Computers and Education: Artificial Intelligence 2 (2021) 100019

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.elsevier.com/locate/caeai

Improving reading and comprehension in K-12: Evidence from a large-scale AI technology intervention in India

Venkat Srinivasan [a,*], Hemavathi Murthy [b]

[a] EnglishHelper Inc. and Innospark Ventures, Boston, USA
[b] EnglishHelper India Pvt Ltd, Gurgaon, India

ARTICLE INFO

Keywords:
Elementary education
Distance education and online learning
Improving classroom teaching
Cooperative/collaborative learning
Teaching/learning strategies

ABSTRACT

This paper presents evidence on the impact of technology-aided instruction on literacy using an AI-based multi-sensory technology platform across a large cross-section of government schools in India. The study focused on reading and comprehension in the English language. The intervention enhances the instructional effectiveness of the teachers and the learning ability of the children within the existing instructional environment without any new instructional design or pedagogy or content. Besides, the intervention is implemented by existing teachers and not outside volunteers.

A total of 1 million children and 15,000 teachers across 5000 government schools in the states of Maharashtra, West Bengal, Punjab, Tamil Nadu, Telangana, Gujarat and Karnataka used the technology for the 2016-17 academic year. Using a randomized control-treatment assessment, the study finds a 20–40% overall gain in learning outcomes in the treatment sample. Gains within individual states and grades vary. Learning outcomes rose across the entire range of proficiency levels in a grade. Ongoing self-administered 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 teacher shortage and inadequate teacher training.

The paper also reports briefly on the growing adoption of the program in several countries outside India. The results hold significant promise for disrupting the low and stagnating literacy levels across government schools in India and other similar environments.

1. Introduction

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. As reviewed subsequently, a 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. More specifically, in developing economies like India, English proficiency takes on a whole new meaning and social significance. It is tied to economic and social well-being. Several studies document these conclusions including the differential income generating capacity of persons with superior English language skills (Azam et al., 2013; Panda, 2017).

The potential role of technology however goes far beyond language learning and extends into foundational literacy especially in K-12 environments. The often used cliché, “learning to read and reading to learn” captures the twin facets of technology impact at foundational levels. Thus, foundational literacy in English has a beneficial impact in the learning of other subjects.

In this paper, we report on the use of an AI based framework for improving literacy. We define AI to mean ‘computational intelligence’ embedded in machines to perform tasks that humans normally do. Such computational intelligence can be either symbolic AI gained from experts or research findings in learning or neuroscience or intelligence gained from empirical observations (data) or both. In our experience, it is rare that intelligence acquired purely from data will ever be ready to be deployed for actual real world use especially in high value problem settings. In the context of literacy, our definition encompasses technology induced enhancement of the learner’s ability to improve their reading and comprehension. Incorporating findings from neuroscience is part of this definition.

Recent studies from neuroscience (learning science) suggest that multi-sensory structured learning education (MSLE) has significant

* Corresponding author.
E-mail address: venkat.srinivasan@innosparkventures.com (V. Srinivasan).

https://doi.org/10.1016/j.caeai.2021.100019
Received 4 December 2020; Received in revised form 5 April 2021; Accepted 5 April 2021

potential to improve the systematic acquisition of reading skills. As reviewed subsequently, 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.

This paper presents the results of a large scale application of multi-sensory technology for reading and comprehension in English across 5000 government schools across 8 States in India covering 1 million children and 15,000 teachers. The impact on learning outcomes was assessed by an independent agency based on a randomized control-treatment (RCT) design. The results indicate that MSLE is highly effective in improving reading and comprehension in English even in highly constrained infrastructural settings. We conducted a follow up assessment on the RCT and found significantly higher levels of impact. The intervention has the potential to radically alter educational outcomes and can be scaled to cover entire populations.

This paper seeks to shed light on the following research questions:

1 Can technology incorporating learning science have beneficial impact on learning outcomes? And in the context of this paper on language learning?
2 Can we demonstrate rigorously the impact of technology on learning outcomes on a large scale?
3 Can we establish the impact of technology without focus on pedagogical content or method?
4 Can we demonstrate a framework or approach for successful introduction of AI technology and overcome change management challenges?
5 Can we address lack of availability of quality teachers and inadequate computing infrastructure with creative educational technology?

