RightToRead

LARGE-SCALE DEPLOYMENT OF TECH-ENABLED ENGLISH READING

REPORT ON LEARNING OUTCOMES 2017

— IN PARTNERSHIP WITH —

USAID
UNITED STATES AGENCY INTERNATIONAL DEVELOPMENT

EnglishHelper™
Learn at your pace, in your context

IL&FS | Education

AIF AMERICAN INDIA FOUNDATION

GRAYMATTERS

STAMP
A initiative of IL&FS | Education

ACKNOWLEDGEMENTS

Government of Maharashtra

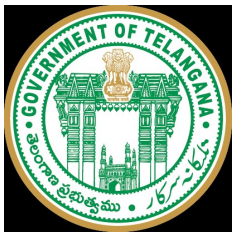
Government of National Capital Territory of Delhi

Government of West Bengal

Government of Tamilnadu

Government of Punjab

Government of Telangana



Government of Gujarat

Government of Karnataka

Contents

1 The RightToRead Program 3
1.1 India's Literacy Crisis 3
1.2 Multi-sensory Linguistic Learning 3
1.3 The RightToRead Program 3
2 Assessment Design 6
2.1 Geography Covered 6
2.2 Control-Treatment Design 6
2.3 Baseline-End line Design 6
2.4 Sampling 7
2.5 Selection of Control 8
3 Assessment Delivery 9
3.1 Communication With School Administrators 9
3.2 Enabling The Assessment Platform 9
3.3 Operationalising Assessments 9
3.4 Data Capture And Transfer 9
4 Analysis Methodology 10
4.1 Testing For Completeness 10
4.2 Testing For Accuracy 10
4.3 Detection Of Anomalies 10
4.4 Back-checks On The Data 10
5 Assessment Outcome 11
5.1 Overall Outcome 11
5.1.1 Improvement in Outcome 12
5.1.2 Change by Quartile 12
5.2 Outcome In West Bengal 14
5.2.1 Sample Size 14
5.2.2 Assessment Outcome – West Bengal 14
5.2.3 Improvement in Outcome – West Bengal 15
5.2.4 Change by Quartile – West Bengal 16
5.3 Outcome In Maharashtra 17
5.3.1 Sample Size 17
5.3.2 Assessment Outcome – Maharashtra 17
5.3.3 Improvement in Outcome – Maharashtra 18

5.3.4 Change by Quartile – Maharashtra 18
5.4 Outcome In Punjab 20
5.4.1 Sample Size 20
5.4.2 Assessment Outcome – Punjab 20
5.4.3 Improvement in Outcome – Punjab 21
5.4.4 Change by Quartile – Punjab 22
5.5 Outcome In Gujarat 23
5.5.1 Sample Size 23
5.5.2 Assessment Outcome – Gujarat 23
5.5.3 Improvement in Outcome – Gujarat 24
5.5.4 Change by Quartile – Gujarat 25
6 The CEFR Analysis 26
6.1 Adapting The CEFR Mapping Framework 26
6.2 Assigning A CEFR Level To Questions 27
6.3 Assigning A CEFR Level To Students 28
6.4 Comparing The CEFR Distribution Between Baseline And End line 28
7 Conclusion 30

Appendix

Assessment Instruments

1 THE RIGHTTOREAD PROGRAM

1.1 INDIA'S LITERACY CRISIS

India has a reading and comprehension skill crisis. Over the last decade, surveys conducted across government schools indicate a consistent and significant gap in reading and comprehension ability of students.

English reading and comprehension skill is also lagging. This is a source of significant concern since English proficiency is considered socially empowering, enhances higher education and job-related opportunities. Across the country, though English is being uniformly introduced into curriculum from early grades, many factors challenge students as they seek to acquire English proficiency.

According to the Annual Status of Education Report, 2016 (ASER 2016)[1], 22.8% of the children from Grade 3 who were assessed could not even read capital letters and only 32% of children from Grade 3 could read short words. The higher grades witnessed a continuing decline in English reading ability; 45.2% of the children from Grade 8 were able to read simple sentences in English compared to 46.7% in 2014 and 60.2% in 2009. Irrespective of grade, only about 60% of children who can read words can explain the meanings of those words.

In this report, we share results from a randomised control assessment for a large scale roll out of ReadToMe™ - a multi-sensory reading and comprehension technology from EnglishHelper™ for learning in English. To this end, students from various Government and aided schools across the states where RightToRead has been deployed were assessed in partnership with independent third parties. This report documents the methodology and findings of these independent assessments.

1.2 MULTI-SENSORY LINGUISTIC LEARNING

Research has shown that we learn most effectively through multiple senses: auditory (sound), visual (sight), tactile (touch), kinesthetic (body movement). Teaching-learning methodology that can stimulate the learner's multiple senses is likely to keep the learner more engaged and achieve better outcomes. This is as true with language acquisition as with any other type of learning. Studies have opined that the human brain was not developed to recognise the letter-speech sound combinations required for reading fluency. Letter-speech sound variations are arbitrary cultural inventions. The brain creates a specialised neural pathway for recognising such arbitrary objects. Multi-sensory stimulation enables the speedy creation of such a neural network.

Given the scale of the challenge India faces this crisis cannot be addressed by just deploying human resource. The advancement and availability of relevant technology make it a compulsive opportunity that must be leveraged to improve the reading and comprehension abilities of children across India's schools.

1.3 THE RIGHTTOREAD PROGRAM

EnglishHelper™ (EH) enables technology-based reading and comprehension improvement for learners across all age groups. Since launching in India in 2011, EH has successfully implemented its reading and comprehension solution, ReadToMe™, in public and private schools across the country.

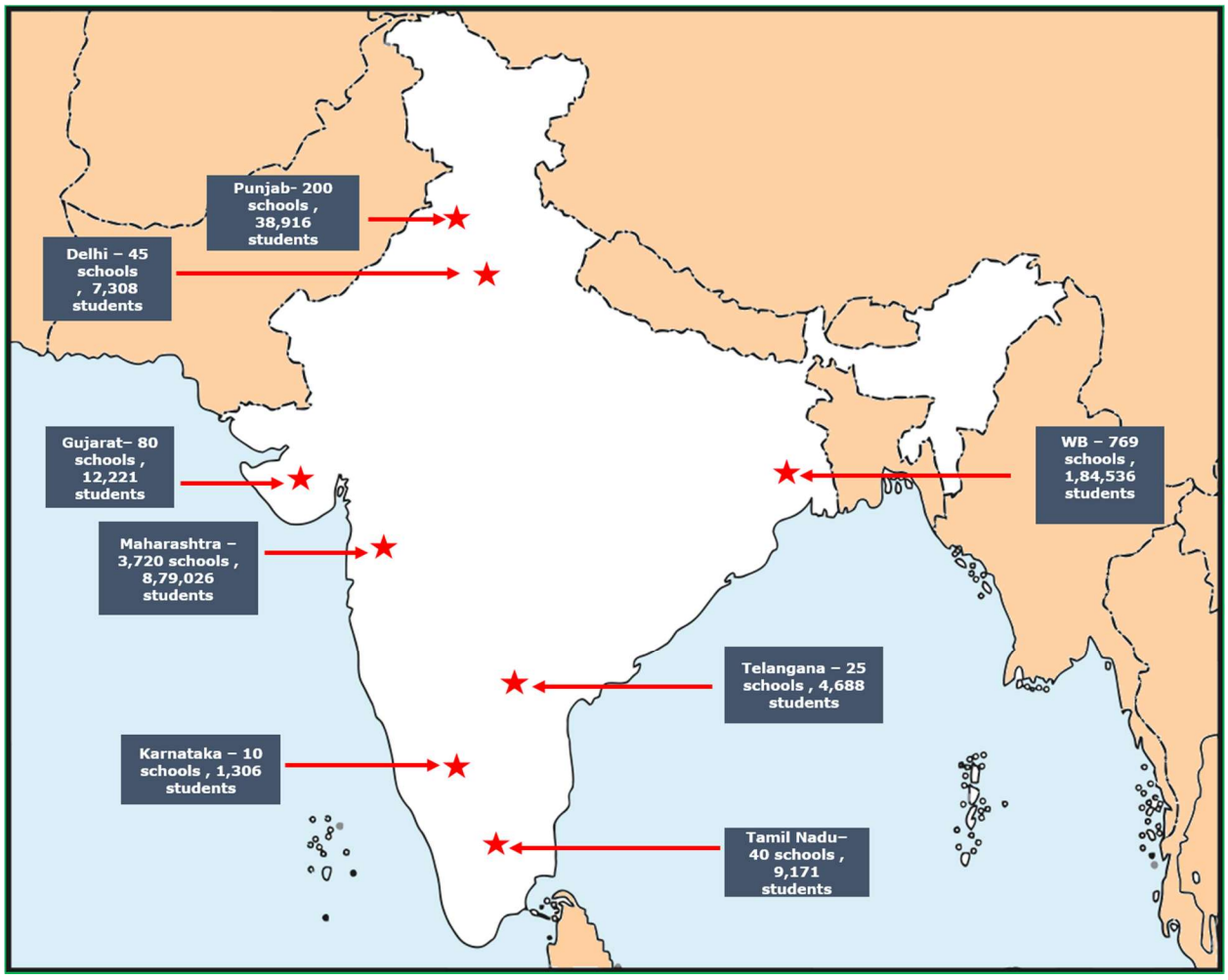
[1] ASER 2016 was conducted in 589 rural districts across India, sampling from every state.

The RightToRead Program is an effort to demonstrate that reading and comprehension technology can play a key role in solving the country's reading crisis. RightToRead has simple yet compelling objectives. The goal is to demonstrate that technology-enabled reading:

- enhances students reading and comprehension skills
- enhances teacher effectiveness
- can be deployed at scale, rapidly and efficiently.

