Improving Learning Outcomes in K-12:
Evidence from a Large-Scale Educational Technology Intervention in India

Venkat Srinivasan
Hemavathi Murthy

EnglishHelper Inc.
August 2018

We acknowledge the helpful comments of M.S. Vijay Kumar, Assoc Dean for Open Learning at MIT. We thank the United States Agency for International Development (USAID) for partially funding this project. We thank Priya Viswanath and Vineet Mehra from the Program Management Office of RightToRead, EnglishHelper for their support of the program. We thank IL&FS Education and American India Foundation (AIF) for being the lead implementation partners. We acknowledge Gray Matters India (GMI) for their instrument design and analysis. We thank Skill Training Assessment Management Partners (STAMP) for providing the technology platform that ensured unbiased data collection and handling. All views expressed are those of the authors and not of any of the institutions they are affiliated with, or of USAID, AIF, IL&FS Education, GMI or STAMP.

Abstract

We present evidence on the impact of technology-aided instruction on learning outcomes for English language learning using an AI-based multi-sensory technology across a large cross-section of government schools in India. Previous studies have reported the impact of technology-aided instruction but with proprietary instructional design and pedagogy. The intervention we report in this study enhances the instructional effectiveness of the teachers and the learning ability of the children within the existing instructional framework and content without any new instructional design or pedagogy or content. Besides, the intervention is implemented by existing teachers not outside volunteers.

A total of 1 million children and 15,000 teachers across 5,000 government schools in Maharashtra, West Bengal, Punjab, Tamil Nadu, Telangana, Gujarat and Karnataka used the technology for the 2016-17 academic year. Using a randomized control assessment, we find a 20-40% overall gain in learning outcomes in the treatment sample. Gains within individual states and grades vary. We find that the distribution of learning outcomes rise across the entire range of proficiency levels in a grade. Ongoing assessments report even higher impact in the 50-60% range. In addition, teachers also reported improving their skills as a result of using the technology, suggesting that the intervention can also alleviate the issue of teacher shortage and inadequate teacher training. The results hold significant promise for disrupting the low and stagnating learning levels across government schools in India and other similar environments.

JEL Codes: I25, O35, O33

1 Introduction

In developing economies like India, English proficiency takes on a whole new meaning and social significance. English is clearly 'aspirational' in such countries. It is tied to economic and social well-being. Several studies document the differential income generating capacity of persons with superior English language skills. The role of technology as a resource for instruction of foreign language is increasing as educators recognize its ability to create both independent and collaborative learning environments for students to acquire and practice a new language. A large number of studies have documented the impact of different technologies on English language learning [ELL] populations generally concluding that the use of technologies have been beneficial.

We should recognize that 'technology' encompasses a wide range of tools, artifacts and practices, from multimedia computers to the internet, from videotapes to chat rooms, from web pages to interactive audio conferencing and more. Liu, et al (2001) reviewed the research on computer-based technology use in second language learning during 1990-2000. They found that findings from numerous studies suggested that the use of visual media supported vocabulary acquisition and reading comprehension, and helped increase achievement scores. They also found that more research needs to be conducted at the K-12 level. Ban et al (2000) review research on the use of technology for English language learners specifically in Content-based classrooms. They argue that it is no longer the question of whether computer technologies have the potential to help ELL students develop English language proficiency. The outstanding issue according to them is how to effectively integrate technology into content-based classroom pedagogy.

Many other researchers have attempted to provide a comprehensive review of studies documenting the impact of technology on language learning [e.g., Cavanaugh, 2001; Chapelle, 1997; Lou, Abrami and d'Apollonia, 2001; Salaberry, 2001]. Zhao points out that there has also been a major paradigm shift in the pedagogical research focus of technology applications in language education recently [Chapelle, 1997, 2001; Pennington, 1996; Salaberry, 2001] – a shift away from traditional drill-and-skill computer-aided instruction (CAI) models toward multimedia, intelligent CAI, and integration models. Zhao (2003) applied meta-analysis on all the qualified empirical studies included in his review and reported a significantly positive impact of technology applications on language learning.

Most of the above studies suffer from several significant limitations. First, the sample sizes of all the studies are relatively small and suffer from a selection bias. Second, there are very few studies that focus on K-12 students; most focus on college learners. Zhao (2003) makes the observation that none of the studies found in the major language education and technology journals is about technology use in K-12 classrooms whereas most such studies in other subject areas (mathematics, science, social studies and language arts) have taken place mainly in K-12 settings. The two populations have very different motivations and at very different stage in their evolution. Third, it seems like in

many cases the instructors designed, implemented and evaluated the assessments, which of course potentially makes the results questionable.

More recently, several experimental studies have shown that simple pedagogical changes can lead to significant improvements in learning levels [Banerjee et al 2007, Duflo et al 2011). These interventions however have largely relied on NGO staff or volunteers to effect the intervention from outside the government school system. Banerjee et al [2016] recognize that to scale such interventions, it will be obviously necessary to implement them within the school system and by teachers not outside volunteers or NGO staff. They report their evaluation of the efforts by the NGO Pratham to scale within the government school system in four Indian States. They find that two scale-up models were effective, with gains in language of 0.14 standard deviation in Haryana, and 0.70 standard deviation in Uttar Pradesh, on all students enrolled in these schools at baseline. Banerjee et al [2016] also point out that it was hard to achieve the impact and change in two other Indian States despite the well-received training sessions and the NGO's support.

In 2013, we effected a multi-sensory technology-based intervention for English language learning which did not require a significant change in instructional approach, no change in pedagogy and implemented by the teachers themselves. This intervention was extended to 20,000 students in 100 government schools across 8 States in India. The intervention demonstrated significant improvement in reading and comprehension levels of students. Encouraged by these findings, we scaled the intervention in 2015-16 to cover 1 million children in 5,000 government schools across 8 States. To properly assess the impact of the intervention, we engaged an independent agency to conduct an assessment using a randomized control-treatment design. The results indicate that multi-sensory technology is highly effective and scalable in English language learning even in highly constrained, infrastructurally challenged settings. The intervention has the potential to radically alter educational outcomes and can be scaled to cover entire populations.

The study is pioneering in many ways. First, it is the first large scale study of the impact of technology for English language learning and specifically reading skills, in a K-12 setting. Second, it supports significant research findings that multisensory methods can be effective in the acquisition of reading skills. Third, the study was conducted in government schools across India in settings that have poor infrastructure, i.e., many students sharing a single computer, limited access to the application, poor classroom facilities, and so on. Fourth, the students who were part of the study have little or no exposure to English outside the classroom. Fifth, the study period was sufficiently long to enable robust conclusions about the impact of the software application.

2 The RightToRead Intervention

The RightToRead intervention was launched by EnglishHelper in 2013 based on its multisensory technology platform for reading and comprehension called ReadToMe™ [RTM].

Multi-Sensory Methods for Language Acquisition

Recent studies from neuroscience suggest that multi-sensory structured learning education [MSLE] has significant potential to improve the systematic acquisition of reading skills. These studies have found that the human brain was not created to recognize the letter-speech sound combinations required for reading fluency. Letter-speech sound variations are arbitrary cultural inventions. The brain creates a specialized neural pathway for recognizing such arbitrary objects. Multi-sensory stimulation enables the speedy creation of such a neural network.

Shams and Seitz (2008) point out that studies of learning and in particular perceptual learning have typically focused on learning of stimuli consisting of a single sensory modality. However, we constantly experience multisensory stimulation in the real world. It is easier to integrate multiple sources of information during learning when the material is physically integrated, auditorily and visually, than when information is presented to each modality separately [Mousavi, Low & Sweller (1995)]. It appears that multi-sensory information processing is part and parcel of object perception and recognition in daily life, whereby the brain integrates the information from different modalities into a coherent percept [Ghazanfar and Schroeder, 2006]. Therefore, it is likely that the human brain has evolved to develop, learn, and operate optimally in multisensory environments. These studies suggest that multisensory training protocols can better approximate natural settings and are more effective for learning.

