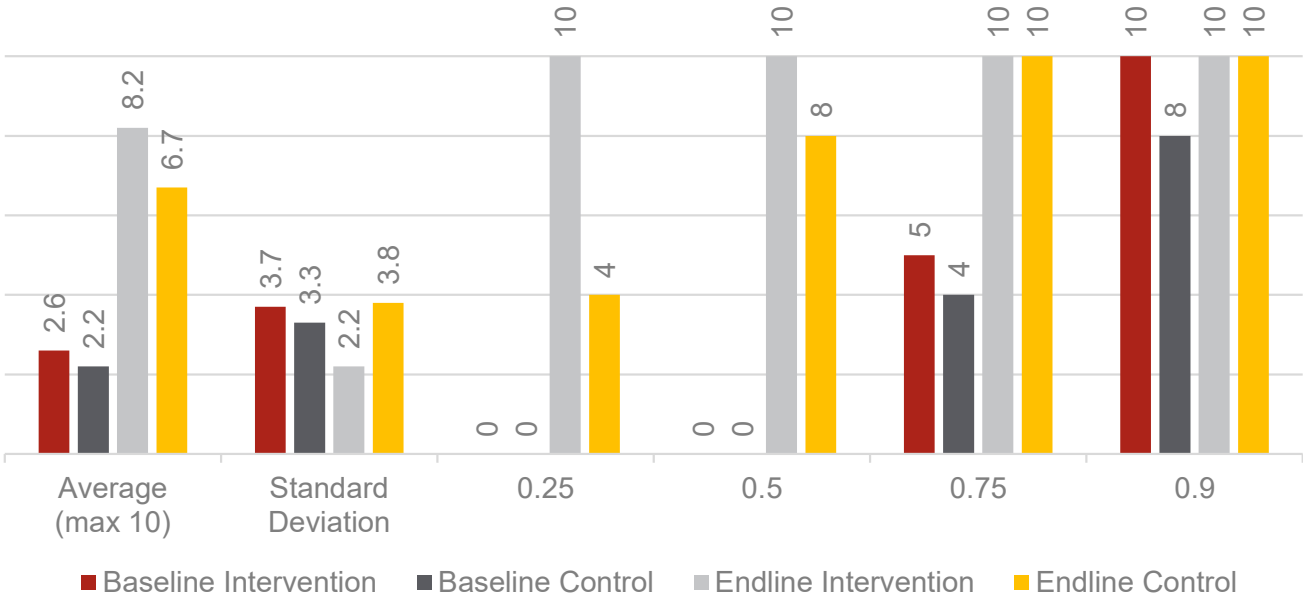
- The objective of this sub-task is to check for the student's ability to identify objects and to check if they have the vocabulary for the objects found/ actions seen in their immediate environment.
- Oral vocabulary is also a pre-school task, and Ei's past data has shown a ceiling effect on this task.
- The decrease in standard deviation shows improvement in vocabulary skills.
- घर and पतंग were identified by almost 100% students in the Endline. They were the most identified in the baseline as well. बक्सा was the least identified in both BL (88.4%) an EL (96.5%).
- Students performed comparatively better in identifying objects than actions in both BL and EL. Though a noteworthy improvement is visible from BL to EL in identifying actions.
- 35% students identified one or more objects in this sub-task in the local dialect (penda for मटका) or in English (house for घर). All these student's answers were marked correct.

As expected, a ceiling effect can be seen in this task. This suggests that the children are quite familiar with their environment.

Gender-wise comparison on this sub-task shows boys performing better than girls in the intervention schools with only 0.1 difference in score. Boys and girls are performing similarly in the control schools.



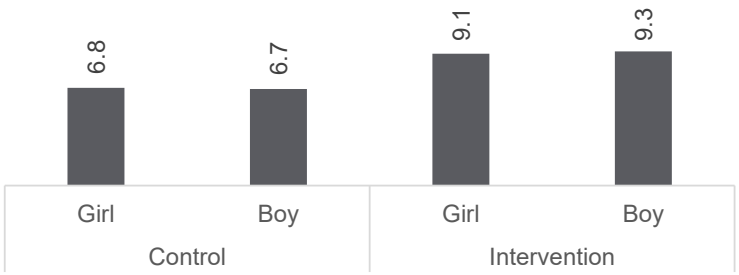
4.2.3. Performance per sub-task: Initial Sound Identification



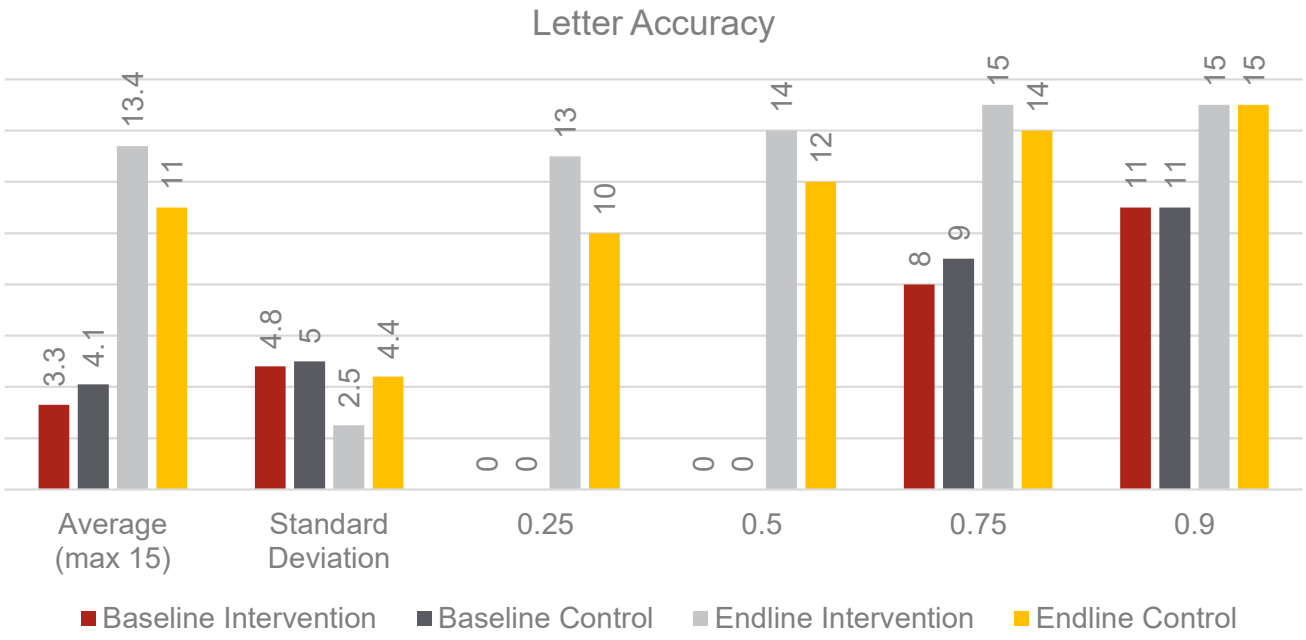
- The objective of this sub-task is to check if the student can understand and differentiate between sounds in a word, for instance identify the first sound in the word आम is आ .
- Most of the students are able to identify the initial sounds of words by the end of Grade 2.
- From 26% in BL to 78% in EL endorses Sonali Nag's framework on 3 phases of akshara learning in which she claims that at first children look at the letter globally (for example, CV is one unit/symbol) and then gradually learns to go deeper and analyze the phonemic marker in the symbol (for example: CV is made up of C and V). Though this framework is for letter recognition, it can be used to understand the high jump in learning in this task.
- Additionally, by Endline students would have developed the ability to comprehend the instruction correctly.

घात and झरना are the least identified while आम and रथ are the most identified in baseline. While तोता and कप are the least identified and ऊंट and एक are the most identified in Endline. घात is the most identified sound in Endline. आम and रथ being the 5th and 7th most identified in Endline.

- Gender-wise comparison on this sub-task shows boys performing better that girls in the intervention schools with only 0.2 difference in score. Girls are performing better than the boys in the control schools with only 0.1 difference in score.