Supported by the United States Agency for International Development (USAID) under the USAID–IPP initiative, EnglishHelper™ has deployed RightToRead to cover government and aided school students during the academic year 2016-17. USAID has established 'reading' as a major education objective. It aims to enable reading improvement for millions of primary school students in India. This large-scale rollout of RightToRead covered over 1 M students in grades 3 to 8 and touched 15,000 teachers in 5,000 schools across 8 States – Punjab, Delhi, Gujarat, Maharashtra, West Bengal, Tamil Nadu, Telangana and Karnataka. Please refer to the following map and table for details by state:

Figure 1: Geographic Coverage – RightToRead



The table below includes the distribution of teachers covered under the program by state:

Table 1: RightToRead *Coverage 2016-17*

State	Number of Students	Number of Teachers	Number of Schools
Maharashtra	8,79,026	11,160	3,720
West Bengal	1,84,536	2,307	769
Punjab	38,916	600	200
Gujarat	12,221	240	80
Tamil Nadu	9,171	120	40
Delhi	7,308	135	45
Telangana	4,688	75	25
Karnataka	1,306	30	10
Total	**11,37,172**	**14,667**	**4,889**

EH implemented RightToRead in the states of Maharashtra and West Bengal in partnership with IL&FS Education. In the remaining six states, EH partnered with American India Foundation (AIF). EH engaged Gray Matters India (GMI) for assessment design and analysis; Skill Training Assessment Management Partners (STAMP) provided the technology platform for conducting the assessments.

EH envisions that the results demonstrated from deploying RightToRead will create a strong case for education policy makers and administrators to leverage technology at scale to improve English reading and comprehension of students in schools.

2 ASSESSMENT DESIGN

To assess the impact of RightToRead in an unbiased manner, a randomised control design was adopted. A randomised Treatment group and Control group of schools was selected. Baseline and End line assessments were administered in these schools. This report reviews various elements of the randomised design and assessments. In addition, the assessments followed a standardised rubric across all states and grades assessed. Details of the assessment instrument and rubrics, with illustrative examples, are presented in the Appendix.

2.1 GEOGRAPHY COVERED

The states of Maharashtra, West Bengal, Punjab and Gujarat account for 98% of the student population covered by the USAID supported RightToRead project. Assessments were conducted in these states.

2.2 CONTROL-TREATMENT DESIGN

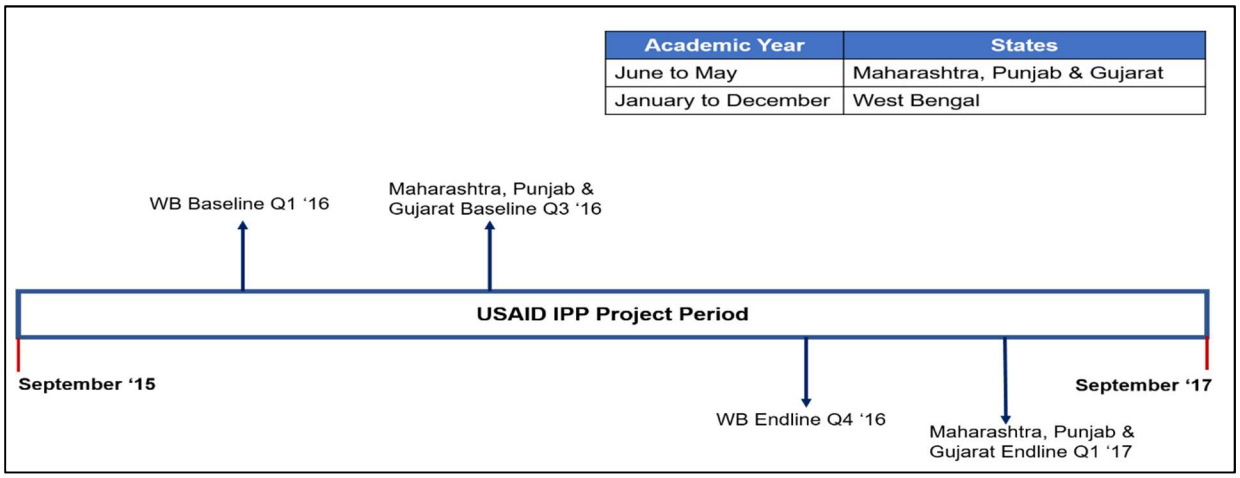
The assessments were conducted across grades 3 to 7. A Control-Treatment design was adopted for comparison of outcomes. Students who underwent technology-enabled reading under the RightToRead program were classified as the Treatment group. Students who were not exposed to technology-enabled English learning constituted the Control group. The design allows analysis of outcomes attributable mainly to the program, compared with learning that may be observed in a defined academic period in the absence of the program.

2.3 BASELINE-END LINE DESIGN

All students were assessed in the early part of the academic year for Baseline results (pre-test) and towards the end of the academic year for End line results (post-test). This enabled measurement of learning outcomes achieved during the academic year. Control and Treatment groups were assessed concurrently.

The schedule for the Baseline and End line assessments for the four states where assessments were undertaken is given below. The timeline for the assessments was dependent upon the academic calendar period stipulated by each state.

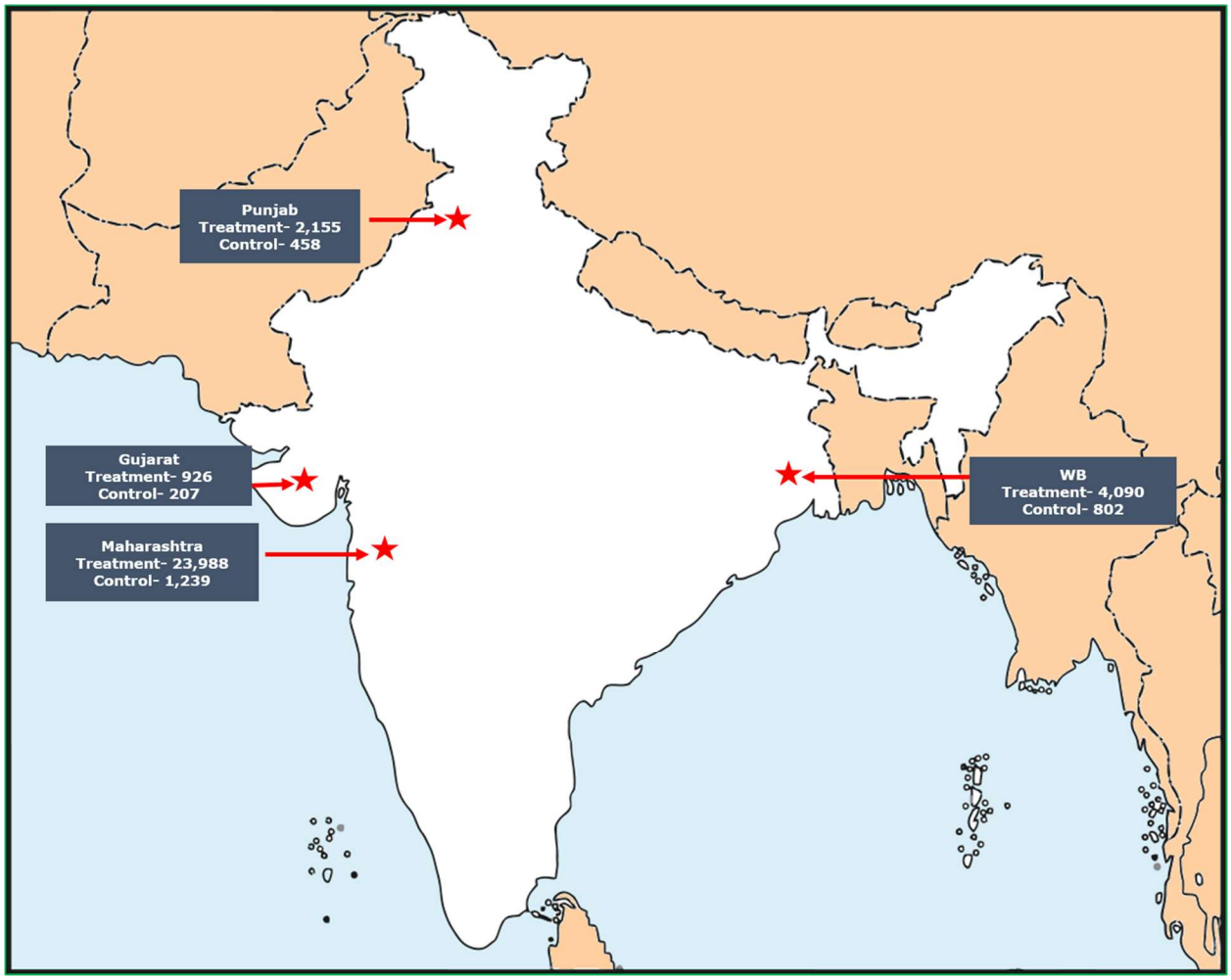
Figure 2: Assessment Schedule (Q = calendar quarter)



2.4 SAMPLING

More than 33,000 students were assessed across the four states and between Treatment and Control groups.

Figure 3: Sample Distribution



For the assessments, the schools in a district where the RightToRead program was implemented, constituted the population for the Treatment group. The districts were then segregated into clusters using the number of schools in the program as a clustering variable. A random sample of clusters was selected for the assessments. In the case of Maharashtra, given the high number of districts and schools that have implemented RightToRead, an additional variable of classification of district (semi-urban or urban) was used for clustering. All schools in a cluster were selected. All students in a school were assessed.