Years after children first learn to decode letters into words, a form of perceptual expertise emerges in which groups of letters are rapidly and effortlessly conjoined into integrated visual percepts, a process which is crucial to fluent reading ability. We need years of explicit instruction and practice before we start to exhibit any fluency in visual word recognition. Blomert and Froyen (2010) point out that in the last decade, neuroimaging studies have identified a brain region that shows specialization for fast visual word recognition (Cohen et al, 2000) in the occipito-temporal cortex. Since fluency and automaticity are the most salient features of experienced reading, it is indeed plausible that a neural network involved in visual object recognition has specialized for recognizing visual letters and word forms (McCandliss et al., 2003).

This contrasts sharply with the way we learn to master spoken language. Infants and young children start to pick up and develop the many complexities of spoken language without explicit instructions at a time when literacy instruction is still far in the future. Recent electrophysiological evidence shows that it takes several years of reading instruction and practice before the first signs of automatic integration of letters and speech sounds appear in normally developing children. Letter–speech sound associations are cultural interventions and therefore biologically arbitrary in nature.

While MSLE holds much promise to accelerate reading skills, there is little empirical evidence on its efficacy in the real world. This is the first large-scale study on the efficacy of multi-sensory technology for language acquisition, and reading and comprehension.

The RightToRead Intervention

ReadToMe™ provides a multi-sensory experience which enhances the learner's engagement and improves retention. RTM trains itself on the curriculum prescribed textbook getting as close as possible to human reading. Human experts can fine tune various aspects of the software including reading speed, voice with various accents, pronunciation strings to reflect context, intonation, etc. It empowers learners with various tools to help develop vocabulary, enable comprehension and practice pronunciation. The EnglishHelper Master Trainer has identified lesson planning methods to most efficiently conduct ReadToMe classes. The core training on lesson planning is to guide teachers on effectively utilising all ReadToMe tools in the stipulated class time. Further, the lesson plan aims to build confidence in the effectiveness of using software and technology enabled learning in classrooms.

In 2013, EnglishHelper launched RightToRead in India by deploying ReadToMe in government schools. In the first phase, RightToRead was implemented across 100 government schools in 6 states covering 20,000 students. Encouraged by the results and leveraging the capability developed in the first phase, RightToRead expanded to 9 states covering 60,000 students in 2014-15. In 2015, the program further expanded to cover over 1 million students in 5,000 government schools for the academic year 2015-16. This was undertaken with the support of USAID under their India Partnership Program.

The goal of RightToRead is to demonstrate that reading & comprehension technology when integrated with the given school curriculum can make a material difference in literacy. Working on the tenets of "Minimum Change and Sustainability", the following actions are practiced:

- ReadToMe is automatically trained on the class text book and state prescribed syllabus which includes digitizing the textbook in a special format so the ReadToMe engine can read it aloud, and fine tuning pronunciation strings for localized context. No additional or new study material is introduced to the students.

- The software is integrated into the school time-table for the regular English class period or their digital lab period. Note that RTM can be used for any subject that is taught in English.

- The existing teachers are trained to use ReadToMe.

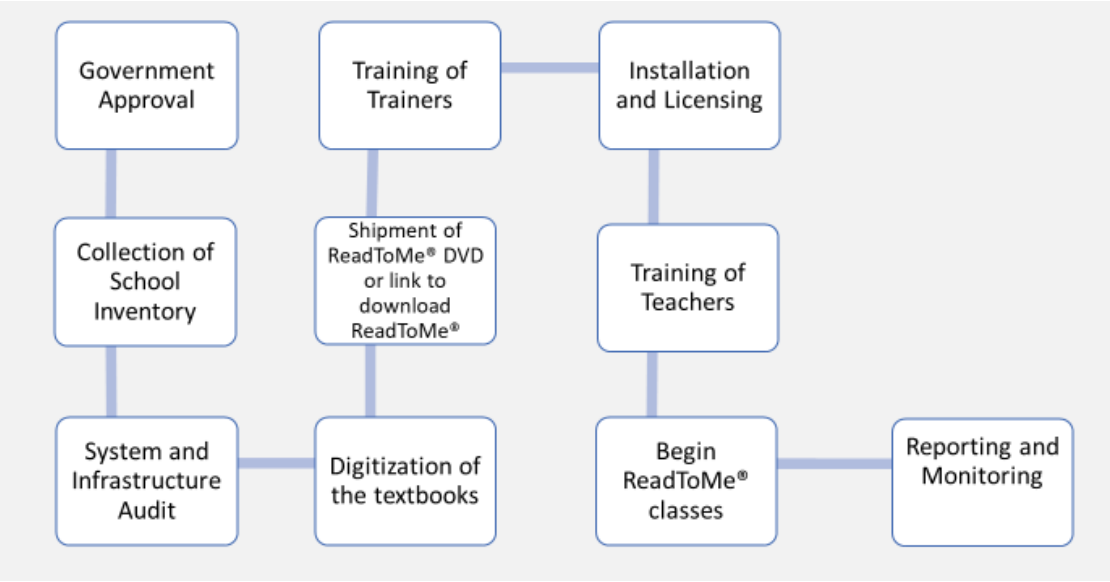
Implementation Model and Process

The RightToRead implementation model relies on leveraging existing infrastructure, personnel, and operating processes. Strong partner support and collaboration is vital for

the implementation and smooth running of RightToRead. Thus, EnglishHelper through its various implementation stages worked closely with its partners to build defined processes.

The first step for the RightToRead implementation is obtaining government approvals to deploy ReadToMe to the schools. This was undertaken with the help of our partner's existing footprint in government schools - IL&FS Education through their ICT program and American India Foundation through their Digital Equalizer program.

On obtaining the necessary government approvals, implementation of RightToRead involved 7 primary steps – acquiring school demographic data, an infrastructure and systems audit of the schools, acquiring school textbooks, training and integration of the school textbooks with RTM, deployment of ReadToMe with partner personnel, installation of RTM, and training of partner personnel and school teachers. Post implementation, the program was monitored to check for effective integration of RTM in school timetables, frequency of its use, and feedback on the product and program. The RTM classes were monitored in four ways - school visits and school calls conducted by the RTM field team, IVR reporting undertaken by the partner field personnel, and expert visits conducted by learning experts and members of the Program Management Office (PMO).



Post – Deployment School Monitoring

Consistent use of ReadToMe is key to achieving the objective of the project. For improved reading and comprehension skills among students the use of ReadToMe in classrooms was suggested to be at least 3 classes per grade per week and a total of 75 classes per grade in the academic year. Thus, EnglishHelper field personnel, learning experts, and members of the PMO constantly monitored the schools to ensure regular and proper use of ReadToMe. Call and visit schedules were developed as per state - specific school calendars. Partner personnel were actively engaged in the monitoring process.

3 Impact Assessment

Rollout Population

This large-scale rollout of RightToRead covered over 1M 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. Fig. 1 shows the students, teachers and schools that were part of the rollout.

Figure I
RightToRead coverage 2016-17

State	No. of Students	No. of Teachers	No. of Schools
Maharashtra	879,026	11,160	3,720
West Bengal	184,536	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	**1,137,172**	**14,667**	**4,889**

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

To assess the impact of RightToRead in an unbiased manner, a randomised control design was adopted. A randomized Treatment group and Control group of schools was selected. Baseline and End line assessments were administered in these schools. 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. This report describes various elements of the randomized 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 Appendix I.

Control-Treatment Design

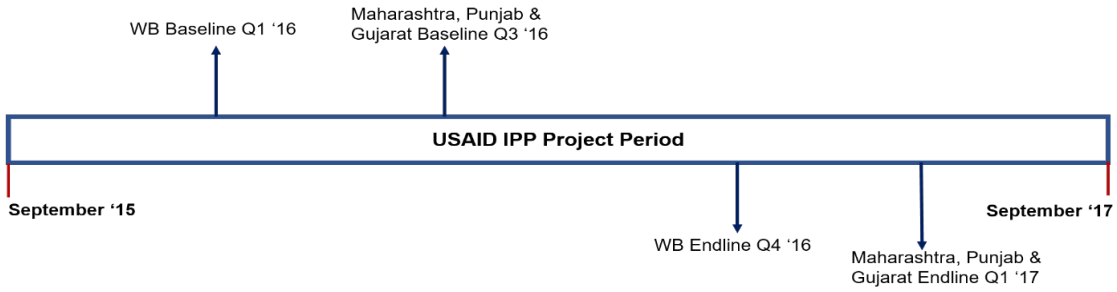
The assessments were conducted across grades 3 to 7. A Control-Treatment design was adopted for comparison of outcomes. Students who underwent technology-enabled reading under the RightToRead program were classified as the Treatment group. Students who were not exposed to technology-enabled English learning 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.