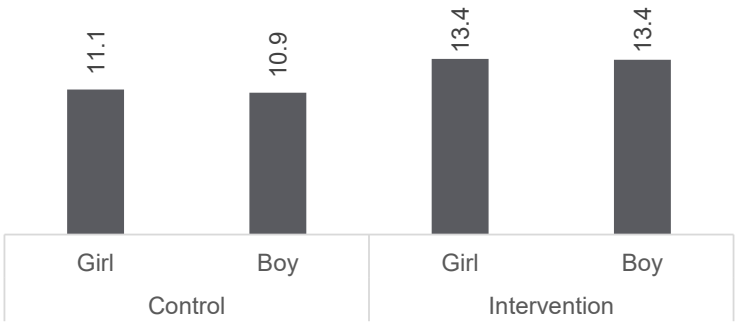
4.2.4. Performance per sub-task: Letter Reading – Accuracy and Fluency

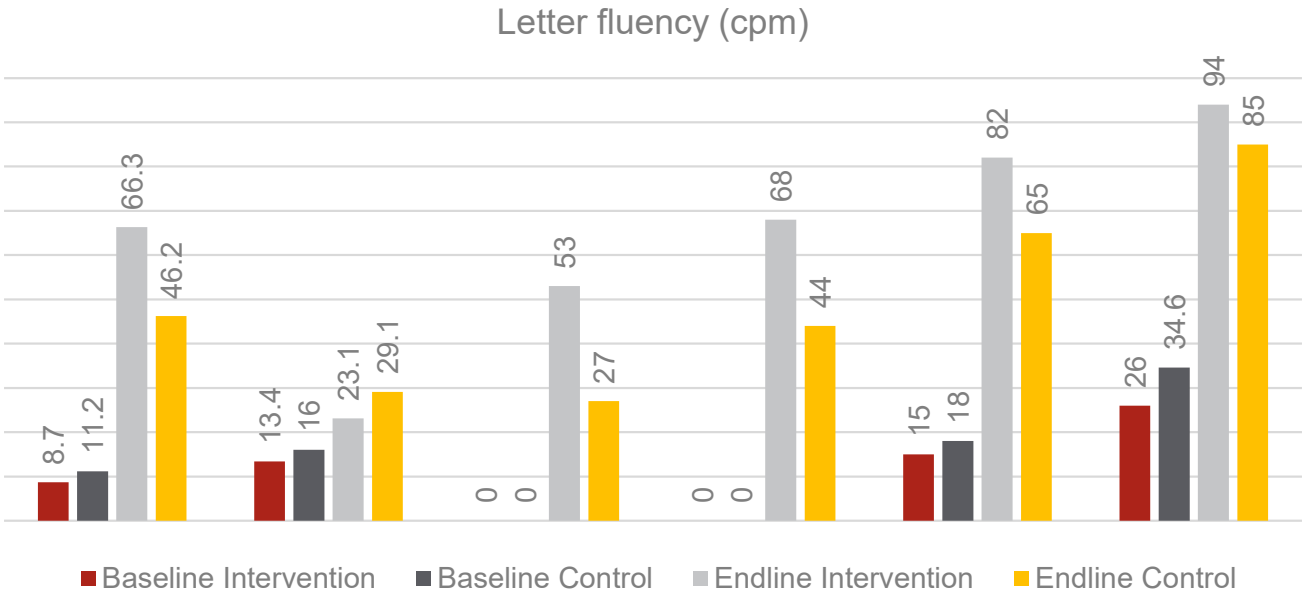


- The objective of this task is to check if the student is able to identify letters accurately. There is no time restriction on the first part of this task.
- A significant rise in accuracy can be seen in letter recognition task from baseline to Endline clearly indicating improvement.
- In all the grades sanyuktakshara and akshara with matra are the least recognized letters. The highest recognized letters in all the grades are high frequency letters. Although, the average of accuracy has exponentially risen, the fact that low frequency letters remain a challenge is alarming.

Round	Least identified	Most identified
BL	वु त्र क्ष मे	आ उ स ख
EL	मे य ज्ञ क्ष	फ स छ आ

- Interestingly, आ being the most identified in BL moves to the fourth most identified letter in EL. Perhaps, students identify the vowel sign more than the vowel itself.
- An increase in percentage of students who scored 100 from 2.2% in baseline to 40% and a decrease in percentage of students who scored 0 from 61% to 1.5% in Endline in intervention group reveals that students have attained proficiency in this skill.
- Gender-wise comparison on this sub-task shows boys and girls performing similarly in the intervention schools. Girls are performing better than the boys in the control schools with only 0.2 difference in score.

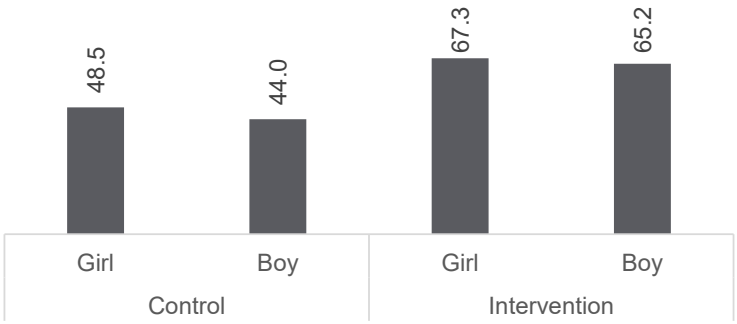




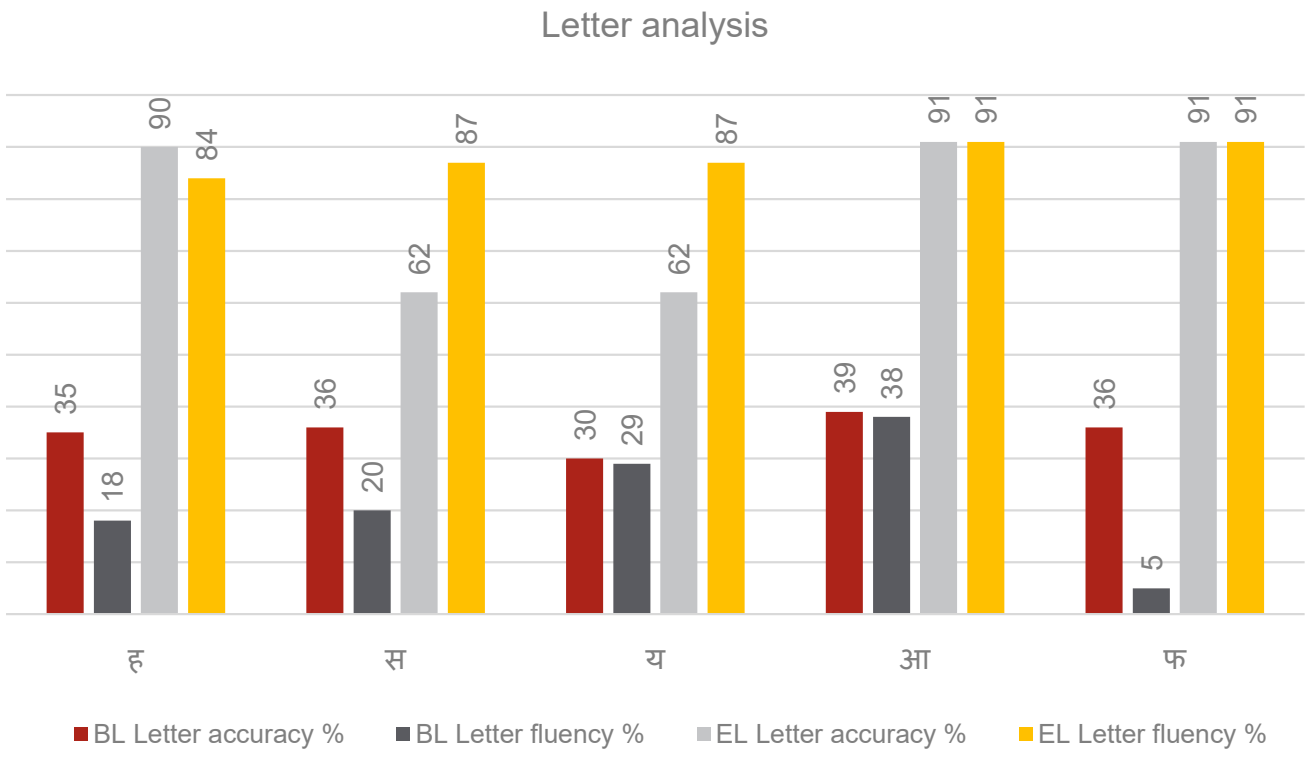
- The objective of this task is to check if the student is able to identify letters with fluency. A one-minute timer was put on this task, and no maximum letter restriction was present.
- Accuracy is given more importance than the pace at which letters are being identified. Until mastery in fluency is attained, students will always struggle with the cognitive load

to identify and blend letters/sounds over automatically (from their memory) reading the word.