The size of the clusters and the final sample was maintained as a minimum proportion of the population (10%). Given the varying population sizes across the states, the final sample proportion to the population was different for each state to allow for minimum sample sizes. For e.g. the total number of Treatment group students (Treatment population) in Gujarat was approximately 12,000 at the beginning of the academic year 2016-17, while that in Maharashtra was approximately 8,80,000 students. Thus, to ensure adequate representation in the smallest sub-group that would be analysed, Gujarat required a higher sampling proportion compared to Maharashtra. Sample sizes were drawn such that the Maximum

Sampling Error would be maintained under 3.5% at a 95% confidence level for any state (and considering the maximum possible variation in responses – 50%).

Table 2: Maximum Sampling Error – State-wise

State	Population Size*	Sample Size	Maximum Sampling Error
West Bengal	1,85,000	4,090	1.5%
Maharashtra	8,80,000	23,988	0.6%
Punjab	39,000	2,155	2.1%
Gujarat	12,000	926	3.1%

*(Number of students covered by RightToRead at the beginning of the academic year, rounded off to the nearest thousand)

2.5 Selection of Control

The choice of Control schools was constrained by government and school permissions as well as by the presence of a matched (to Treatment) sample in the same district.

3 ASSESSMENT DELIVERY

The delivery of the assessments in schools was the team-effort of STAMP, EH, implementing partner/s and school stakeholders. Every stage in the delivery process conformed to stringent data collection and data integrity protocols.

3.1 COMMUNICATION WITH SCHOOL ADMINISTRATORS

Upon identification of schools for assessment, school administrators were informed of the dates, procedure and requirements of the assessment. Details pertinent to the assessments, including student-teacher data and the school academic calendar, were obtained. Baseline and End line assessments were scheduled such that they did not conflict with examination dates, vacations and holidays (the latter two are important to avoid low attendance).

3.2 ENABLING THE ASSESSMENT PLATFORM

Upon receipt from GMI, the assessment instruments were digitally rendered on STAMP's proprietary assessment platform – "LinQ". Each student from the databases provided by the schools was assigned a unique ID and linked with the relevant test instrument to ensure assessment integrity. On completion of the digital rendering, the app was tested in the school environment. App testing encompassed clarity of visuals, correct rendering of questions and answer options, details of response capture, data validation and load testing in environments with varying connectivity.

3.3 OPERATIONALISING ASSESSMENTS

Resource allocation for assessments was underpinned by the goal of assurance of integrity and fair practice. For this purpose, protocols for observation and back-checks were developed to which all stakeholders adhered strictly.

STAMP and EH trained field personnel on the process for assessments. Additionally, the need to create a low stakes environment both for students as well as stakeholders in the school, was emphasized to obtain valid results.

After the training, the assessment platform was provided to the relevant District Coordinators and School Co-ordinators/ Computer Instructors through the cloud and downloaded to local devices. Depending on the availability and capacity of computer laboratories, assessments were conducted on computers in the ICT laboratories. Wherever this was not feasible, assessments were conducted on tablets.

Additional oversight was provided by EH personnel who visited at least 10% of schools in each state during the assessment process.

3.4 DATA CAPTURE AND TRANSFER

On completion of the assessments at each school, the student submissions were available on the assessment app as a 'read only – protected file' which were uploaded by the relevant implementing partner field personnel. Each file was uniquely identified by school name and school code. Subsequently, STAMP extracted the data from these files, processed it on their proprietary assessment engine and shared outcomes with GMI.

4 ANALYSIS METHODOLOGY

The submissions from students were used to score correct and incorrect responses. The total test scores were converted into percentage (%) correct for each student. All scores represented in this report are grade averages of percentage correct.

Percentage improvement is calculated using: [(End line% - Baseline %) / Baseline %)] X 100

Data was analysed using Microsoft Excel. IBM SPSS was used for higher order analysis. All the data transferred by STAMP to GMI was checked for completeness, accuracy, anomalies and a sample was also subjected to back-checks.

4.1 TESTING FOR COMPLETENESS

STAMP delivered a summary sheet accompanying every parcel of data transferred in Microsoft Excel. This summary sheet was also provided to EH. This was checked by EH for the counts of schools and students against field reports. GMI checked for the counts of schools and students on the data set received.

4.2 TESTING FOR ACCURACY

Computations on the data undertaken by STAMP were validated by GMI to ensure the accuracy of the transformed data.

4.3 DETECTION OF ANOMALIES

All data was checked to ensure absence of non-valid entries. For instance, in a case where all response options can take values of A, B, C, D or missing, a value of E would be an anomaly in the data and once/ if detected was duly reconciled. As a corollary, absence of a valid value from all records [for instance, absence of option C from all records for question 25 (example)] was also considered an anomaly and once/if detected was investigated to completion.

4.4 BACK-CHECKS ON THE DATA

A minimum of 10% of the records in the final data files was matched against the root data capture files to ensure data quality. Additionally, student muster rolls were recorded manually and transferred to a Microsoft Excel file. Every muster roll was duly back-checked prior, for generating a unique Student ID. Data between the Baseline and the End line was matched on unique Student ID at the state level to ensure a threshold 60% match. It has been observed, that 10% - 20% absenteeism of students is normal on the day of the assessment. Coupled with student transfers/drop-outs 60% match between the two data sets was stipulated for the RightToRead assessments.

Once the validity of data was established, data was analysed at various levels, following a top-down approach:

- State
- Grades within a State
- Schools within each grade
- Gender within a school

The Assessment results were consolidated into two groups: Grades 3 to 5 and Grades 6 and 7. The outcomes presented in this report adhere to this consolidated analysis.

5 ASSESSMENT OUTCOME

The assessments undertaken in the four states of Maharashtra, West Bengal, Gujarat and Punjab have reinforced that ReadToMe™ has a positive impact on English reading and comprehension among children undergoing the RightToRead program. Across the total of over 33,000 students assessed spanning five grades in the different states, students undergoing ReadToMe™ classes (Treatment) were consistently seen to score higher in the End line as compared to students who were not exposed to such a technology-enabled platform for English learning (Control).

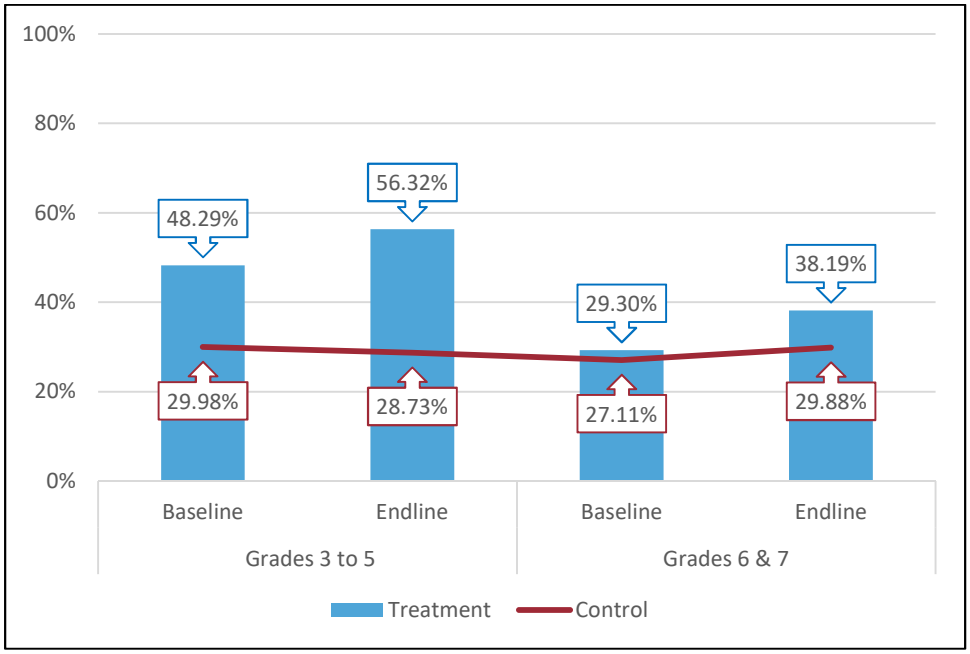
In the subsequent sections, we examine the outcome at an overall level followed by a state-wise analysis. Throughout our analysis, we review two cohorts – Grades 3 to 5, and Grades 6 and 7. State-wise analysis presents each grade assessed in that state as a cohort. Outcomes for Treatment groups are compared with those for Control groups.

5.1 OVERALL OUTCOME

The primary grades of 3 to 5 witnessed a 17% improvement in English scores in one academic year for the Treatment group as compared to a 4% decline among the Control group. Improvement in grades 6 and 7 was over 30% in the Treatment group as compared to 10% in the Control group.

Figure 4 illustrates the change in the mean percentage correct responses of students between Baseline and End line for the two cohorts – Grades 3 to 5, and Grades 6 and 7.