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 II
Assessment schedule (Q = calendar quarter)

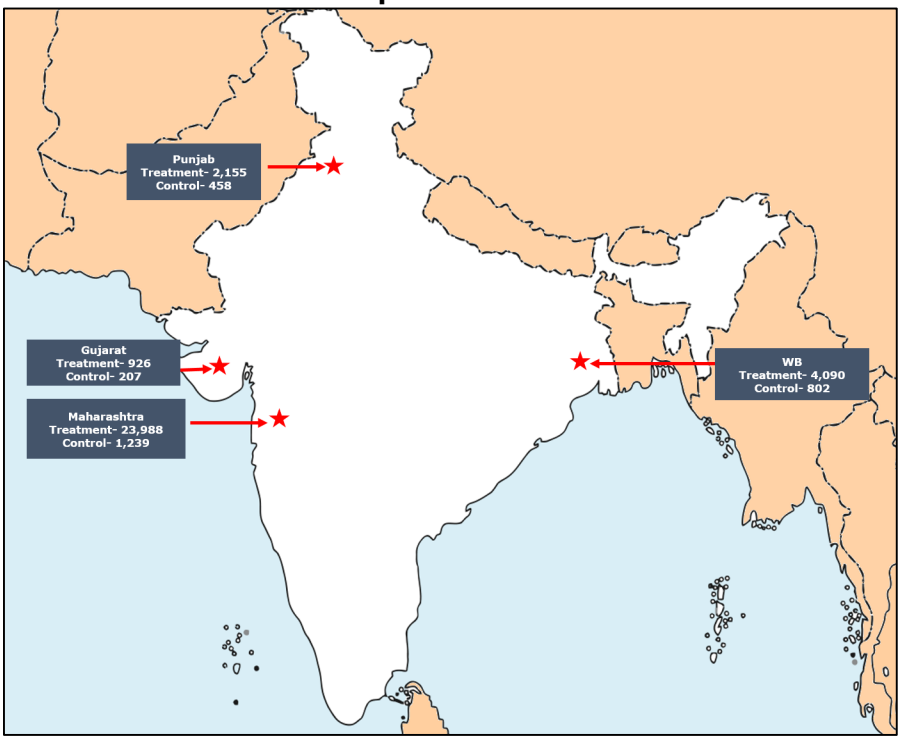
Academic Year	States
June to May	Maharashtra, Punjab & Gujarat
January to December	West Bengal



Assessment Sample

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

Figure III
Sample distribution



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 - 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 880,000 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%).

Figure IV
Maximum State-wise Sampling Error

State	Population Size*	Sample Size	Max 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

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.

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

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).

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.

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 EnglishHelper 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 EnglishHelper personnel who visited at least 10% of schools in each state during the assessment process.

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.

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. All the data transferred by STAMP to GMI was checked for completeness, accuracy, anomalies and a sample was also subjected to back-checks.

STAMP delivered a summary sheet accompanying every parcel of data transferred. This summary sheet was also provided to EnglishHelper. This was checked by EnglishHelper for the counts of schools and students against field reports. GMI checked for the counts of schools and students on the data set received. Computations on the data undertaken by STAMP were validated by GMI to ensure the accuracy of the transformed data.

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.

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.

4 Results

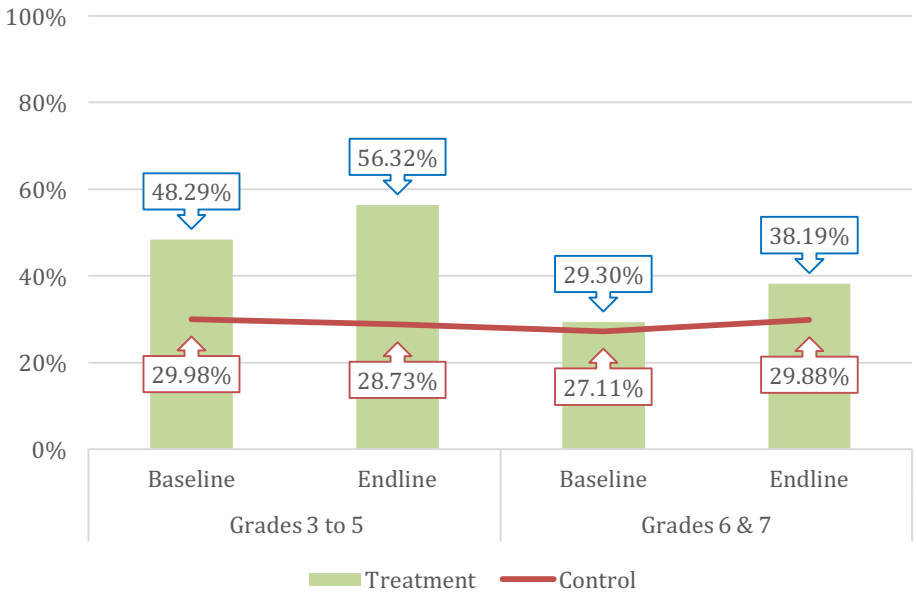
The assessments demonstrate that ReadToMe has a positive impact on English reading and comprehension among children undergoing the RightToRead program. Across the total of over 33,000 students assessed spanning five grades in the different states, students undergoing ReadToMe classes (Treatment) were 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 discuss 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.

Overall

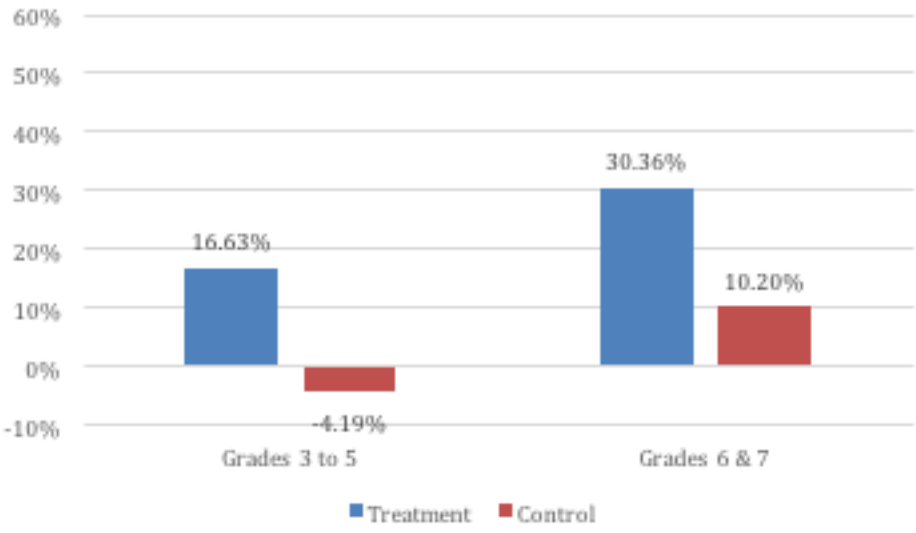
The primary grades of 3 to 5 witnessed a 17% improvement in English scores in one academic year for the Treatment group as compared to a 4% decline among the Control group. Improvement in grades 6 and 7 was over 30% in the Treatment group as compared to 10% in the Control group. Figure IV illustrates 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 IV
Change in scores from Baseline to End line – Overall



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

Figure V
Improvement of Treatment and Control – Overall



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.

Figure VI 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.

Figure VI
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 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.

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 for the overall analysis. Please refer to the section on overall analysis for explanation and interpretation of the figures presented in the state-wise analyses. In addition, the state-wise analyses also present the sample sizes per grade.