- The low scores can be attributed to lack of teacher's efficiency and knowledge about the importance of fluency.
- Gender-wise comparison on this sub-task shows girls performing better than the boys in the intervention schools with a 2.1 difference in score. Similarly, girls are performing better than the boys in the control schools with 4.5 difference in score.

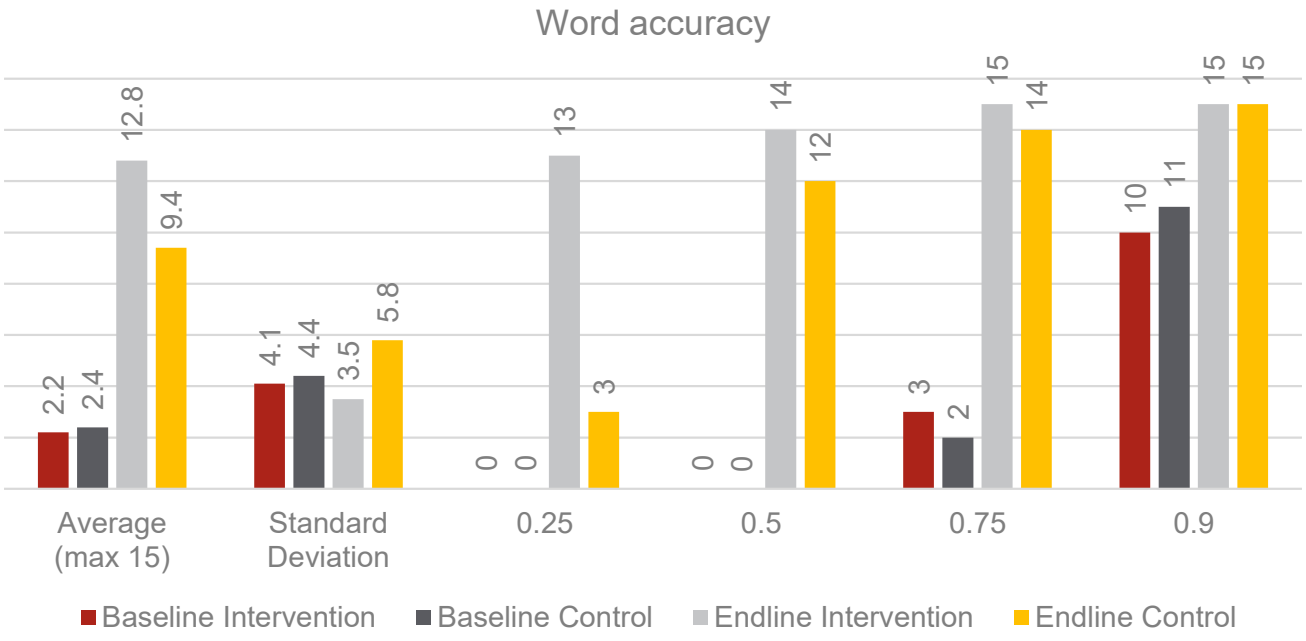


As opposed to accuracy, children did poorly in fluency. In baseline 27% of the students identified the letters accurately while only 13% could identify fluently. The Endline result surely shows a significant rise to 80% students identifying the letters accurately to 56% identifying fluently. Yet, the difference between accuracy and fluency is high.



- Performance of स, आ, and फ are similar in both accuracy and fluency signifying students' mastery on these letters.
- Fascinatingly, य which was 3rd on accuracy task and 12th on fluency task, shows a stark difference and a jump from 62% of the students identifying the letter in accuracy task to 87 % identifying in fluency task.
- Despite of ह being further down the list in letter fluency task, the difference in performance is not that high; from 90 % in accuracy task to 84% in fluency task.

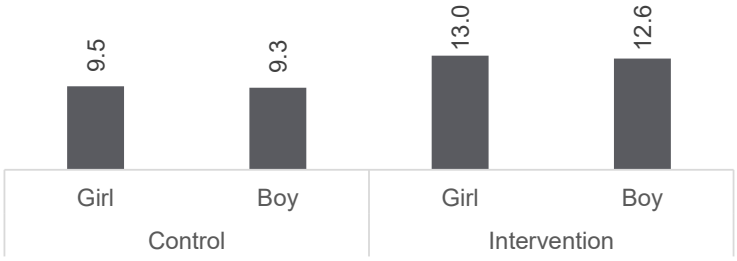
4.2.5. Performance per sub-task: Word Reading – Accuracy and Fluency

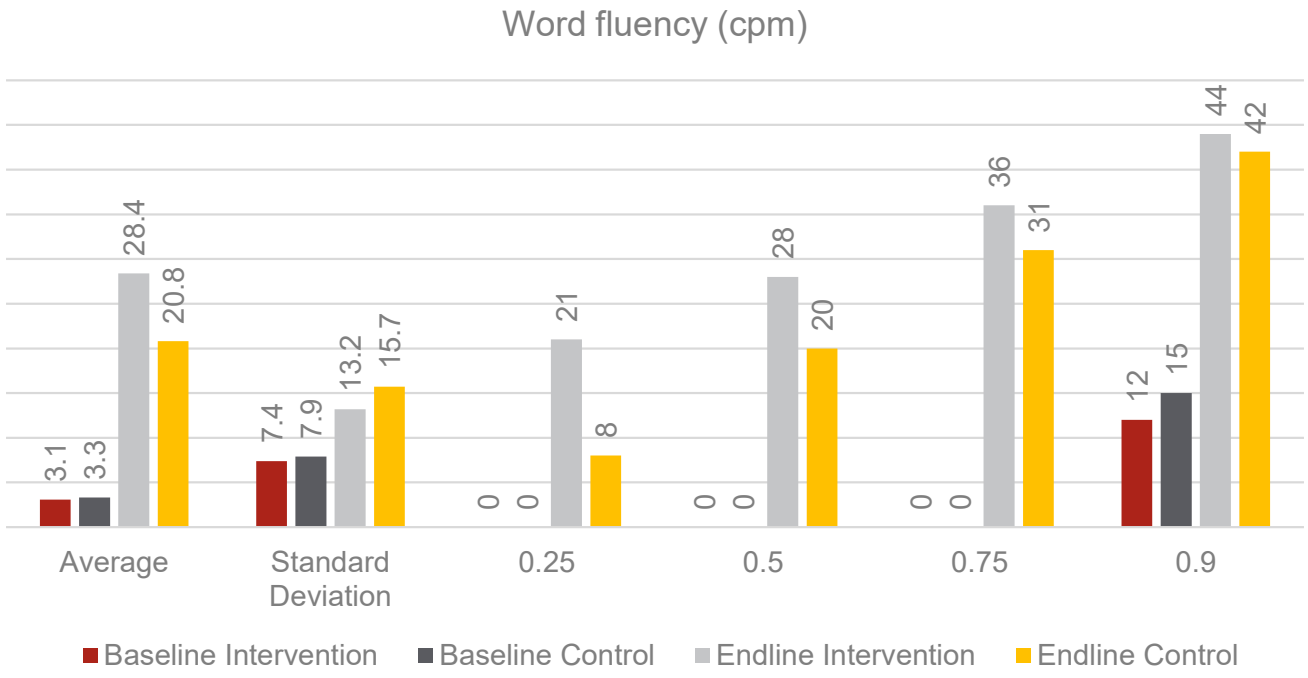


- The objective of this task is to check if the student can identify words (appropriate to their grade level) with accuracy.
- Less than 10% students have scored 0 on this task
- In BL words with sanyuktakshara and half letter and matra words are the least recognized words. In EL both four letter no matra words are also the least recognized words.