The study is pioneering in many ways. First, it is the first large scale study of the impact of AI technology for reading and comprehension in English, in a K-12 setting. Second, it supports significant research findings that multisensory methods can be effective in the acquisition of reading and comprehension skills. A secondary finding is that the technology is effective in language learning. 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.

Reading and comprehension are foundational skills in education. If a learner is a poor reader and has difficulty understanding written material, it doesn't matter what subject it is. While this study is focused on English as a subject, the technology is equally applicable to all subjects.

2. Prior research

There is widespread recognition that AI can and will play a significant role in education at large. We are not aware of any study which specifically uses AI technology for improving reading and comprehension skills. However, there are a plethora of studies on the use of technology for language learning generally and for learning English as a second language. There are also studies in related fields like intelligent tutoring systems and learning levels in poor k-12 schools with similar infrastructural challenges which are useful to review to provide fuller background and motivation.

Liu et al. (2002) 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.

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; bib_Lou_et_al_2001Lou et al., 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, 2001bib_Chapelle_1997bib_Chapelle_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. Based on a meta-analysis of a number of empirical studies, Zhao (2003) reported a significantly positive impact of technology applications on language learning.

In a more recent comprehensive effort, Golonka et al. (2014) review over 350 studies on (including classroom-based technologies, individual study tools, network-based social computing, and mobile and portable devices) and find that, in spite of an abundance of published research, **evidence of efficacy of technology for foreign language learning is limited**. However, they report that indirect support for technology's impact in foreign language learning can be deduced from studies on computer-assisted pronunciation training, in particular, automatic speech recognition (ASR). They also found that chat significantly increased both the amount of learners' language production and its complexity.

John Seely Brown (Iiyoshi & Vijay Kumar, 2010) recognizes that all learning does not happen inside the classroom and argues forcefully for blurring the distinction between formal learning (classrooms) and informal learning out of class. In the context of language learning, Lai (2011, 2013, 2017) has examined this in great detail. Lai (2017) provides critical insights into theoretical approaches for supporting technology assisted out of class learning methods and in-depth guidelines for fostering such learning. These have become particularly relevant in the context of the Covid 19 pandemic. In a related study, Yang (2013) reviews studies focusing on mobile assisted language learning - short messaging service (SMS), microblogging, ambient intelligence and augmented reality, GPS and tablet computing and concludes that mobile technologies in general have had a positive impact on language learning. Godwin-Jones (2019) examine the online informal language learning trend with learners active in the "digital wilds" like social media and caution against narrow studies that consider narrow contexts to identify factors influencing language learning.

More recently, Abdulrahaman et al. (2020) systematically review the literature on the use of multimedia tools in teaching and learning with the objective to understand which tools and what circumstances lead to the most effective outcomes. Not surprisingly, they find that the success of multimedia tools used depended on how well the tools were designed reflecting an understanding of the objective the tools sought to achieve and the appropriateness of the technologies for the given target student population. Ahmadi (2018) more specifically reviewed the literature on English language learning. The review concludes that effective use of new technologies improves learners' language skills.

As can be observed from the above review, the impact of technology on language learning has been the subject of considerable scrutiny by researchers. However, most of the above studies suffer from several significant limitations. First, the sample sizes of prior 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. Fourth, as reported by Abdhulrahaman and Ahmadi (2020), most of the tools reviewed were focused on the pedagogical content related to different subjects. None of the studies were about AI technologies based on learning science principles. None were content agnostic and none were about foundational literacy.

Chan and Zary (2019) point out in the case of medical education, how with wide application of AI technologies for teaching and learning, instructors are able to get rid of repetitive and tedious tasks and to reply to students timely, thus advancing the adaptive and personalized teaching process. The intervention reported in this study also offers a very similar benefit to teachers except the context here is that the teachers are freed up to focus on the students and not be in front of the class reading aloud repetitively.

In a related context, 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.

Hwang (2014) proposes a set of criteria for smart learning environments. The intervention in this study is consistent with several of the criteria proposed, the most important perhaps being that the actual intervention implemented in the first phase of this study completely reflects the learner's online and real world status. While the criteria in Hwang (2014) do not explicitly include the integration of technologies that incorporate findings from neuroscience on learning, it does refer to "computer mind tools" which is where this study belongs. Very few studies have measured the impact of such tools and at scale.