Figure 4: Change in Scores from Baseline to End line – Overall

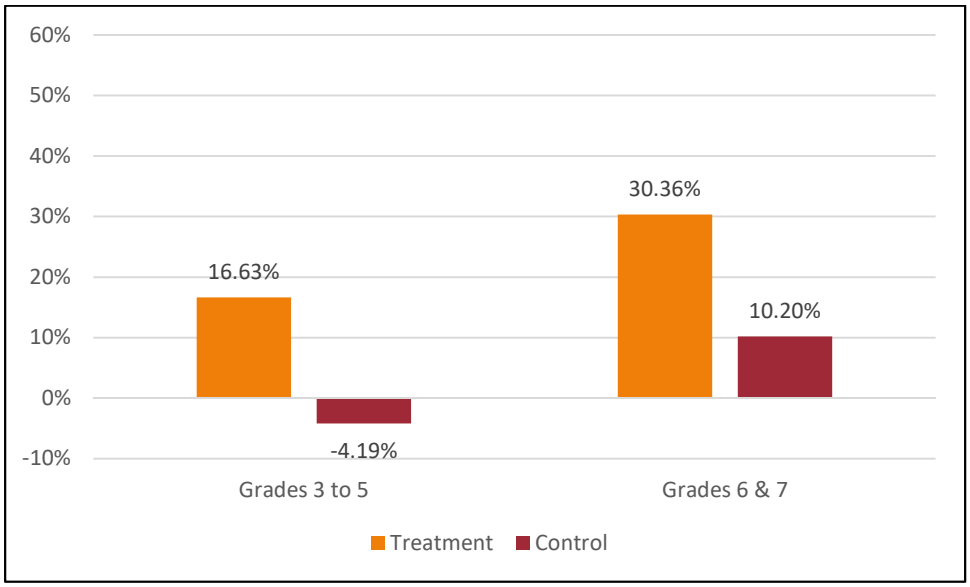


5.1.1 Improvement in Outcome

Improvement is calculated as change in scores between End line and Baseline as a percentage of Baseline scores.

Both cohorts witnessed a 20% improvement of the Treatment group compared to that of the Control group. Across the grade-within-state cohorts, we see improvement compared to Control ranging from 8% to 40%.

Figure 5: Improvement of Treatment and Control – Overall



5.1.2 Change by Quartile

Examination of the overall outcomes has established that there is improvement in the mean assessment scores between the Baseline and End line for both grade cohorts of the Treatment group.

It is equally important to examine the nature of this improvement across various quartiles of the students to determine whether students at all learning levels are benefitting from the program. Table 3 presents the change in the quartiles. The column labelled 'Change' indicates whether the limits of the quartile have increased (green upward arrow), remained unchanged (yellow side arrow) or declined (red downward arrow). It also presents the numerical difference between the End line and the Baseline to understand the extent of the change.

To illustrate, in Grades 3 to 5, the 75th percentile limit or the top one-fourth of students' scores has increased by 4.6 points from Baseline to End line, in the Treatment group. In comparison, the top one-fourth of students' scores in the Control group has declined by 5 points.

Table 3: Improvement in Quartiles – Overall

Grades 3 to 5						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	100.00%	100.00%	**0.00%**	85.00%	78.13%	**-6.88%**
Percentile 75	62.07%	66.67%	**4.60%**	42.50%	37.50%	**-5.00%**
Median	47.50%	50.00%	**2.50%**	32.50%	31.25%	**-1.25%**
Percentile 25	35.00%	37.50%	**2.50%**	20.00%	21.05%	**1.05%**
Minimum	0.00%	0.00%	**0.00%**	2.50%	0.00%	**-2.50%**

Grades 6 & 7						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	97.83%	100.00%	**2.17%**	86.96%	92.50%	**5.54%**
Percentile 75	44.68%	57.14%	**12.46%**	36.96%	38.30%	**1.34%**
Median	34.78%	40.00%	**5.22%**	29.79%	29.79%	**0.00%**
Percentile 25	26.09%	28.57%	**2.48%**	21.28%	22.86%	**1.58%**
Minimum	0.00%	0.00%	**0.00%**	0.00%	0.00%	**0.00%**

In both grade cohorts, the Treatment group exhibits increase in the scores in every quartile [75th percentile, Median (50th percentile) and 25th percentile]. In comparison, the Control group shows increase only in the 25th percentile, and to a lower extent than the Treatment group in Grades 3 to 5. While the Control group exhibits increase in each quartile in Grades 6 and 7, the increase is much lower than that in the Treatment group. The median however, has remained unchanged.

To summarize, the Treatment group exhibits greater increase than the Control group across all quartiles, indicating that ReadToMe^{TM} benefits students across all learning levels.

The following four sub-sections examine outcomes for individual grade cohorts in every state. It is important for the grade-within-state cohorts to exhibit patterns of improvement in line with that of overall outcomes to conclude that the program is effective across geographies and grade-levels.

5.2 Outcome In West Bengal

This section presents a grade-wise analysis of outcomes of the assessments in West Bengal. The analysis format is identical to that followed under **Section 5.1 Overall Outcome**. Please refer to that section for explanation and interpretation of the tables and figures presented in the state-wise analyses. In addition, the state-wise analyses also present the sample sizes per grade.

5.2.1 Sample Size

The sample sizes presented here reflect the sampling achieved after the End line.

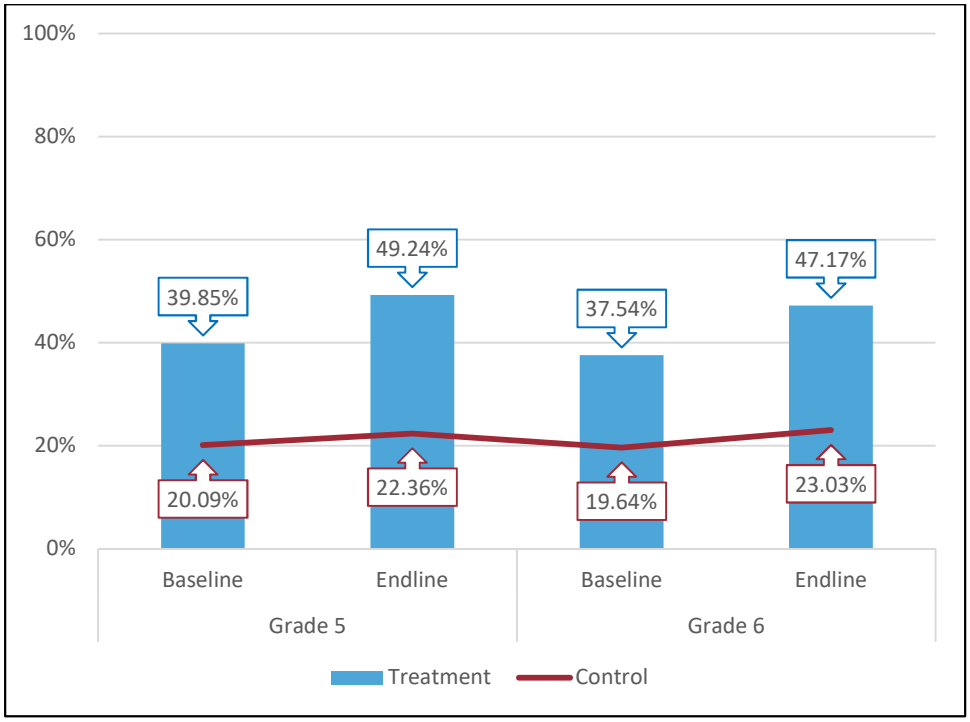
Table 4: Sample Size – West Bengal

	Grade 5		Grade 6	
	Treatment	**Control**	**Treatment**	**Control**
Number of students	2,106	400	1,984	402

5.2.2 Assessment Outcome – West Bengal

Treatment and Control groups in both grades exhibit an increase in scores from Baseline to End line.

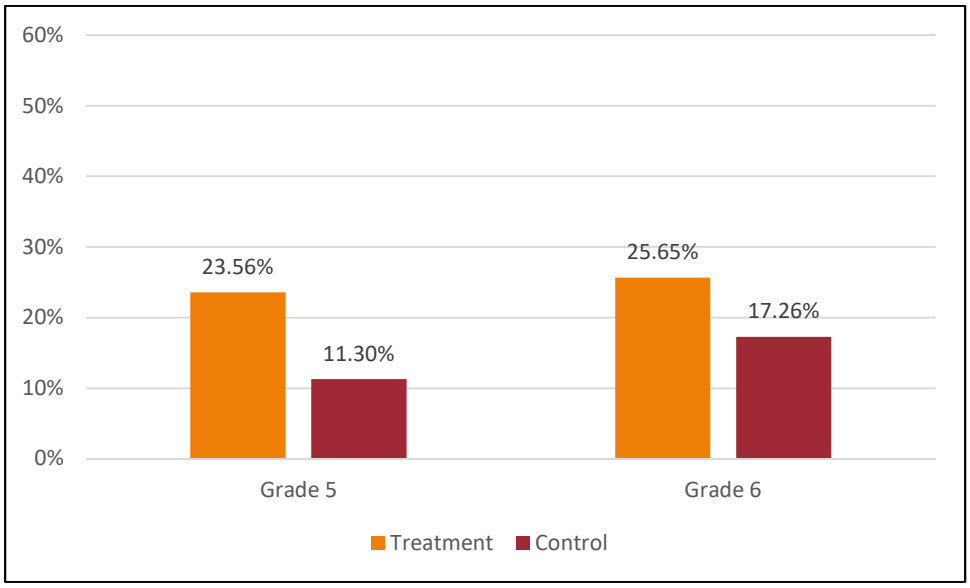
Figure 6: Change in scores from Baseline to End line – West Bengal



5.2.3 Improvement in Outcome – West Bengal

The Grade 5 Treatment group exhibits an improvement of 24% over Baseline scores (12% higher than the Control group); Grade 6 exhibits an improvement of 26% (8% higher than the Control group).

Figure 7: Improvement of Treatment and Control – West Bengal



5.2.4 Change by Quartile – West Bengal

The quartile bounds for both grades of the Treatment group exhibit increases ranging from 2% to 16%. These increases are much higher than the corresponding quartile bound increases for the Control group (the Maximum score achieved by both grades of the Control group having declined).