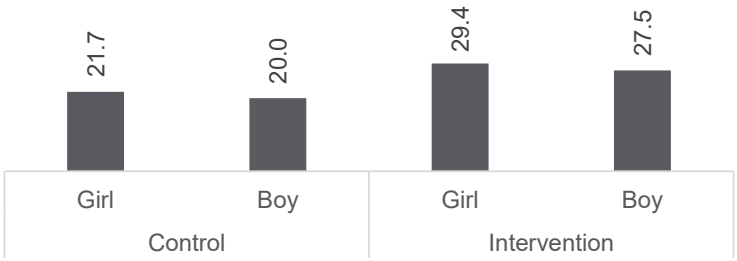
Round	Least identified	Most identified
BL	जूठी रंग छड़ी अम्मा	चल एक बतख बादल
EL	झगड़ा मस्ती सरकस नटखट	चल एक बतख छाता

- The highest recognized words in all the grades are 2 letter words with no matra. Same as letter accuracy, the average has exponentially risen from 18% in BL to nearly 75% in EL with std dev being nearly the same.
- In EL, in letter accuracy छ has moved to the most identified letter and so has in word accuracy.
- Gender-wise comparison on this sub-task shows girls performing better than the boys in the intervention schools with a 0.4 difference in score. Similarly, girls are performing better than the boys in the control schools with 0.2 difference in score.

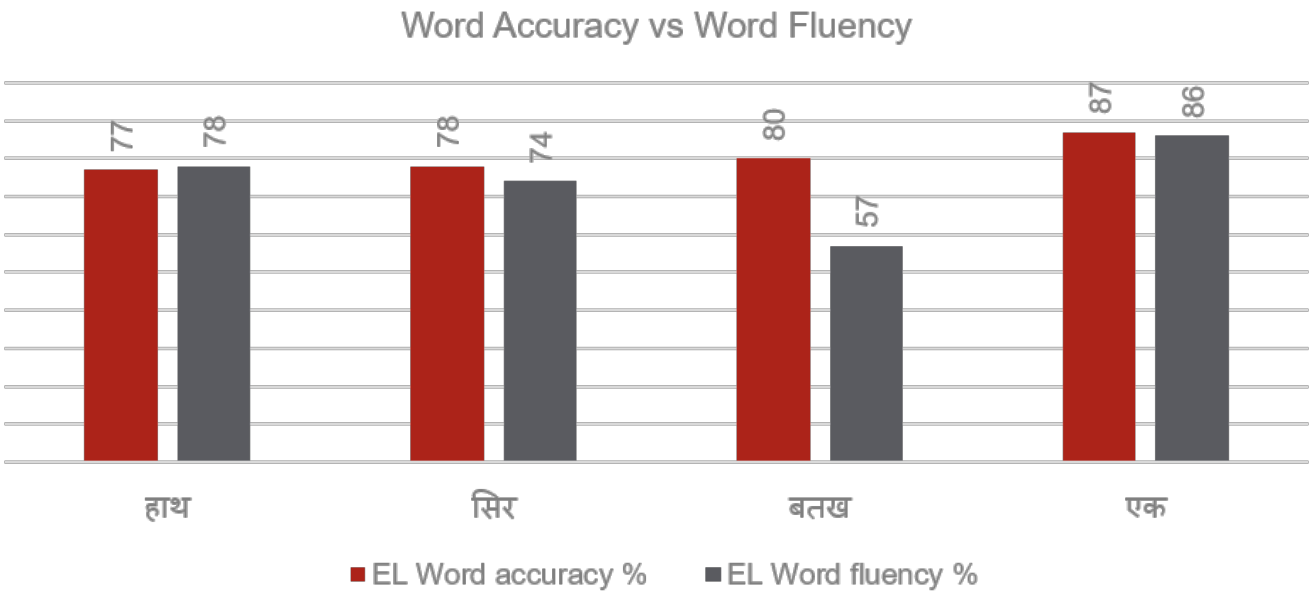




- The objective of this task is to check if the student can identify words (appropriate to their grade level) with fluency. A one-minute timer was present on this task.
- As opposed to accuracy (74%), children have done incompetently in fluency (~25%), which was expected.
- Gender-wise comparison on this sub-task shows girls performing better than the boys in the intervention schools with a 1.9 difference in score. Similarly, girls are performing better than the boys in the control schools with 1.7 difference in score.

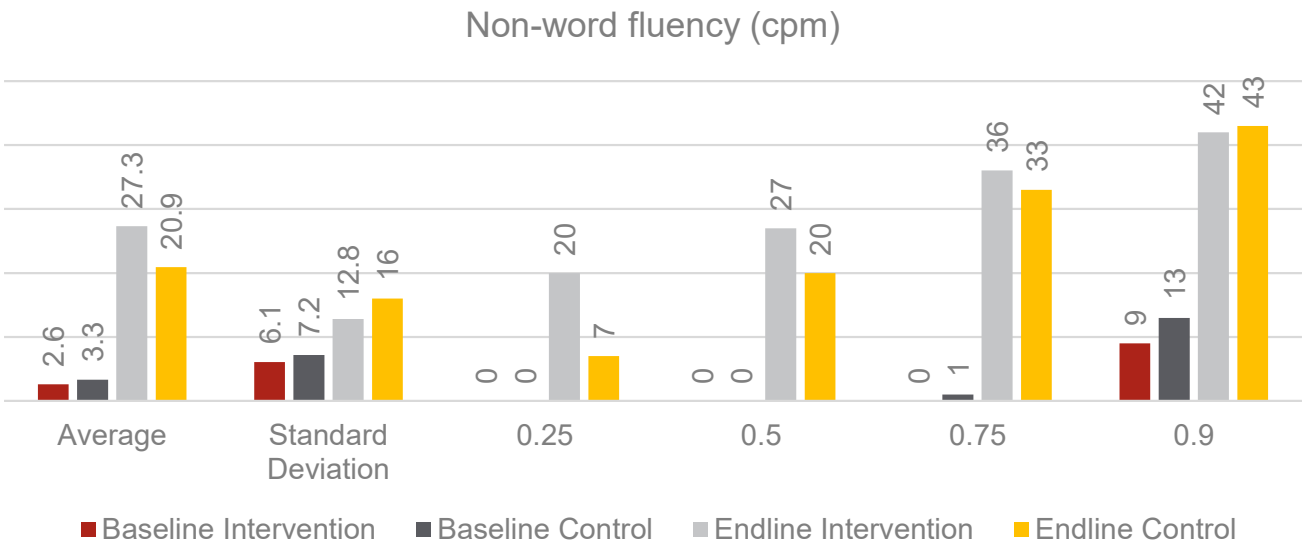


Impressively, the percentage of students who have scored 0 has come down to almost 0% in EL as opposed to 66% in BL in intervention group. This shows that mostly all the children can read at least 1 word correctly in the fluency task. This can also be seen in the percentile scores.

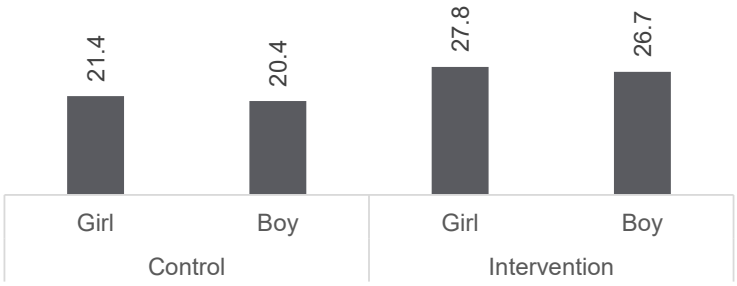


- Performance of हाथ and एक are similar in both accuracy and fluency signifying students' mastery on these words.
- सिर had similar performance and was 11th word on accuracy task and 14th on fluency task.
- बतख was further down the list in fluency task (24th), which is likely to explain lower performance on the same letter.