Sample Size

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

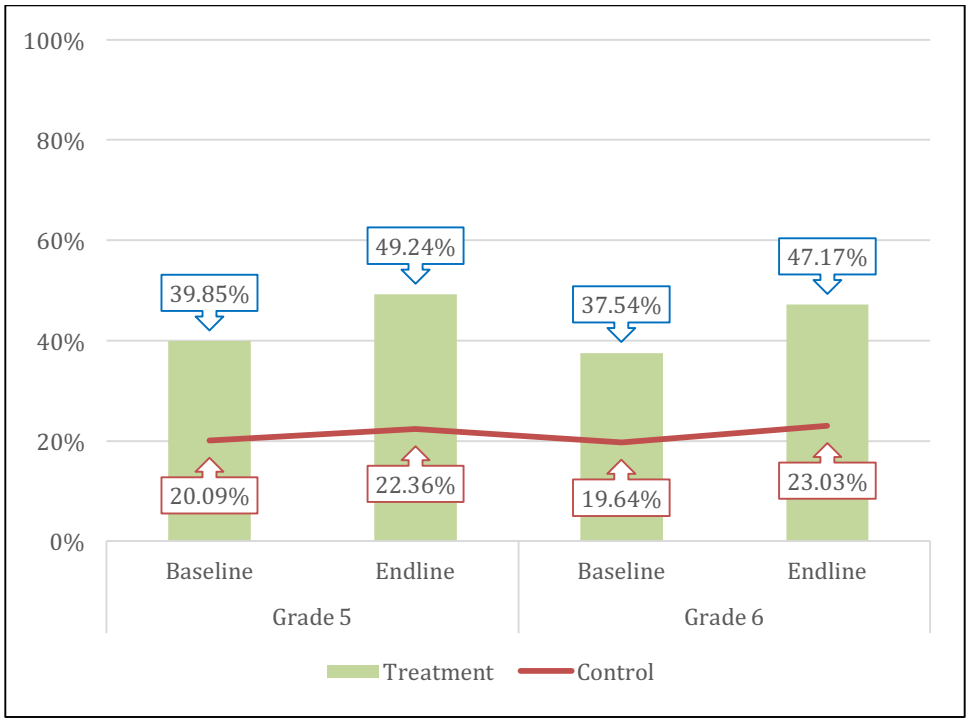
Figure VII
Sample size – West Bengal

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

Assessment Outcome

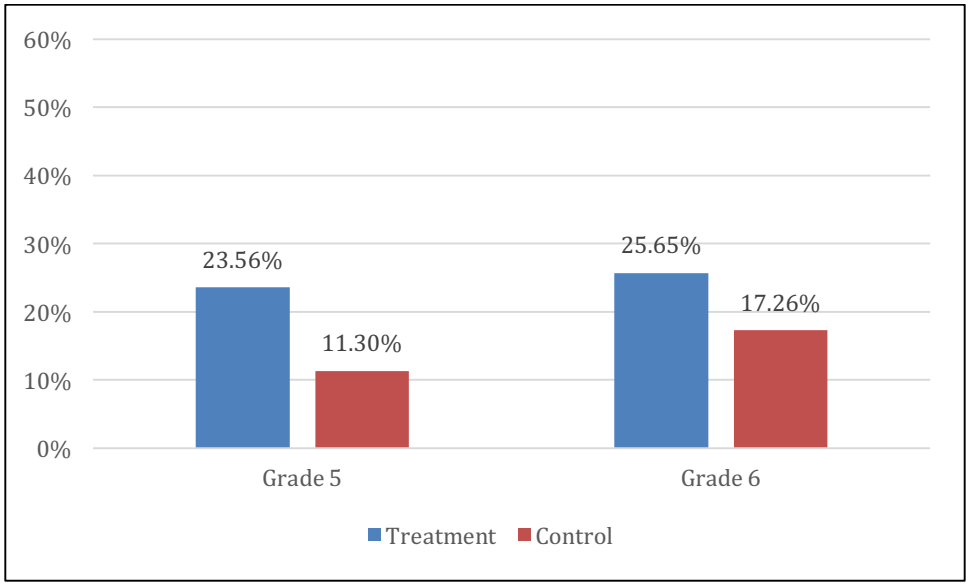
Treatment and Control groups in both grades exhibit an increase in scores from Baseline to End line.

Figure VIII
Change in scores from Baseline to End line – 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 IX
Improvement of Treatment and Control – West Bengal



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).

Figure X
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%**

Maharashtra

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

Sample Size

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

Figure XI
Sample size – Maharashtra

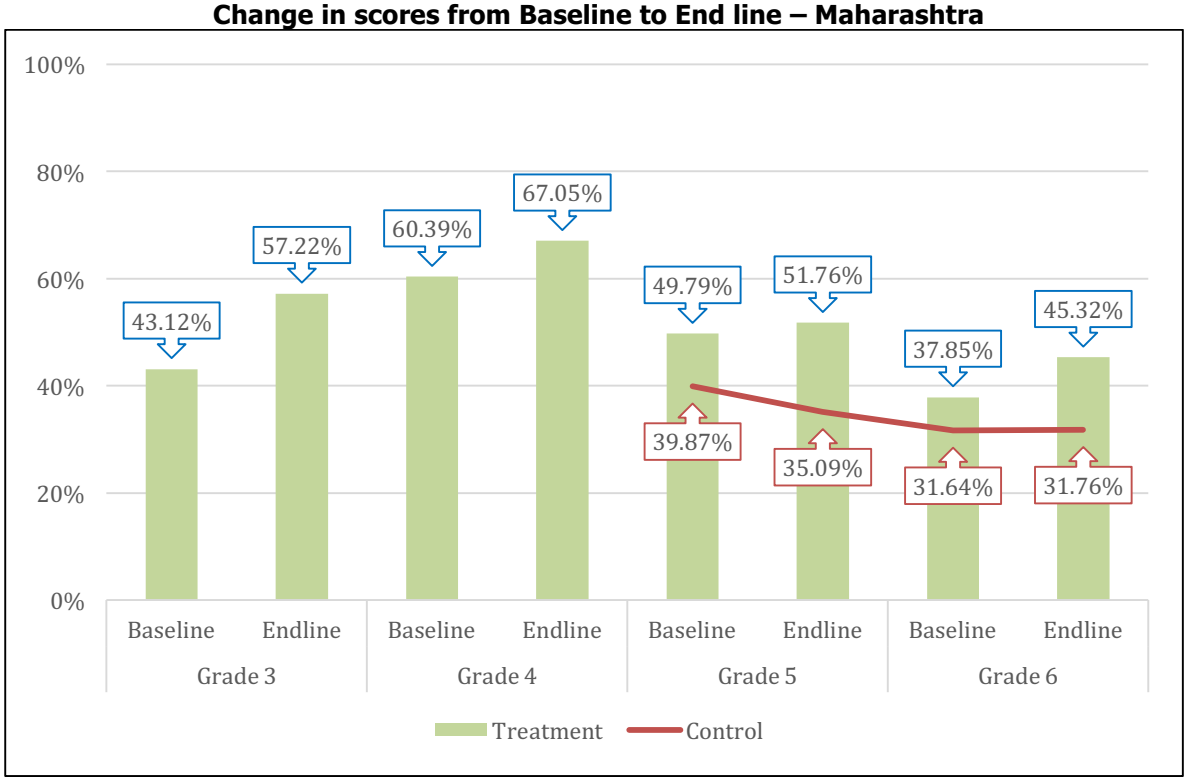
	Grade 3	Grade 4	Grade 5		Grade 6	
	Treatment	**Treatment**	**Treatment**	**Control**	**Treatment**	**Control**