Table 5: Improvement in Quartiles – West Bengal

Grade 5						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	94.74%	100.00%	**5.26%**	81.58%	73.68%	**-7.89%**
Percentile 75	50.00%	65.79%	**15.79%**	26.32%	31.58%	**5.26%**
Median	36.84%	50.00%	**13.16%**	18.42%	21.05%	**2.63%**
Percentile 25	26.32%	34.21%	**7.89%**	10.53%	13.16%	**2.63%**
Minimum	2.63%	0.00%	**-2.63%**	2.63%	0.00%	**-2.63%**

Grade 6						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	95.83%	97.87%	**2.04%**	81.25%	72.34%	**-8.91%**
Percentile 75	47.92%	63.83%	**15.91%**	25.00%	29.79%	**4.79%**
Median	33.33%	42.55%	**9.22%**	18.75%	23.40%	**4.65%**
Percentile 25	22.92%	29.79%	**6.87%**	12.50%	14.89%	**2.39%**
Minimum	2.08%	0.00%	**-2.08%**	2.08%	0.00%	**-2.08%**

5.3 Outcome In Maharashtra

This section presents a grade-wise analysis of outcomes of the assessments in Maharashtra. The analysis format is identical to that followed under **Section 5.1 Overall Outcome**. Please refer to that section for explanation and interpretation of the tables and figures presented in the state-wise analyses. In addition, the state-wise analyses also present the sample sizes per grade.

5.3.1 Sample Size

The sample sizes presented here reflect the sampling achieved after the End line.

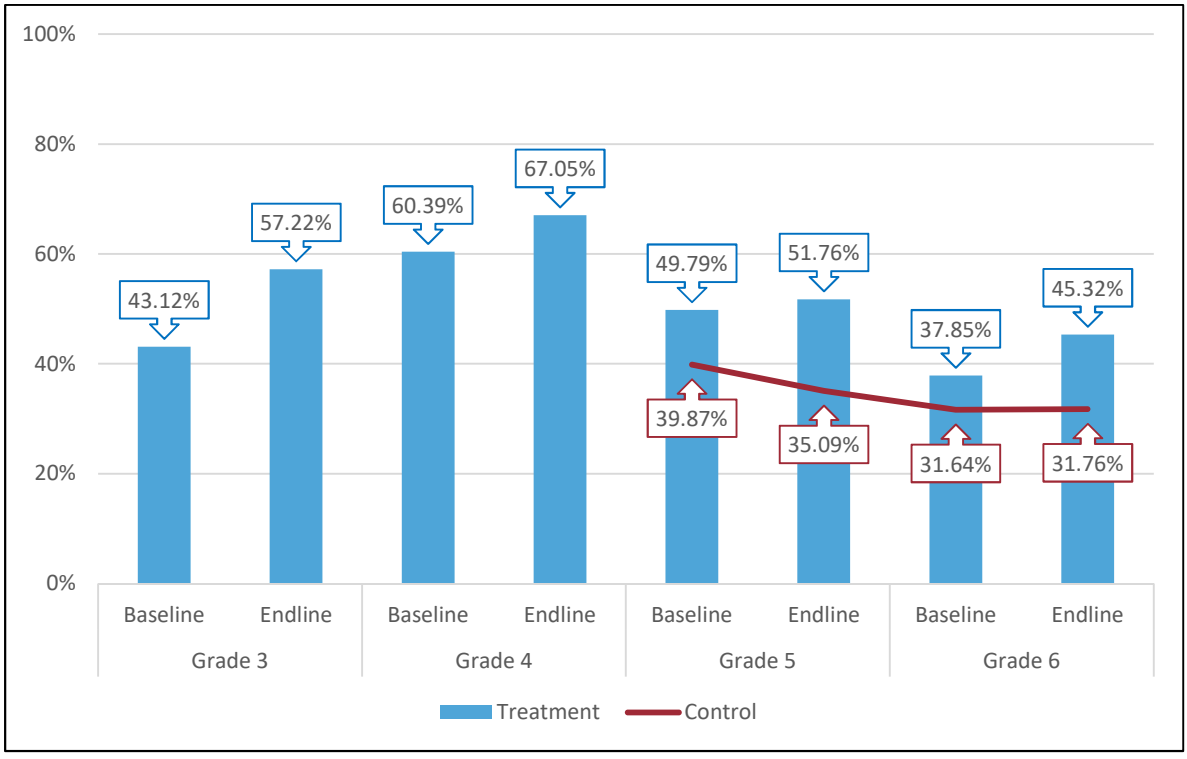
Table 6: Sample Size – Maharashtra

	Grade 3	Grade 4	Grade 5		Grade 6	
	Treatment	**Treatment**	**Treatment**	**Control**	**Treatment**	**Control**
Number of students	1,005	1,049	10,147	629	11,787	610

5.3.2 Assessment Outcome – Maharashtra

The Treatment group of all the grades shows an increase from Baseline to End line. Grade 5 Control exhibits a decline from Baseline to End line while Grade 6 Control remains unchanged.

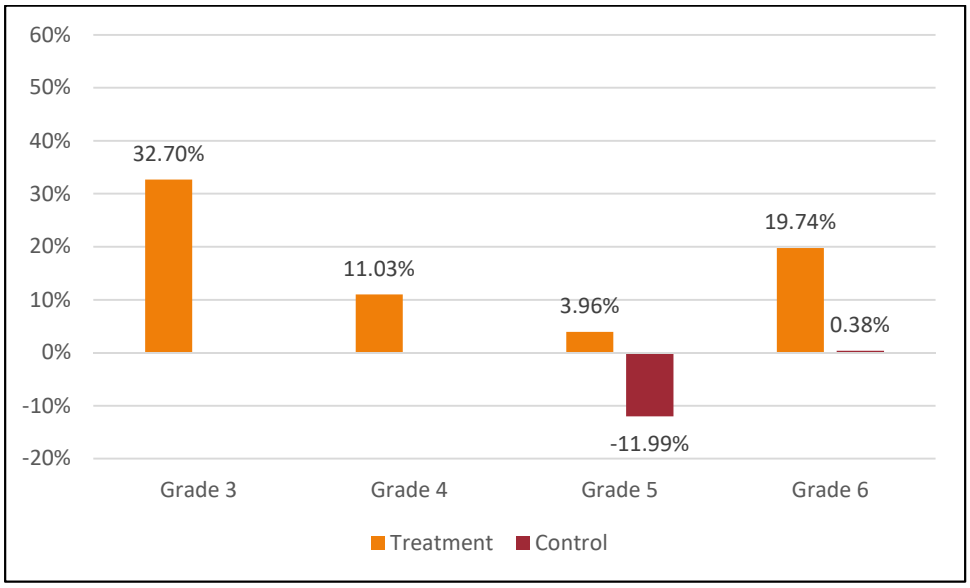
Figure 8: Change in scores from Baseline to End line – Maharashtra



5.3.3 Improvement in Outcome – Maharashtra

All grade cohorts of the Treatment group exhibit improvement, with Grade 3 exhibiting higher than the overall mean improvement. The improvement in Grade 5 appears low at 4%, but should be interpreted considering the decline of 12% in the Control group.

Figure 9: Improvement of Treatment and Control – Maharashtra



5.3.4 Change by Quartile – Maharashtra

The quartile bounds for Grade 3 show almost uniformly high improvement indicating that the improvement is spread across all learning levels. In Grade 4, the lower quartile bound (25th percentile) shows maximum improvement; the program has benefitted the lowest learning levels most in this group. The 25th percentile of Grade 5 remains unchanged, while that for Grade 6 exhibits only marginal improvement, indicating potential for improvement at this level. Control groups in grades 5 and 6 (except the median in Grade 6) show a decline from Baseline levels.

Table 7: Improvement in Quartiles – Maharashtra

Grade 3			
	Treatment		
	Baseline	Endline	Change
Maximum	89.66%	96.67%	**7.01%**
Percentile 75	58.62%	76.67%	**18.05%**
Median	41.38%	60.00%	**18.62%**
Percentile 25	24.14%	36.67%	**12.53%**
Minimum	0.00%	0.00%	**0.00%**

Grade 4			
	Treatment		
	Baseline	Endline	Change
Maximum	100.00%	100.00%	**0.00%**
Percentile 75	79.31%	83.33%	**4.02%**
Median	65.52%	70.00%	**4.48%**
Percentile 25	41.38%	53.33%	**11.95%**
Minimum	0.00%	0.00%	**0.00%**

Grade 5						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	100.00%	100.00%	**0.00%**	85.00%	78.13%	**-6.88%**
Percentile 75	60.00%	65.63%	**5.63%**	47.50%	40.63%	**-6.88%**
Median	47.50%	50.00%	**2.50%**	37.50%	34.38%	**-3.13%**
Percentile 25	37.50%	37.50%	**0.00%**	30.00%	28.13%	**-1.88%**
Minimum	0.00%	0.00%	**0.00%**	2.50%	3.13%	**0.63%**

Grade 6						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	97.83%	100.00%	**2.17%**	86.96%	74.29%	**-12.67%**
Percentile 75	45.65%	60.00%	**14.35%**	39.13%	37.14%	**-1.99%**
Median	34.78%	40.00%	**5.22%**	30.43%	31.43%	**0.99%**
Percentile 25	28.26%	28.57%	**0.31%**	26.09%	25.71%	**-0.37%**
Minimum	0.00%	0.00%	**0.00%**	0.00%	0.00%	**0.00%**

5.4 Outcome In Punjab

This section presents a grade-wise analysis of outcomes of the assessments in Punjab. The analysis format is identical to that followed under **Section 5.1 Overall Outcome**. Please refer to that section for explanation and interpretation of the tables and figures presented in the state-wise analyses. In addition, the state-wise analyses also present the sample sizes per grade.

5.4.1 Sample Size

The sample sizes presented here reflect the sampling achieved after the End line.