4.2.6. Performance per sub-task: Non-word Reading Fluency

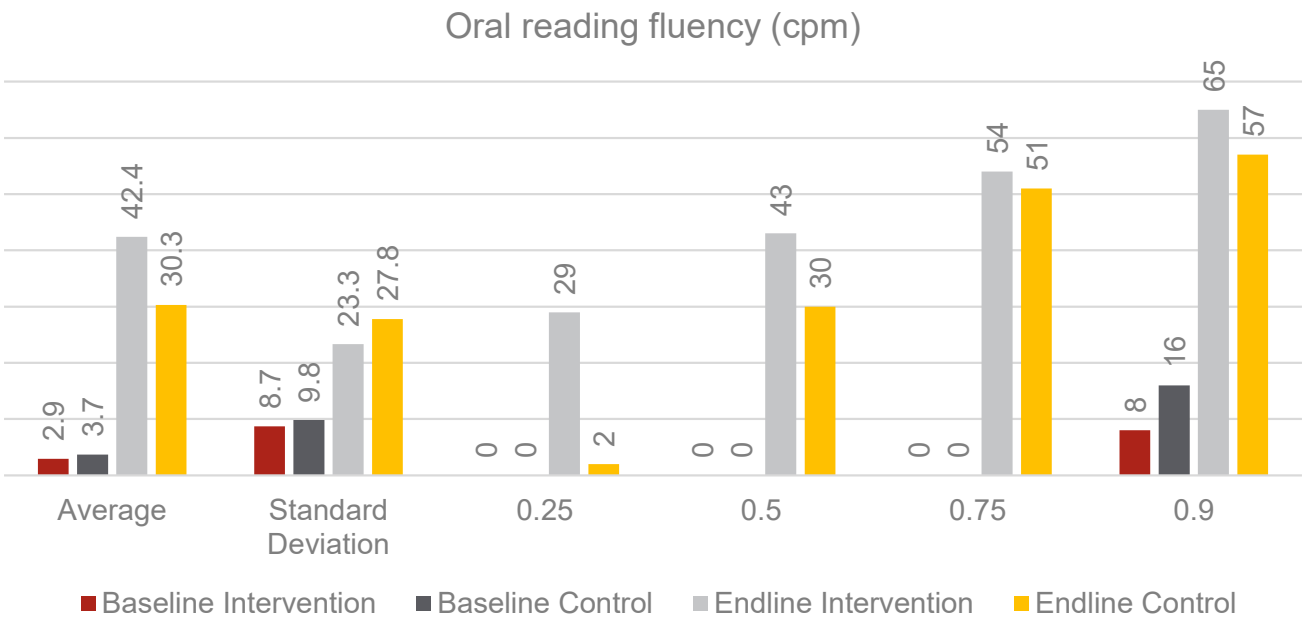


- The objective of this sub-task is to check whether the student is able to blend and create words which they have not seen and hence have no visual memory of.
- Overall, 47% students scored below the average score of 26.9.
- It was observed that most of the students were able to read the initial words in the order which were usually 2-3 letters non-words. As the difficulty level of the non-words increased, the students were not able to read it fluently.
- This is an important skill to analyse the student responses to gain information on the accuracy of letter-sound relationships, accuracy in unitizing letter-sounds into a whole word, and fluency and automaticity with correctly reading whole words.
- Gender-wise comparison on this sub-task shows girls performing better than the boys in the intervention schools with a 1.1 difference in score. Similarly, girls are performing better than the boys in the control schools with 1.0 difference in score.

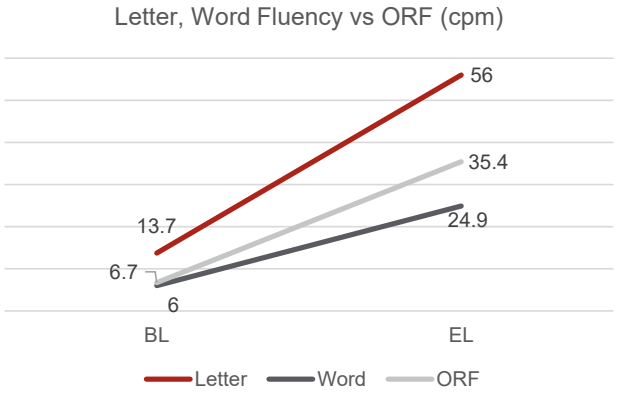


The result of non-word fluency is the same as word fluency with children doing just a little better in word fluency. This shows that children who has attained word decoding skills are able to apply it to non-words and decode them.

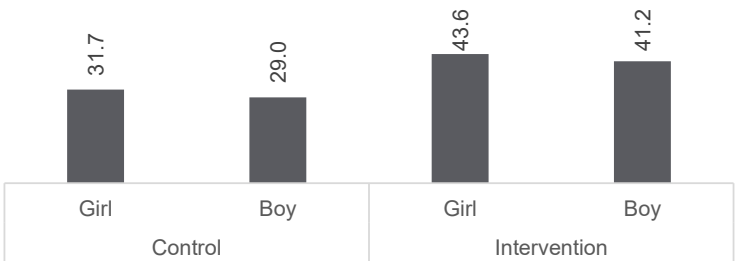
4.2.7. Performance per sub-task: Oral Reading Fluency



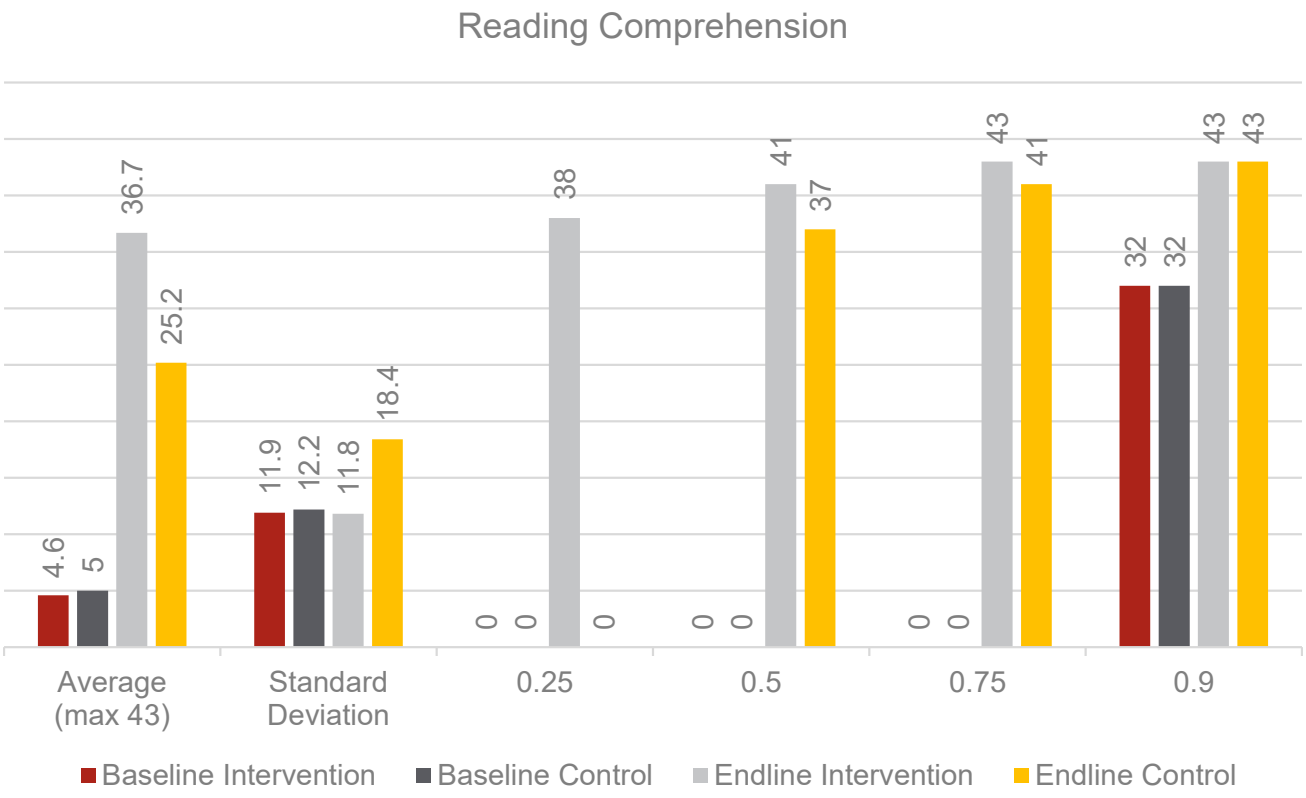
- The objective of this sub-task is to check if the student is able to read connected sentences with the necessary prosodic behaviour & interjections.
- This task is a timed task and must be calculated words read per minute.
- Interestingly, children have done much better in ORF than word or non-word fluency tasks, though the scores are similar in baseline, by Endline a massive rise can be seen.