Number of students	1,005	1,049	10,147	629	11,787	610

Assessment Outcome

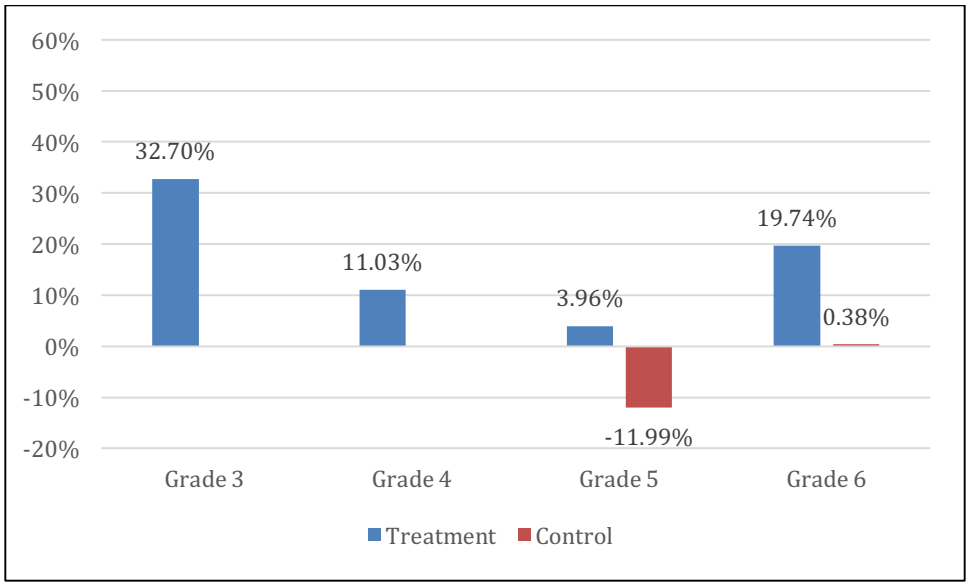
The Treatment group of all the grades shows an increase from Baseline to End line. Grade 5 Control exhibits a decline from Baseline to End line while Grade 6 Control remains unchanged.

Figure XII
Change in scores from Baseline to End line – 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 XIII
Improvement of Treatment and Control – Maharashtra


60%
50%
40%
30%
20%
10%
0%
-10%
-20%
32.70%
11.03%
3.96%
-11.99%
19.74%
0.38%
Grade 3
Grade 4
Grade 5
Grade 6
Treatment
Control

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.

Figure XIV
Improvement in Quartiles – Maharashtra

	Grade 3			**Grade 4**		
	Treatment			**Treatment**		
	Baseline	Endline	**Change**	Baseline	Endline	**Change**
Maximum	89.66%	96.67%	**7.01%**	100.00%	100.00%	**0.00%**
Percentile 75	58.62%	76.67%	**18.05%**	79.31%	83.33%	**4.02%**
Median	41.38%	60.00%	**18.62%**	65.52%	70.00%	**4.48%**
Percentile 25	24.14%	36.67%	**12.53%**	41.38%	53.33%	**11.95%**
Minimum	0.00%	0.00%	**0.00%**	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%**

Punjab

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

Sample Size

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

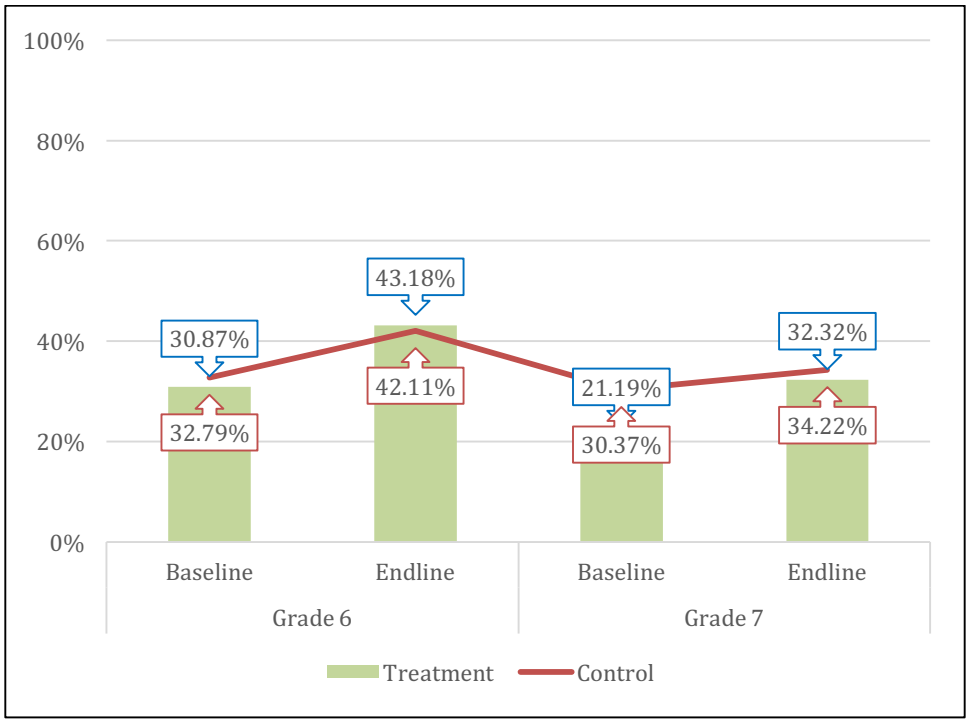
Figure XV
Sample size – Punjab

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

Assessment Outcome

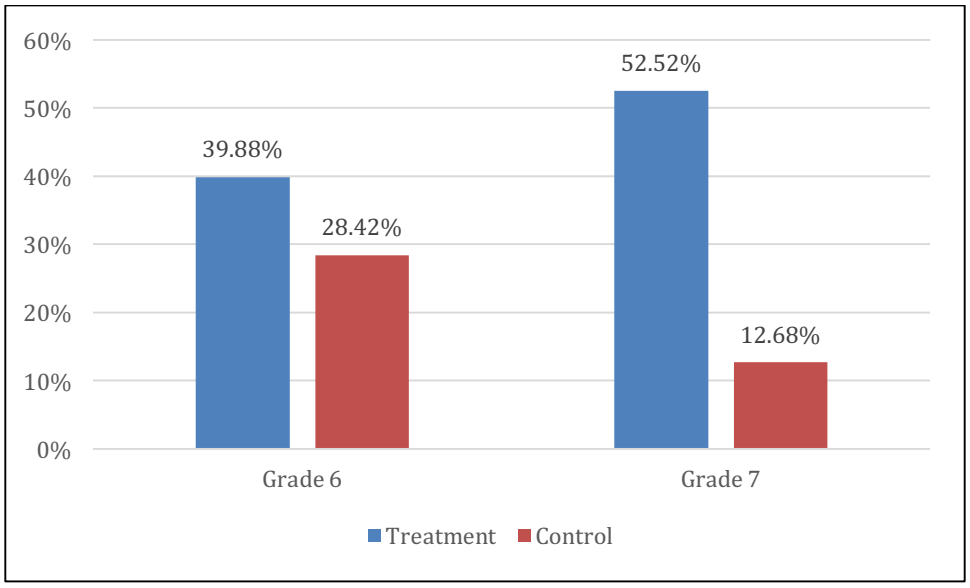
In the Baseline, the Treatment groups of both grades 6 and 7 had lower scores than the corresponding Control groups. However, by the End line, the Treatment group of Grade 6 has scored higher than the Control group and the Treatment group of Grade 7 is lower than the Control group by only 2%.

Figure XVI
Change in scores from Baseline to End line – 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 XVII
Improvement of Treatment and Control – Punjab



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).