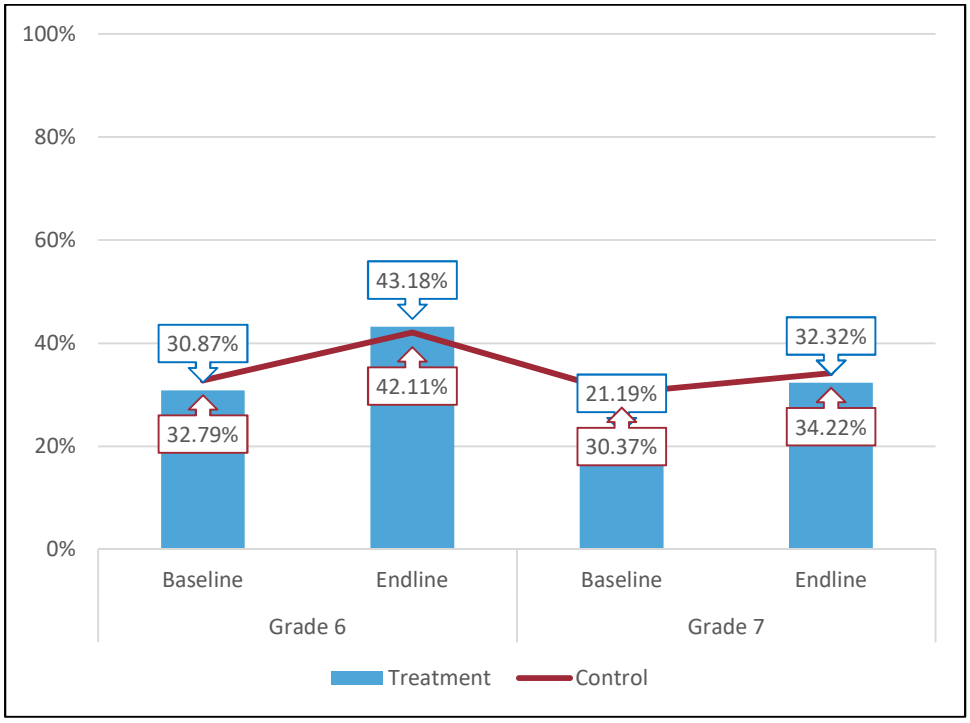
Table 8: Sample Size – Punjab

	Grade 6		Grade 7	
	Treatment	**Control**	**Treatment**	**Control**
Number of students	961	199	1,194	259

5.4.2 Assessment Outcome – Punjab

In the Baseline, the Treatment groups of both grades 6 and 7 had lower scores than the corresponding Control groups. However, by the End line, the Treatment group of Grade 6 has scored higher than the Control group and the Treatment group of Grade 7 is lower than the Control group by only 2%.

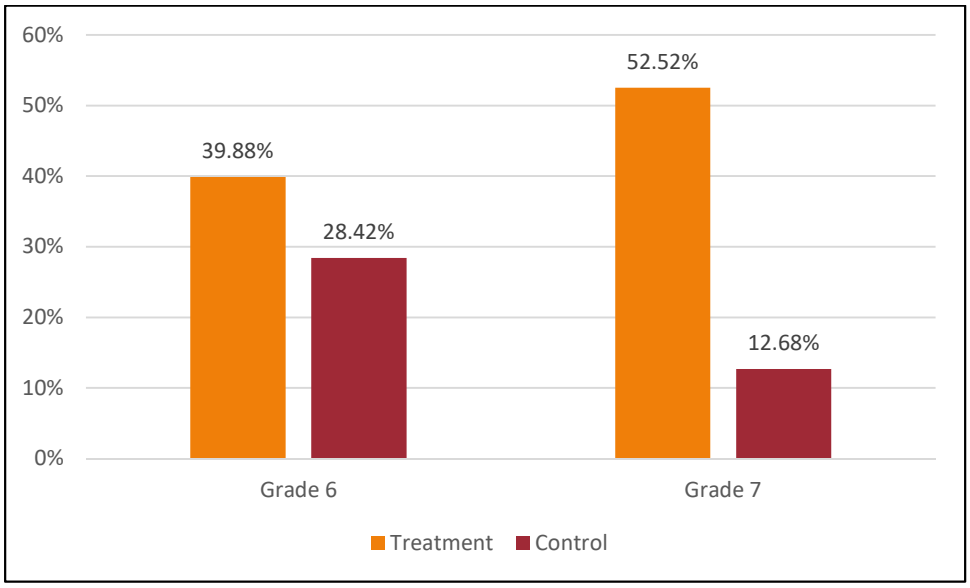
Figure 10: Change in scores from Baseline to End line – Punjab



5.4.3 Improvement in Outcome – Punjab

The Grade 6 Treatment group exhibits an improvement of 40% over Baseline scores (11% higher than the Control group). Grade 7 exhibits an improvement of 53% (40% higher than the Control group); the highest across all state and grade cohorts.

Figure 11: Improvement of Treatment and Control – Punjab



5.4.4 Change by Quartile – Punjab

Grade 6 exhibits uniform increases across all quartiles indicating the program has benefitted all students equally. Grade 7 of the Treatment group exhibits higher increases than the Control group except in the 25th percentile (where the increase in Treatment was marginally lower than that in Control).

Table 9: Improvement in Quartiles – Punjab

Grade 6						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	85.00%	95.00%	**10.00%**	65.00%	92.50%	**27.50%**
Percentile 75	42.50%	52.50%	**10.00%**	42.50%	50.00%	**7.50%**
Median	32.50%	42.50%	**10.00%**	35.00%	42.50%	**7.50%**
Percentile 25	20.00%	30.00%	**10.00%**	27.50%	32.50%	**5.00%**
Minimum	2.50%	0.00%	**-2.50%**	2.50%	0.00%	**-2.50%**

Grade 7						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	76.60%	76.09%	**-0.51%**	80.85%	71.74%	**-9.11%**
Percentile 75	31.91%	47.83%	**15.91%**	36.17%	43.48%	**7.31%**
Median	25.53%	36.96%	**11.42%**	31.91%	36.96%	**5.04%**
Percentile 25	17.02%	23.91%	**6.89%**	23.40%	30.43%	**7.03%**
Minimum	2.13%	0.00%	**-2.13%**	2.13%	2.17%	**0.05%**

5.5 Outcome In Gujarat

This section presents a grade-wise analysis of outcomes of the assessments in Gujarat. The analysis format is identical to that followed under **Section 5.1 Overall Outcome**. Please refer to that section for explanation and interpretation of the tables and figures presented in the state-wise analyses. In addition, the state-wise analyses also present the sample sizes per grade.

5.5.1 Sample Size

The sample sizes presented here reflect the sampling achieved after the End line.

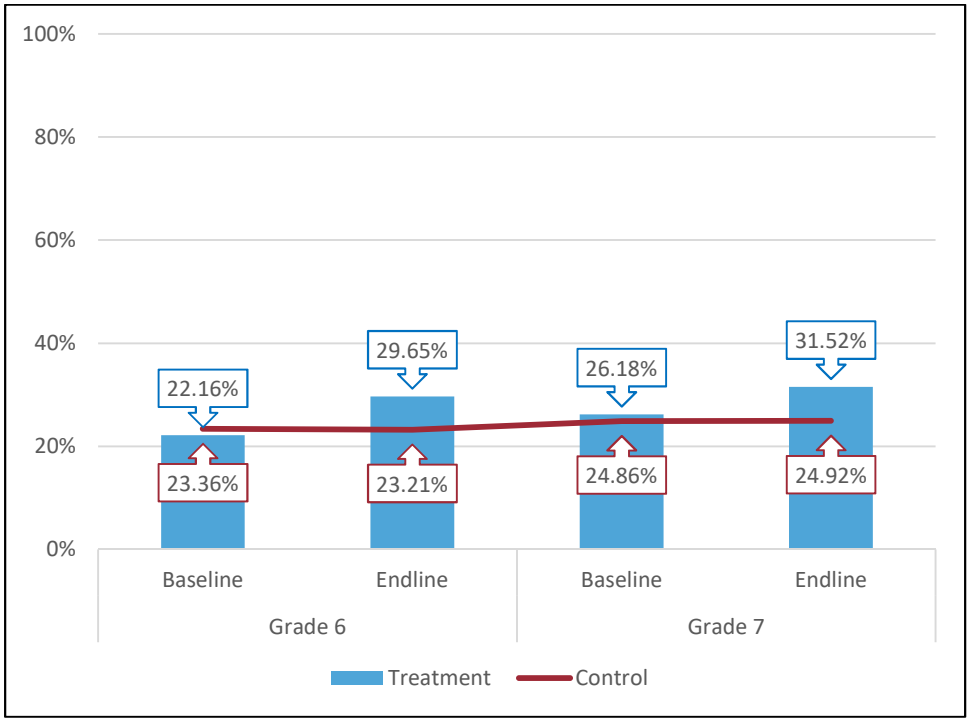
Table 10: Sample Size – Gujarat

	Grade 6		Grade 7	
	Treatment	**Control**	**Treatment**	**Control**
Number of students	480	96	446	111

5.5.2 Assessment Outcome – Gujarat

In the baseline the Treatment group of Grade 6 scored lower than the Control group; by the End line the Treatment group scored higher. The Control group has remained largely unchanged in both grades while Treatment group shows significant positive change in scores.