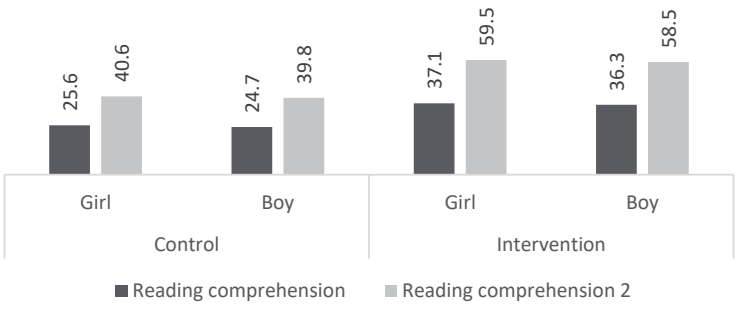
- This rise in the score can also be because maybe the students enjoyed reading the passage.
- Gender-wise comparison on this sub-task shows girls performing better than the boys in the intervention schools with a 2.4 difference in score. Similarly, girls are performing better



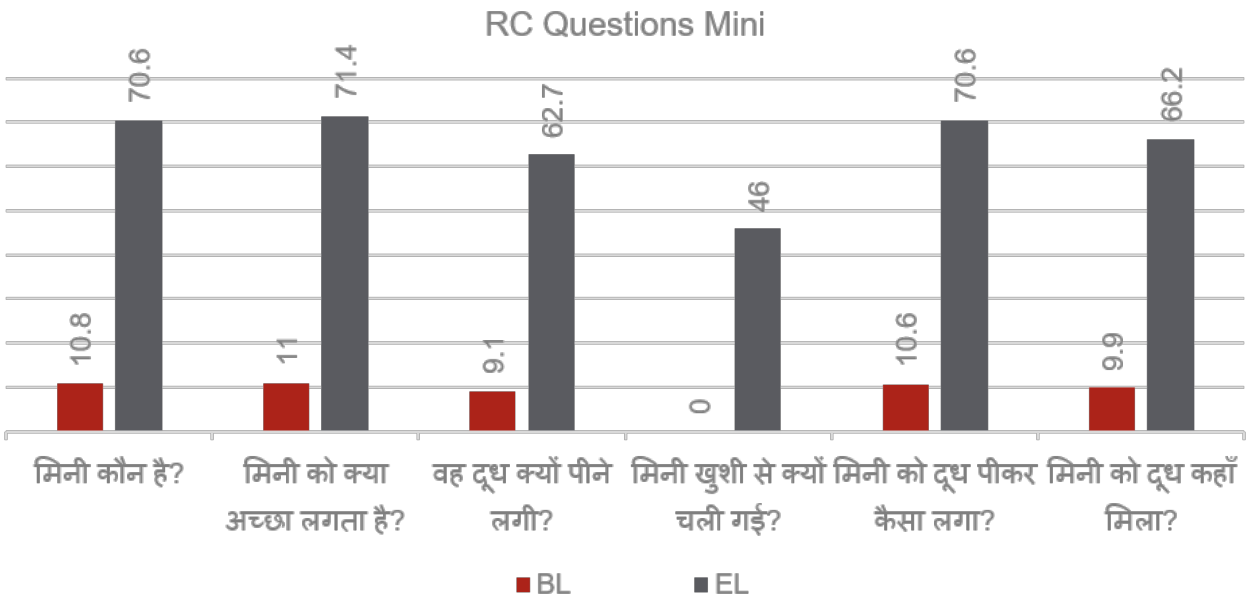
4.2.8. Performance per sub-task: Reading comprehension Accuracy and Answers



- The objective of this sub-task is to understand if the student is able to read a passage and respond to the questions asked from the story.
- Gender-wise comparison on this sub-task shows girls performing better than the boys in the intervention schools with a 0.8 and 1.0 difference in score. Similarly, girls are performing better than the boys in the control schools with 0.9 and 0.8 difference in score.

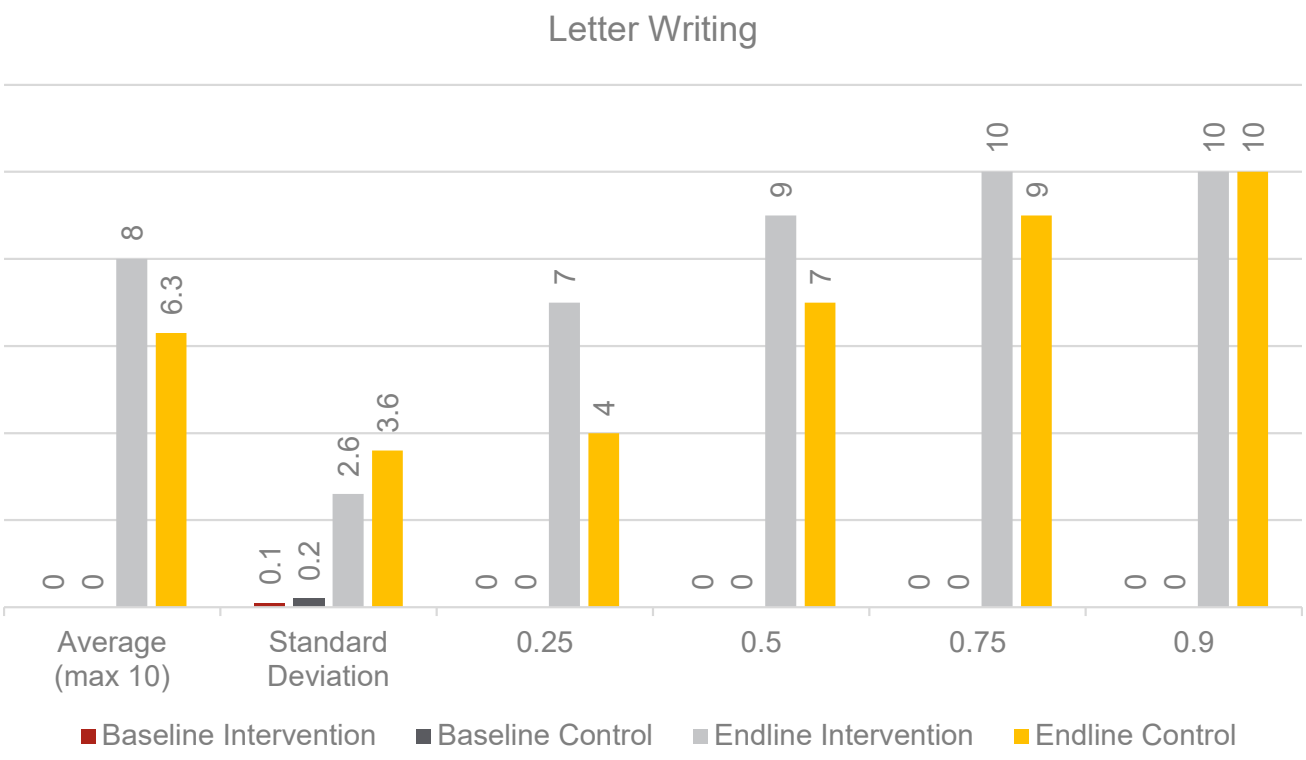


- The passage on Mini was the same in BL and EL. The graph clearly reveals that a huge improvement is there in students from BL to EL. Students are able to answer the questions which shows they are able to make meaning.
- Post discussion, the question क्या मनी को दूध अच्छा लगता है? was replaced with मनी ख़ुशी से क्यों चली गई? as the latter question is more inferential.