Figure XVIII
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%**

Gujarat

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

Sample Size

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

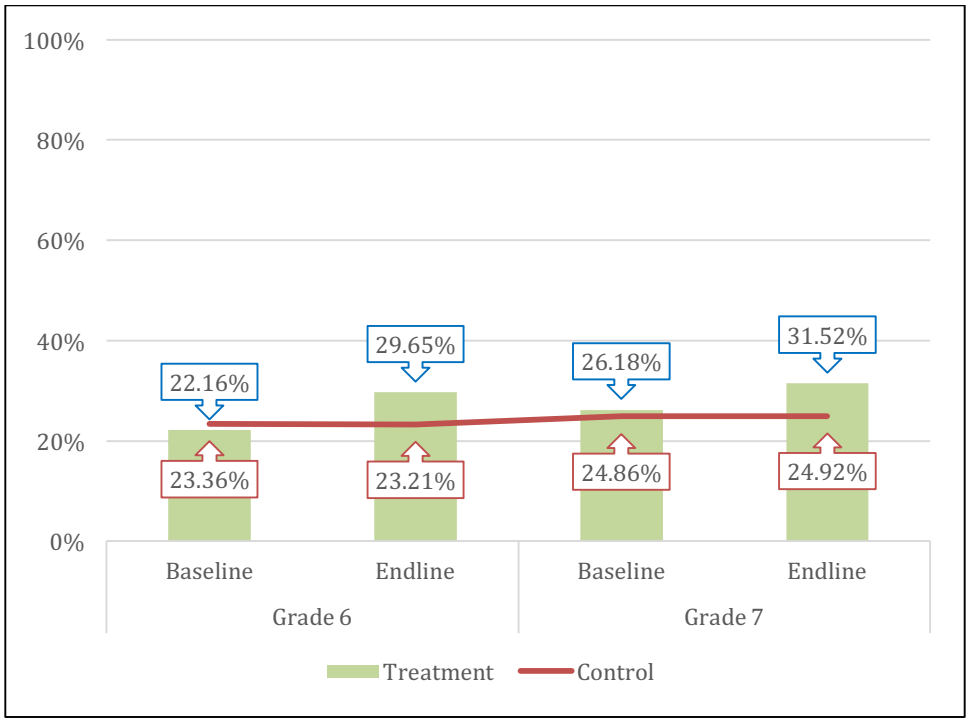
Figure XIX
Sample size – Gujarat

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

Assessment Outcome

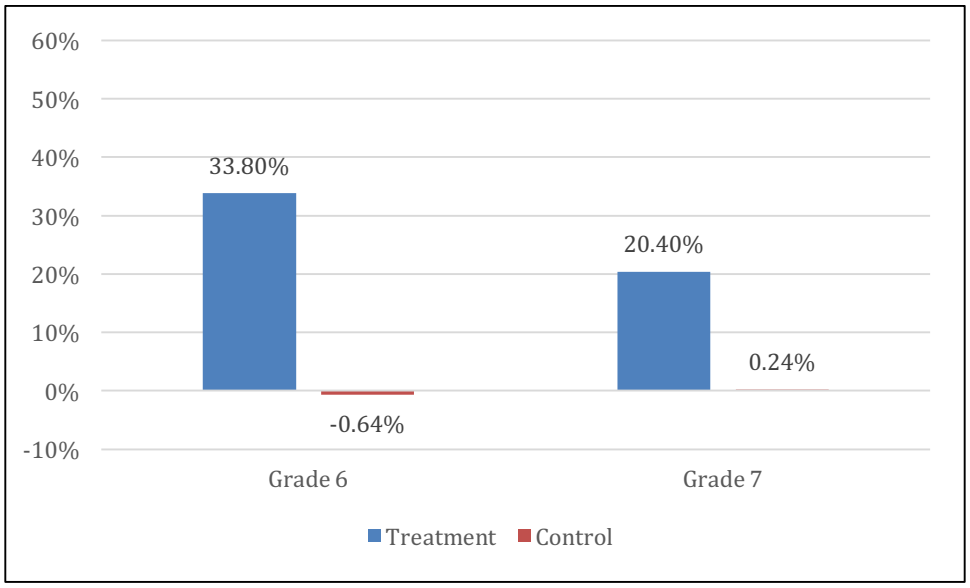
In the baseline the Treatment group of Grade 6 scored lower than the Control group; by the End line the Treatment group scored higher. The Control group has remained largely unchanged in both grades while Treatment group shows significant positive change in scores.

Figure XX
Change in scores from Baseline to End line – 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 XXI
Improvement of Treatment and Control – Gujarat



Change by Quartile

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).

Figure XXII
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%**

CEFR Normalization

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.

In order to normalize the outcomes from each of the assessments (across grades and states) EnglishHelper 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 was adopted to measure the learning outcomes aligned with the CEFR.

1 Adapting and establishing the mapping framework

2 Assigning a CEFR level to every question
3 Assigning a CEFR level to every student
4 Comparing the CEFR distribution of students in the Baseline and in the End line

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:

Figure XXIII
Adapted CEFR for 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. Can understand short, simple texts on familiar matters of a concrete type which consist of high frequency everyday or job-related language

A2	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

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 Figure XXIII.

Figure XXIV
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%

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.

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, and indicate positive impact of the RightToRead program using the ReadToMe platform, integrated with the curriculum, on the English learning of students.

Figure XXV and Figure XXVI
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 XXV
Comparison of the CEFR achievement between Baseline and End line – Grades 3 to 5

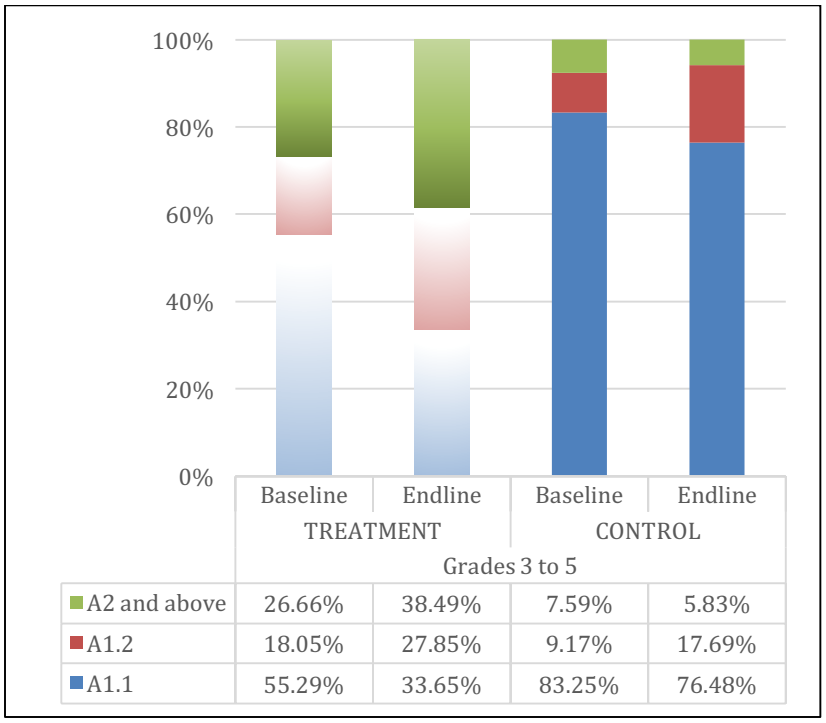
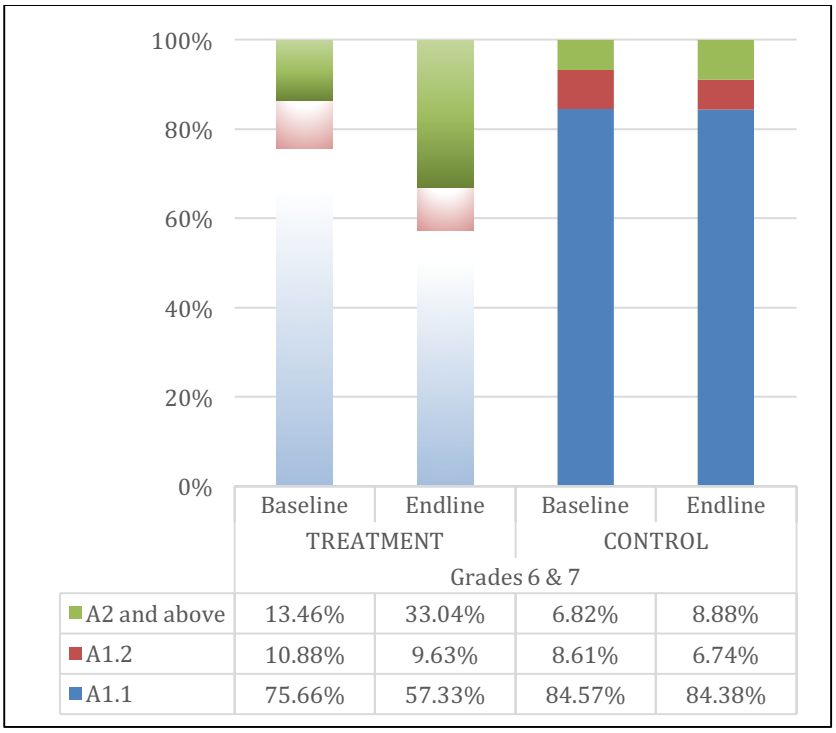


Figure XXVI
Comparison of the CEFR Achievement between Baseline and End line – Grades 6 & 7



5 Conclusion

We have reported the results of a large scale intervention to improve English literacy in K-12 schools in India. The intervention had reached 1 million students at the time of the study and since then has expanded to 2.5 million students besides several countries. The intervention comprised a software program involving the use of AI and engaging multiple senses to effectively optimize the cognitive load on students.