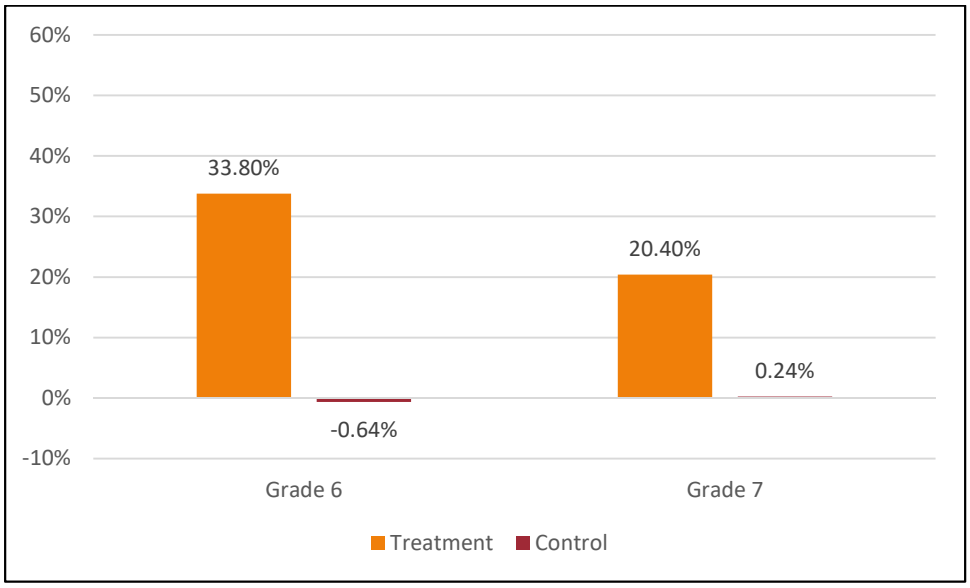
Figure 12: Change in scores from Baseline to End line – Gujarat



5.5.3 Improvement in Outcome – Gujarat

The Grade 6 Treatment group exhibits an improvement of 34% over Baseline scores (34% higher than the Control group); Grade 7 Treatment group exhibits an improvement of 20%. Little to negative improvement is exhibited by the Control groups.

Figure 13: Improvement of Treatment and Control – Gujarat



5.5.4 Change by Quartile – Gujarat

Highest increases were observed, when compared with other states, in the Maximum scores across both grades in the Treatment group. In every quartile bound, Treatment exhibits a higher increase than Control (or, in the case of the Minimum in Grade 6, a lower decline).

Table 11: Improvement in Quartiles – Gujarat

Grade 6						
	Treatment			**Control**		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	50.00%	88.57%	**38.57%**	50.00%	45.71%	**-4.29%**
Percentile 75	28.57%	37.14%	**8.57%**	30.95%	28.57%	**-2.38%**
Median	23.81%	28.57%	**4.76%**	26.19%	22.86%	**-3.33%**
Percentile 25	19.05%	20.00%	**0.95%**	21.43%	17.14%	**-4.29%**
Minimum	2.38%	0.00%	**-2.38%**	9.52%	0.00%	**-9.52%**

Grade 7						
	Treatment			Control		
	Baseline	Endline	Change	Baseline	Endline	Change
Maximum	70.21%	80.95%	**10.74%**	44.68%	57.14%	**12.46%**
Percentile 75	31.91%	35.71%	**3.80%**	31.91%	28.57%	**-3.34%**
Median	27.66%	30.95%	**3.29%**	27.66%	26.19%	**-1.47%**
Percentile 25	23.40%	23.81%	**0.41%**	23.40%	19.05%	**-4.36%**
Minimum	2.13%	7.14%	**5.02%**	6.38%	2.38%	**-4.00%**

6 THE CEFR ANALYSIS

To maintain contextual relevance of the assessment instruments with the text book, separate instruments were developed for each grade in each state. This has resulted in multiple sets of instruments albeit, adhering to a standard rubric.

It is necessary to be able to synthesize the outcomes from each of the assessments (across grades and states) to evaluate the overall impact of the RightToRead program using the ReadToMe™ platform. EH has adopted the Common European Framework of Reference (CEFR) as a benchmark to measure the outcomes. The CEFR is also intended as the benchmark for future assessments, providing a common platform for comparison across segments, geographies and over time.

The following process, undertaken in the order mentioned, was adopted to measure the learning outcomes aligned with the CEFR.

- Adapting and establishing the mapping framework
- Assigning a CEFR level to every question
- Assigning a CEFR level to every student
- Comparing the CEFR distribution of students in the Baseline and in the End line

6.1 ADAPTING THE CEFR MAPPING FRAMEWORK

The adaptation of the CEFR for the RightToRead assessments is driven by and dependent on the following factors:

- The design of the assessment instruments was primarily driven by contextual relevance for the student. To that effect, all the text and reading comprehension stimuli (passages) were familiar to students.
- The objective of the assessments was to test students' learning at a fundamental level, considering that most students are first-generation learners.
- Students were tested only for their reading skills. To that extent, the mapping relates to the Reading component of the Common European Framework of Reference for Languages (CEFR).

The CEFR describes language proficiency (related to listening, speaking, reading and writing) on a six-level scale:

- A1-A2 for Basic User
- B1-B2 for Independent User
- C1-C2 for Proficient User

The CEFR defines specific competencies of a language learner at each of these levels in the form of "Can do" statements. It also allows for branching and defining sub-competencies, such as A1.1 and A1.2. Considering that there is an aggregation of assessment objectives mapping to the A1 level in the RightToRead assessments, especially in Reading Comprehension, two branches for reading under the CEFR A1 level were defined as A1.1 and A1.2. Consequently, the overall adapted framework for the assessments reads as follows:

Table 12: Adapted CEFR for Overall Reading Comprehension

	OVERALL READING COMPREHENSION
C2	*Can understand and interpret critically virtually all forms of the written language including abstract, structurally complex, or highly colloquial literary and non-literary writings.* *Can understand a wide range of long and complex texts, appreciating subtle distinctions of style and implicit as well as explicit meaning.*
C1	*Can understand in detail lengthy, complex texts, whether or not they relate to his/her own area of speciality, provided he/she can reread difficult sections.*
B2	*Can read with a large degree of independence, adapting style and speed of reading to different texts and purposes, and using appropriate reference sources selectively. Has a broad active reading vocabulary, but may experience some difficulty with low-frequency idioms.*
B1	*Can read straightforward factual texts on subjects related to his/her field and interest with a satisfactory level of comprehension.*
A2	*Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language*
	Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items.
A1	*Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required.*
A1.2	Can interpret information in familiar text
	Can link given information to locate details in a text
	Can comprehend vocabulary in context
	Can synthesize information and makes simple inferences from familiar and different types of text
A1.1	Can recognise first letter from a familiar picture, missing letters from a familiar word, repeated letters
	Can complete sentences meaningfully by recognising missing words
	Can identify name of a given picture
	Can identify synonyms and antonyms of familiar words
	Can identify contrasting words
	Can retrieve explicitly stated information from a familiar text

6.2 Assigning CEFR Levels To Questions

The English Profile Project (www.englishprofile.org) funded by Cambridge University Press and Cambridge English Language Assessment, among others, has compiled a list of words with their associated CEFR levels and a list of grammatical forms that are used by students at various CEFR levels. These are called the English Vocabulary Profile (http://www.englishprofile.org/wordlists) and the English Grammar Profile (http://www.englishprofile.org/english-grammar-profile), respectively. These have been used as the fundamental guiding principles when assigning CEFR levels to questions that satisfy the Letter Recognition, Word Recognition and Vocabulary constructs in the assessment instruments.

All Reading Comprehension questions in the assessment instruments were assigned a CEFR level using the "can do" statements presented in

Table 12.

Table 13: The CEFR Composition of Assessment Instruments

	Grades 3 to 5	Grades 6 & 7
A1.1	48.22%	35.39%
A1.2	25.89%	19.15%
A2 and above	25.89%	45.46%

6.3 Assigning CEFR Levels To Students

Each student was assessed on the level of achievement at every CEFR level. This was measured as the proportion (percentage) of questions that a student answered correctly for each CEFR level. For a student to be deemed as having achieved a CEFR level, the student should have scored more than 50% of the questions, at that level, correctly. Thus, a single CEFR level was assigned to each student.

6.4 Comparing The CEFR Distribution Between Baseline And End line

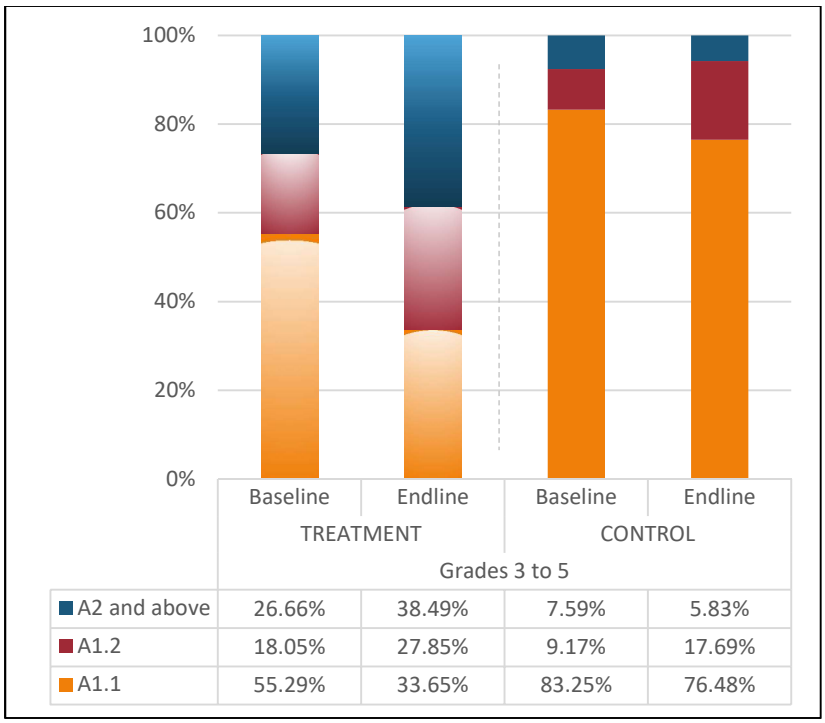
Having assigned a CEFR level to every student, the distribution of students across the CEFR levels of the Baseline was compared with that of the End line for both the Treatment and the Control groups.