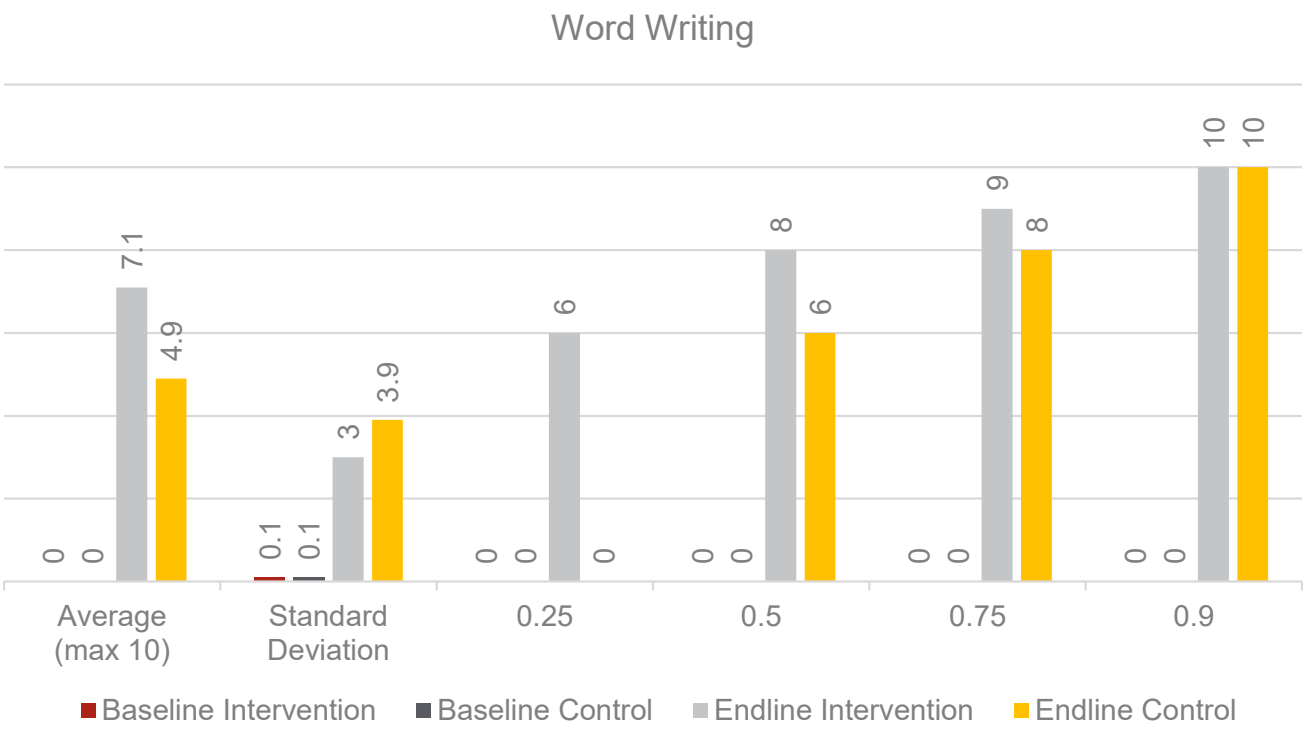


The average for passage and questions match and the rise from BI to EL is staggering. Fascinatingly, by EL students are comprehending the text read.

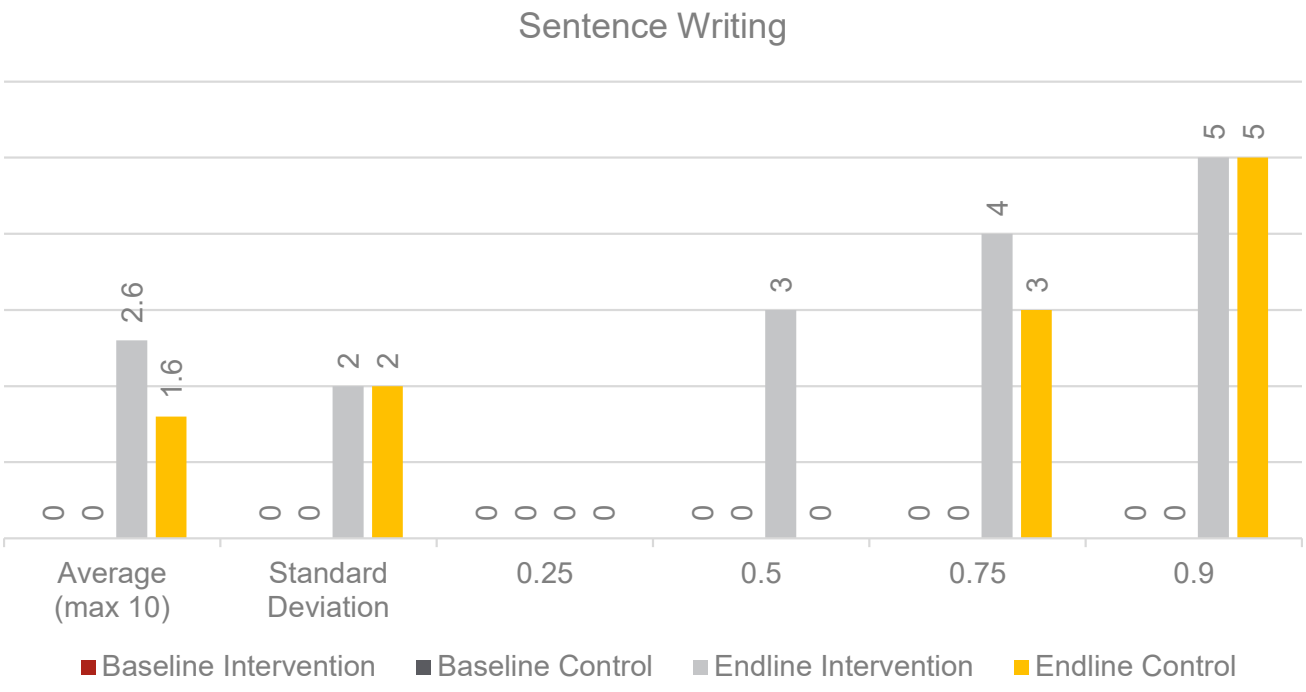
4.2.9. Performance per sub-task: Dictation – Letters, Words, Sentences



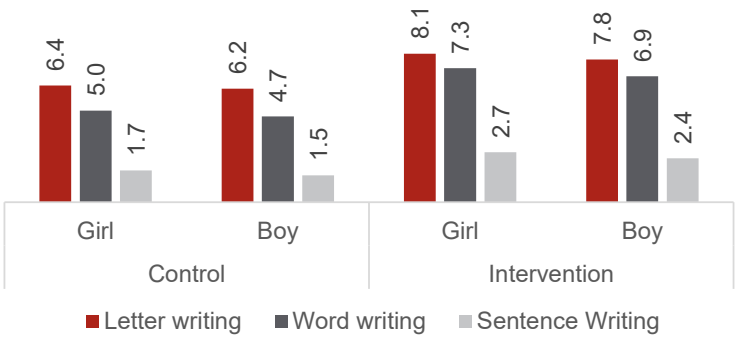
- The objective of this sub-task is to understand if the student is able to write the letters for the sounds heard.
- Intervention students have shown a significant improvement from the Baseline.



- The objective of this sub-task is to understand if the student is able to write the words with correct spellings when dictated to them.
- Writing tasks were clearly above the ability level of children at Grade 1 but we have seen a significant improvement in the Endline



- The objective of this sub-task is to understand if the student is able to write sentences that are dictated to them.
- Students have not mastered the ability to write the sentences.
- Gender-wise comparison on this sub-task shows girls performing better than the boys both in intervention and control schools across all writing tasks.



5. Appendix

Performance by Gender

Baseline					
Sub-task #	Sub-tasks	Girl		Boy	
		Control	Intervention	Control	Intervention
1	Listening comprehension	2.9	2.9	2.9	2.8
2	Oral vocabulary	8.6	9.0	9.0	9.1
3	Initial sound identification	2.2	2.7	2.1	2.5
4a	Letter accuracy	4.3	3.5	3.9	3.2
4b	Letter fluency (cpm)	11.7	9.3	10.9	8.1
5a	Word accuracy	2.6	2.5	2.3	2.0
5b	Word fluency (cpm)	3.3	3.3	3.4	2.8
6	Non-word fluency (cpm)	3.7	2.9	3.1	2.3
7	Oral reading fluency (cpm)	4.2	3.6	3.3	2.3
8	Reading comprehension	5.7	5.3	4.5	4.0
9	Reading comprehension 2	0.0	0.0	0.0	0.0
10a	Letter writing	0.0	0.0	0.0	0.0
10b	Word writing	0.0	0.0	0.0	0.0
10c	Sentence Writing	0.0	0.0	0.0	0.0