The impact assessment was conducted by an independent agency using a randomized control-treatment design. 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. The assessments included a Control group to identify comparable improvement in reading and comprehension of students not exposed to technology-enabled reading of their text books.

These independent assessments encompassing four key parameters in English provide a picture of the impact of the intervention on the 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 demonstrated between 8% and 40% higher improvement in scores as compared to students who were not 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.

To enabled comparison across grades and regions, we further normalized the assessment instrument and results according to the CEFR. We created a mapping between the raw assessment results and CEFR levels. The Treatment group shows reading improvement across all levels. This analysis enables the comparison of the CEFR level of students with the required proficiency prescribed by the curriculum.

We have continued to assess the ongoing impact of the intervention. These results which were done in house demonstrate a much higher level of impact in the assessed population. Students with the intervention show a 50-60% improvement in their literacy levels.

EnglishHelper 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.

Appendix I
Assessment Instrument

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)

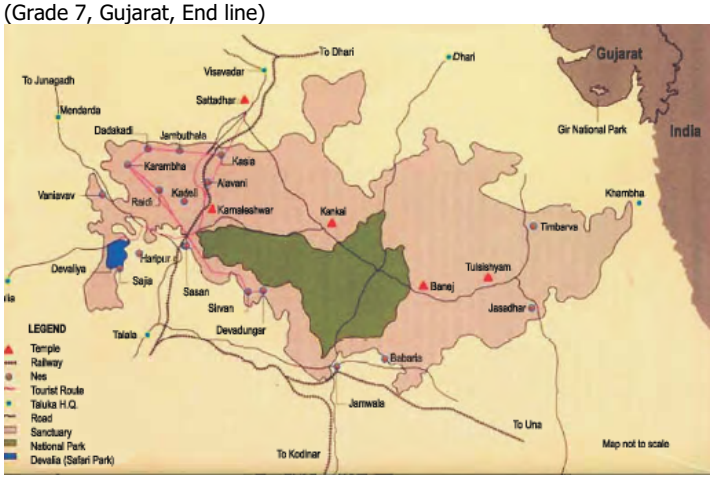
KNIFE

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.*

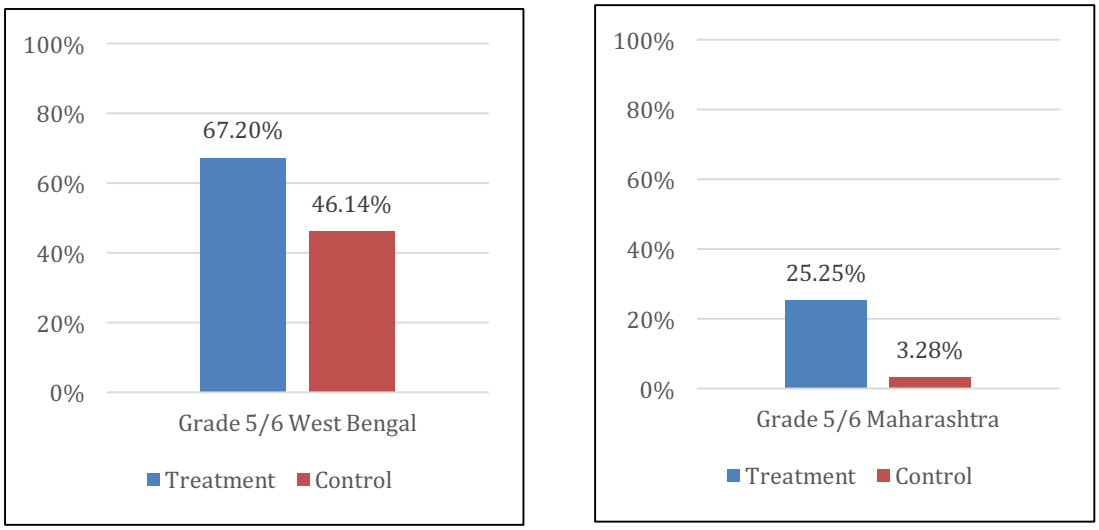
Appendix II
Ongoing Assessment Results

Two-year Assessments

As an extension of the Baseline-End line research conducted in 2016-17, Grade 6 students from West Bengal and Maharashtra were assessed again towards the end of their respective academic years in 2017-18. (The students were part of the Grade 5 assessments in 2016-17.) The objective of the 2-year End line was to ascertain the presence of sustainable improvement on outcomes and hence, impact of the RightToRead program.

The improvement of mean scores obtained by the students in 2017-18 over the Baseline scores, is presented below.

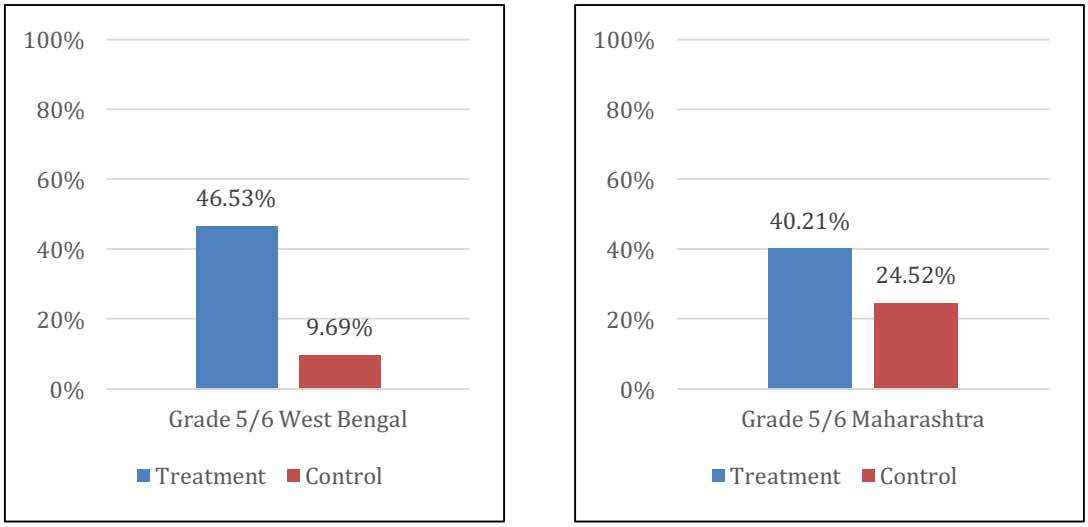
Figure XXVII
Two-year improvement of Treatment and Control – West Bengal and Maharashtra – Grades 5 / 6



In both states, Treatment group has achieved between 21 and 22% more improvement in reading and comprehension scores, over the Control group.

Remarkable gains were also observed in the CEFR proficiency of students. The figure below presents the change in the percentage of students at the A1.2 and higher CEFR levels between the 2-year End line and the Baseline for each cohort.

Figure XXVIII
Two-year change in Percentage of students in higher CEFR levels (A1.2 and higher): Treatment and Control – West Bengal and Maharashtra – Grades 5 / 6



More than 46% of students achieved higher proficiency by the 2-year End line in the West Bengal Treatment group compared to fewer than 10% in the Control group. The program in West Bengal enabled more than 36% more students to achieved higher CEFR proficiency, in the assessed group. Similarly, in Maharashtra, the program enabled more than 15% more students to achieve higher CEFR proficiencies, in the assessed group.