The Treatment group, in Grades 3 to 5, exhibits a clear progression of students from A1.1 to A1.2 and A2 and above, from Baseline to End line. The corresponding Control group exhibits a comparatively lower progression in the percentage of students with progression primarily to the A1.2 level; percentage in A2 and above having dipped from the Baseline. More than 21% of the students have moved from A1.1 to the higher CEFR levels in the Treatment group, while fewer than one-third that number (7%) have done so in the Control group. Additionally, upward movement in the Treatment group continues through the spectrum, with numbers in the A2 and above level increasing by 8%. However, these numbers decline in the Control group. The Treatment group of Grades 6 & 7 also shows a marked progression to the A2 and above level in the End line (with over 18% progressing); the Control group shows very little change in comparison, from Baseline to End line.

The nature of these results is consistent with the assessment outcomes outlined in Section 5, and indicate positive impact of the RightToRead program using the ReadToMe™ platform, integrated with the curriculum, on the English learning of students.

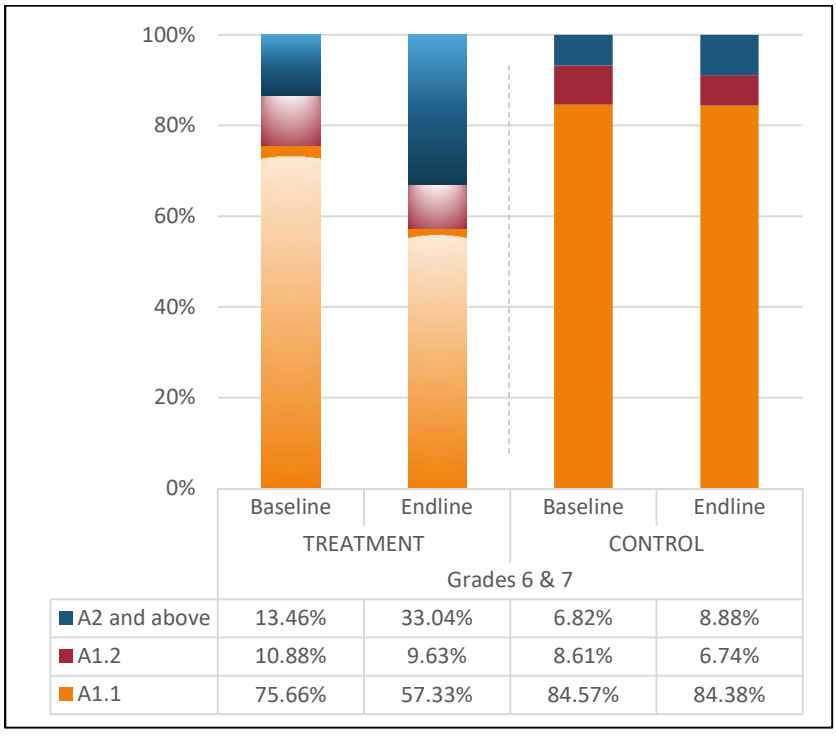
Figure 14 and Figure 15 illustrate the change in the CEFR achievements of students from Baseline to End line. An increase in height in the blue and brown bars indicates improvement in the higher CEFR levels.

Figure 14: Comparison of the CEFR Achievement between Baseline and End line – Grades 3 to 5



	Baseline	Endline	Baseline	Endline
	TREATMENT		CONTROL	
	Grades 3 to 5			
A2 and above	26.66%	38.49%	7.59%	5.83%
A1.2	18.05%	27.85%	9.17%	17.69%
A1.1	55.29%	33.65%	83.25%	76.48%

Figure 15: Comparison of the CEFR Achievement between Baseline and End line – Grades 6 & 7



	Baseline	Endline	Baseline	Endline
	TREATMENT		CONTROL	
	Grades 6 & 7			
A2 and above	13.46%	33.04%	6.82%	8.88%
A1.2	10.88%	9.63%	8.61%	6.74%
A1.1	75.66%	57.33%	84.57%	84.38%

7 CONCLUSION

The assessments for the RightToRead program covered more than 33,000 students for each test i.e. almost 70,000 tests across four states were conducted during Baseline and End line assessments. These assessments were deployed to provide representation of the impact of the RightToRead program which has reached more than a million students. The assessments included a Control group to identify comparable improvement in English reading and comprehension of students not exposed to technology-enabled reading of their English text books.

These independent assessments encompassing four key parameters in English provide a true picture of the impact of the ReadToMe™ platform on the English learning outcomes of students. The primary grades of 3 to 5 witnessed a 21% higher improvement in English scores in an academic year for the Treatment group as compared to the Control group. Improvement in grades 6 and 7 was higher by over 20% in the Treatment group as compared to the Control group. Overall, students undertaking ReadToMe™-enabled classes demonstrate between 8% and 40% higher improvement in scores as compared to students who have not been exposed to the program.

Improvement was observed across all levels consistently compared with the Control group, indicating that ReadToMe™ positively impacts learning across grades (ages) and for students at all levels of English proficiency.

The CEFR analysis enables a standardised comparison of students across grades and regions by allotting a CEFR level to each student. The Treatment group shows reading improvement across all levels. This analysis will now enable the comparison of the CEFR level of students with the required proficiency prescribed by curriculum.

In conclusion, the assessments validate that the use of the ReadToMe™ platform has a positive impact on English reading and comprehension of students undergoing the RightToRead program. In addition, the process provides valuable insights for enabling proficiency gap identification for leveraging RightToRead learning benefits.

Appendix – Assessment Instruments

The assessment instruments developed by GMI were grade- and state-specific to maintain contextual relevance for students. All instruments followed a standard rubric appropriate for each grade level. This enabled examination of the reading proficiency of students across various segments and over time.

The instruments, consisting of 40 questions, on an average across grades and states, were designed to test students on four parameters:

- Letter Recognition (4 questions)
- Word Recognition (4 questions)
- Vocabulary (12 questions)
- Reading Comprehension – of two levels of complexity:
    - focus on retrieval of information (15 questions), and
    - focus on synthesis and inference (5 questions)

All instruments were piloted before deploying them on field. The final instruments were sent to Skill Training Assessment Management Partners (STAMP) with the questions, answer options and answer keys for integration into the assessment platform.

Each parameter has been briefly explained below and illustrated with a sample question. The answer key to the question is in *italics.*

1. LETTER RECOGNITION

- Test the ability of students to identify missing letters, repeated letters and silent letters in a word
- Presented in word form or as a sentence
- May or may not be supported by a visual

Write the letter that is **silent** in the word given in the box. (Grade 5, West Bengal, Baseline)

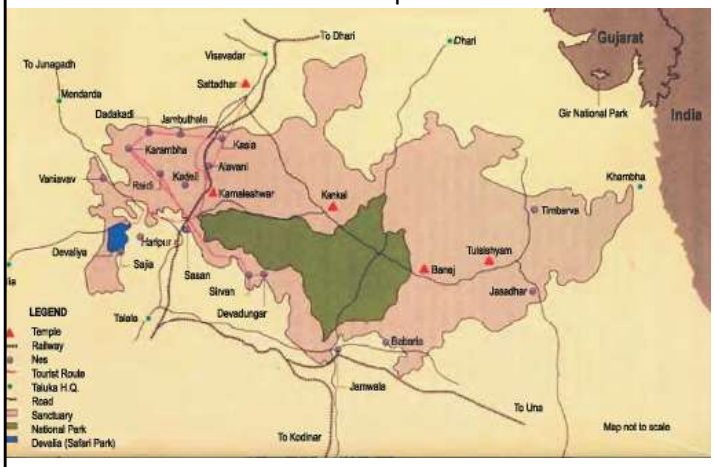
KNIFE

A. I
B. E
C. *K*
D. F

2. WORD RECOGNITION

- Test the ability of students to identify a word
- May be supported by a visual
- Or, require selection of outlier word out, given a list of words

Choose the correct word for the picture. (Grade 7, Gujarat, End line)



A. outline

B. *map*

C. plan

D. graph

3. VOCABULARY

- Test the ability of students to identify a word that completes a sentence meaningfully in various grammatical contexts
- Comprehend the synonym or antonym of a given word
- Understand vocabulary in context
- May or may not be supported by a visual

The word “sharpen” cannot be used for which of the following? (Grade 6, Maharashtra, Baseline)

A. a sword

B. a pencil

C. a knife

D. *a bottle*

4. READING COMPREHENSION

- Test the ability of students to
- Retrieve explicitly stated information from a text
- Locate detail in a text in the presence of competing information
- Make a simple inference from a narrative
- Synthesize information from a dialog text

A clean confession

A relative and I became very fond of smoking. Not that we saw any good in smoking, or liked the smell of a cigarette. We simply imagined a sort of pleasure in sending out clouds of smoke from our mouths. My uncle had the habit, and we thought we should copy his example. But we had no money. So we began stealing stumps of cigarettes thrown away by my uncle.

The stumps, however, were not always available and could not give out much smoke either. So we began to steal coppers* from the servant's pocket money in order to purchase cigarettes. But the question was where to keep them. We could not of course smoke in the presence of elders. We managed somehow for a few weeks on these stolen coppers.

Why did the boy and his relative smoke? (Grade 7, Punjab, End line)

A. They had no money.
B. They liked the smell of cigarettes.
C. They thought smoking is good.
D. *They liked sending out clouds of smoke.*

EnglishHelper™

Learn at your pace, in your context

EnglishHelper Education Technologies Pvt. Ltd.

Suite # 301-323 V3S Mall, Mangalam Place Sector -3, Rohini, New Delhi – 110085

info@englishhelper.com

© EnglishHelper™ 2017. All Rights Reserved