Endline					
Sub-task #	Sub-tasks	Girl		Boy	
		Control	Intervention	Control	Intervention
1	Listening comprehension	3.0	3.0	3.2	3.4
2	Oral vocabulary	9.5	9.5	9.8	9.9
3	Initial sound identification	6.8	6.7	9.1	9.3
4a	Letter accuracy	11.1	10.9	13.4	13.4
4b	Letter fluency (cpm)	48.5	44.0	67.3	65.2
5a	Word accuracy	9.5	9.3	13.0	12.6
5b	Word fluency (cpm)	21.7	20.0	29.4	27.5
6	Non-word fluency (cpm)	21.4	20.4	27.8	26.7
7	Oral reading fluency (cpm)	31.7	29.0	43.6	41.2
8	Reading comprehension	25.6	24.7	37.1	36.3
9	Reading comprehension 2	40.6	39.8	59.5	58.5
10a	Letter writing	6.4	6.2	8.1	7.8
10b	Word writing	5.0	4.7	7.3	6.9
10c	Sentence Writing	1.7	1.5	2.7	2.4

Comparison with Indian Studies

Project Name	Implementing Partner	Period of Performance	Study Name	Indicators	Effect Size (SD)
DIB Haryana	LLF	April 2020-April 2022	DIB Haryana	ORF	0.67
				RC	0.88
Scaling Up Early Reading Intervention	Reading Intervention Room to Read (R2R)	September 2015 – September 2020	SERI Chattisgarh	ORF	0.43
				RC	0.33
			SERI Uttarakhand	ORF	0.24
				RC	0.2
Nurturing Early Literacy	Centre for microFinance (CmF)	October 2015 – September 2019	NEL Maharashtra	ORF	-0.04
				RC	-0.02
			NEL Rajasthan	ORF	-0.02
				RC	-0.01
			NEL Karnataka	ORF	0.04
				RC	0.01
Teacher Innovations in Practice Schools and Teachers Innovating for Results	(STIR) Education	October 2014 – September 2018	TIP Uttar Pradesh	ORF	-0.04
				RC	-0.04
Start Early: Read in Time	CARE	July 2014–July 2018	SERT Odisha	ORF	0.14
				RC	0.06
			SERT Uttar Pradesh	ORF	0.06
				RC	0.09
RightToRead	EnglishHelper	September 2015 – September 2019	RTR Maharashtra	ORF	0.11
				RC	-0.03

Comparison with International Studies

Country (program)	Effect size, Letter Fluency	Effect size, ORF	Effect size, RC (%)
India, DIB Haryana	1.07	0.67	0.67
Egypt, Arab Rep.	1.07	0.55	.
Nigeria	0.78	0.66	0.6
Kenya	0.68	0.41	0.4
Mali	0.63	0.33	.
Uganda	0.55	0.35	0.35
Liberia (EGRA Plus)	0.52	0.8	0.82
Jordan	0.48	0.46	.
Tonga	0.31	0.24	0.14
Kyrgyz Republic	0.23	0.27	0.19
Congo, Dem. Rep.	0.01	0.13	.
PNG	.	0.3	0.12
South Africa	.	0.8	0.59

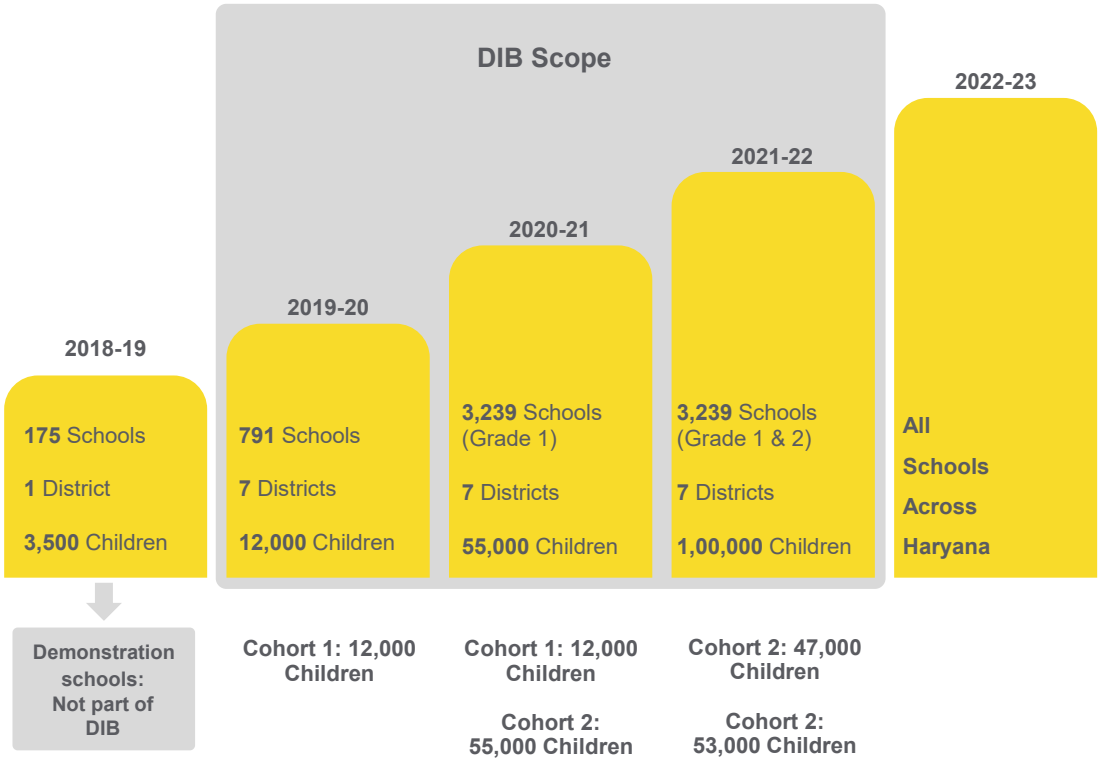
Country (program)	DiD, Letter Fluency	DiD, ORF (cwpm)	DiD, RC (%)
India, DIB Haryana	22.5	12.9	28
Malawi (MTPDS)	20.14	8.88	11
Egypt, Arab Rep.	19	12	.
Liberia (EGRA Plus)	14.8	21.2	25.2
South Africa	14.32	7.21	9
Jordan	11.6	6.3	.
Liberia (LTTP II)	11.1	7.03	7.67
Mozambiq ue	8.8	3.03	3.18
Nigeria	6.1	4.9	0.1
Malawi (EGRA)	5.35	1.01	-0.01
Uganda	3.4	3.04	0.24
Philippines	3.12	2.16	3.96
Congo, Dem. Rep.	0.17	3.07	3
Tanzania	.	0.07	-1.5

Brief introduction to Impact Bonds

Impact bonds are a form of results-based financing instrument that Social Finance pioneered in 2010 in the U.K. Impact Bonds are now being used globally to shift the focus of social and development funding towards outcomes, build a culture of monitoring and evaluation, and encourage performance management and adaptability in intervention design. Impact bonds blend impact investing, results-based financing, and public-private partnerships. Most impact bonds involve the following key actors: (a) the service providers who deliver social services to the population in need; (b) the risk investors who provide upfront capital; (c) the outcome funder who repays the investor their principal plus an agreed upon return on investment if target outcomes are achieved; (d) an independent evaluator who verifies whether pre-agreed outcomes have been achieved; and (e) the intermediary/ convener who designs and structures the instrument, facilitates fundraising and negotiations among the parties.

Scope of the DIB

The DIB will scale up the existing program of LLF in the state of Haryana. In 2017, Haryana School Shiksha Pariyojna Parishad (HSSPP) 1 and LLF entered into an agreement to strengthen the academic support provided to Hindi teachers by the state resource personnel (including the Block Resource Persons (BRP), Assistant BRPs and others). The ultimate aim is to develop a state specific early literacy learning package and to improve students' language learning outcomes. A school-level intervention was successfully completed in the academic session April 2018- March 2019, covering a total of 175 schools and 3,500 students in Grade 1.



DIB Targeted Outcomes

The DIB aims to systemically strengthen the capacities of teachers and teacher educators to improve student outcomes in early grade language, that can then be further replicated by the government across the state. The DIB will target that by the end of Grade 2 students can:

- Read grade-level text fluently, and
- Understand appropriate texts as well as write simple sentences.

These outcomes will be evaluated using an assessment tool based on the Early Grade Reading Assessment (EGRA) that is widely used globally to examine gaps in reading competencies among students