Relationship between ReadToMe sessions and Improvement

In the academic year 2017-18, the Government of Telangana State, the Government of Uttar Pradesh and the South Delhi Municipal Corporation (SDMC) piloted ReadToMe in schools in their respective regions. EnglishHelper conducted randomised sample assessments in these schools, designed as Baseline - End line studies. This would allow measurement of learning outcomes achieved in one academic year among schools using ReadToMe.

EnglishHelper recommends at least 75 class sessions of ReadToMe in an academic year. In all three regions, due to various constraints, fewer than the recommended number of sessions were conducted.

The figure below shows the mean number (and percent) of ReadToMe sessions that were conducted in each region in the academic year.

Figure XXIX
Percent of ReadToMe sessions conducted in the academic year

	Mean Number of ReadToMe® Sessions	% of Recommended ReadToMe® Sessions
Uttar Pradesh	15	20.0%
Telangana	32	42.7%
Delhi	56	74.7%

In Uttar Pradesh and Telangana, assessments were conducted among students from Grades 6 to 8. In Delhi, assessments were conducted among students from Grades 3 to 5.

Presented below is the total (End line) count of students in each region.

Figure XXX
Sample Size 2017-18

	Number of Students
Uttar Pradesh	249
Telangana	518
Delhi	1521

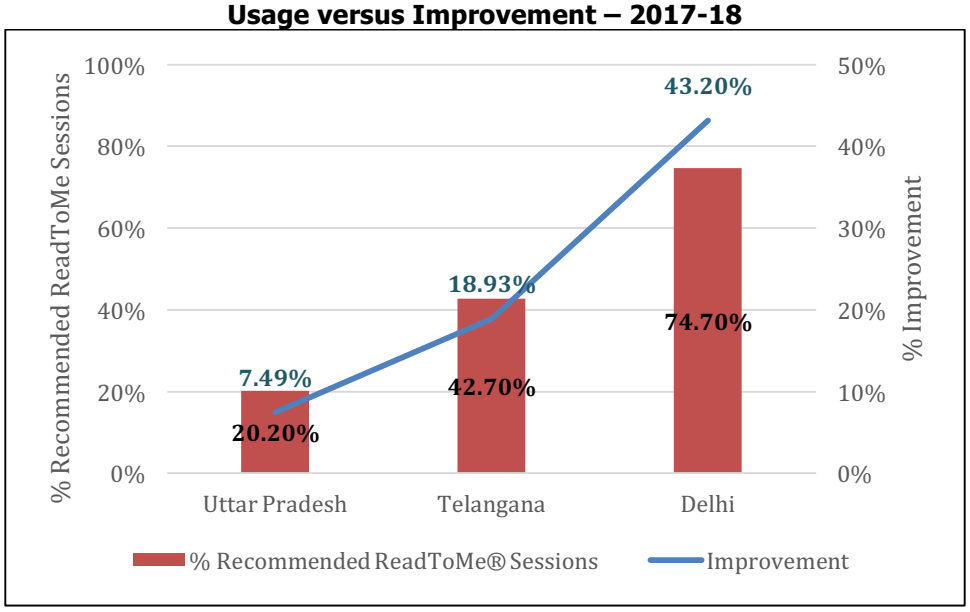
It is important to note that despite the different grades and regions, on instruments with high reliabilities, the mean Baseline scores were nearly at par. Hence, comparisons of improvement remain meaningful. The figure below presents the mean Baseline scores in each region.

Figure XXXI
Mean Baseline scores

	Mean **Baseline Score**
Uttar Pradesh	33.77%
Telangana	38.24%
Delhi	34.64%

The figure below plots the mean improvement in every region against the mean percentage of recommended ReadToMe sessions that were conducted in the region in the academic year.

Figure XXXII
Usage versus Improvement – 2017-18



We observe that with higher ReadToMe usage (bars), significantly higher gains in improvement (line) can be achieved in the Learning Outcome. Therefore, we can conclusively state that there exists a positive relationship between ReadToMe usage and improvement in learning outcomes, across grades and geographies.

References

Andrabi, T., Das, J., Khwaja, A. I., Vishwanath, T., and T. Zajonc, "Learning and Educational Achievements in Punjab schools (LEAPS): Insights to Inform the Education Policy Debate," (World Bank, Washington, DC, 2007).

ASER Centre, "Annual Status of Education Report (Rural) 2005," (Mumbai: Pratham Resource Center, 2006).

ASER Centre, "Annual Status of Education Report (Rural) 2008," (Mumbai: Pratham Resource Center, 2009).

ASER Centre, "Annual Status of Education Report (Rural) 2011," (New Delhi: ASER Centre, 2011).

ASER Centre, ASER District Performance Tables 2011, Retrieved on 24 July 2016 from: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER_2011/DPT_2011/2011districtpage.pdf

ASER Centre. (2012). ASER District Performance Tables 2012. Retrieved on 22 August 2016 from: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER_2012/DPT_2012/aser2012districttablesapril42013-allstates.pdf

ASER Centre. (2014). Annual Status of Education Report 2014 – National Findings. New Delhi: ASER Centre.

ASER Centre. (2016). Annual Status of Education Report 2016 – National Findings. New Delhi: ASER Centre.

Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., and S. Khemani. (2010). Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India. American Economic Journal: Economic Policy, 2(1):1–30.

Banerjee, A.V., Cole, S., Duflo, E., and L. Linden. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. The Quarterly Journal of Economics, 122(3):1235– 1264.

Banerjee, A.V., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M. and Walton, M. (2016). Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India, MIT Department of Economics Working Paper, Cambridge, Mass, ZDB-ID 21886593. - Vol. 16, 08

Blomert, L., & Froyen, D. Multi-sensory learning and learning to read. *International Journal of Psychophysiology*, 77, 195-204.

Byrne, B., R, F.-B., 1989. Phonemic awareness and letter knowledge in the child's acquisitions of the alphabetic principle. Journal of Educational Psychology, 80, 313-321.

Cavanaugh, C.S., (2001). The effectiveness of interactive distance education technologies in k-12 learning: A meta-analysis. *International Journal of Educational Telecommunications*, 7(1), 73-88.

Chapelle, C.A. (1997). CALL in the year 2000: Still in search of research paradigms. *Language Learning & Technology,* 1(1), 19-43.

Chapelle, C.A. (2001). Computer applications in second language acquisition: Foundations for teaching, testing and research. Cambridge: Cambridge University Press.

Ghazanfar, A.A., and Schroeder, C.E., 2006. Is the neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10, 278-285.

Hedge, L.V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando: Academic Press.

Iandoli, J.J., (1990). CALL and the profession: The current state. *French Review*, 64(2), 261-272.

Liu, M., Moore, Z., Graham, L. & Lee, S., (2001). A Look at the Research on Computer-Based Technology Use in Second Language Learning: Review of Literature from 1990-2000. *Journal of Research on Technology in Education*.

Lopez, O.S., (2010). The Digital Learning Classroom: Improving English Language Learners' academic success in mathematics and reading using interactive whiteboard technology. Computers and Education, 54, 901-915.

Lou, Y., Abrami, P.C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449-521.

McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The Visual Word Form Area: expertise for reading in the fusiform gyrus. Trends in Cognitive Sciences, 7(7), 293-299.

McCandliss, B.D., Posner, M.I., and Givon, T., 1997. Brain Plasticity in Learning Visual Words. Cognitive Psychology, 33, 88-110.

McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P (2007). "The Hawthorne Effect: a randomized, controlled trial". *BMC Med Res Methodol* ***7***: 30.

Mousavi, S.Y., Low, R., and Sweller, J., (1995). Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. *Journal of Educational Psychology*, 87(2), 317-334.

Pennington, M.C. (1996). The Power of the computer in language education. In M.C. Pennington (Ed.), The Power of CALL (pp 1-14). Houston, TX: Athelstan.

Pusack, J.P., & Otto, S.K., (1990). Applying Instructional Technologies. *Foreign Language Annals*. 23(5), 409-417.

Salaberry, M.R. (2001). The use of technology for second language learning and teaching: A retrospective. *The Modern Language Journal*, 85 (1), 39-56.

Wilcoxon, F., (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), 80-83.