Hwang et al. (2020) propose a framework for AI applications in Education in different teaching and learning settings – intelligent tutor, intelligent tutee, intelligent learning tool/partner, and policy making advisor. To our knowledge, there are no AI applications which attempt to comprehensively improve reading and comprehension skills of k-12 learners. There are certainly several applications in math which attempt to provide personalized guidance based on pre-specified normative pathways. The intervention reported in this paper focuses on improving reading and comprehension in English. It fits in both the intelligent tutor and intelligent learning tool categories. It can also serve as a policy making advisor in the future.

3. The RightToRead intervention

The RightToRead intervention was launched in 2013 based on a multisensory technology platform for reading and comprehension called ReadToMe™ (RTM).

3.1. The AI framework

ReadToMe™ (RTM) provides a multi-sensory experience which has been proven to improve retention and knowledge formation in the brain. The RightToRead intervention is essentially an intelligent tutor and can also be thought of as an intelligent learning companion in the taxonomy proposed by Hwang et al. (2020). Fig. 1 shows the overall RTM AI based technology. RTM trains itself on the curriculum prescribed textbook getting as close as possible to human reading. The text book or other curriculum content is converted to a proprietary multi-sensory structure by the RightToRead AI Engine. In this conversion, it adds knowledge such as localization of pronunciation strings, attach images different words, for the sense in which they are being used, to facilitate understanding and a knowledge graph that shows the word's relationship with other words (concepts). Based on learning objectives, learning science is integrated in the form of interleaving and repetition of the major concepts.

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 to help develop

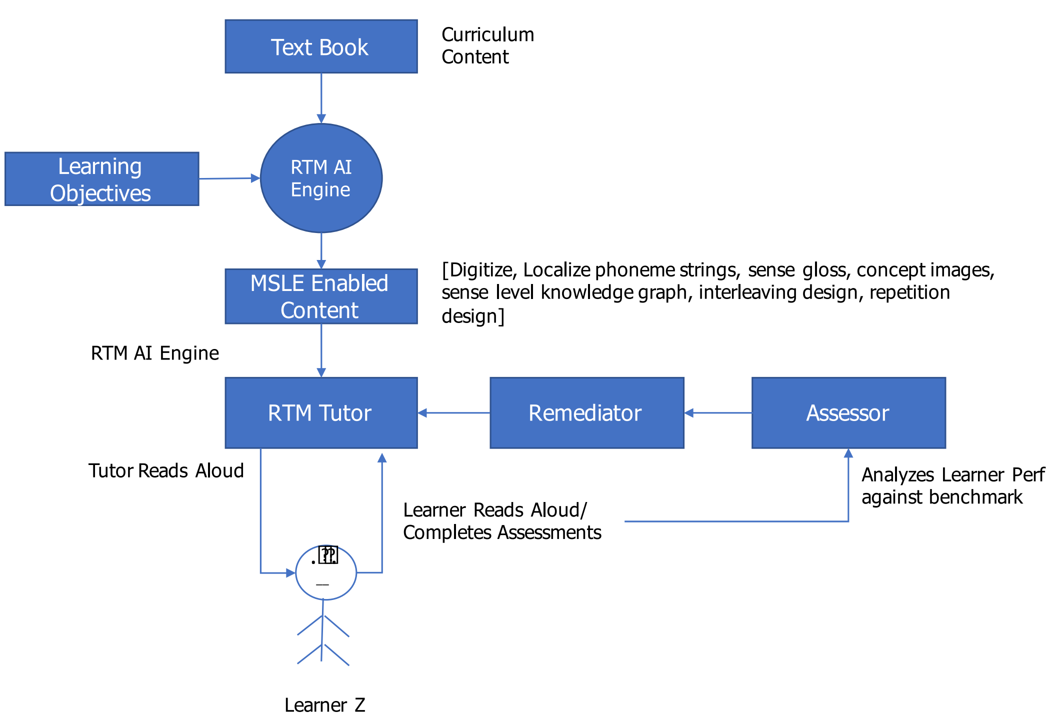


Fig. 1. The RTM AI framework.

vocabulary, enable comprehension and practice pronunciation. At the end of this step, RightToRead product can read the text aloud retaining full visual view of what is being read as the automated reading proceeds word to word. As the tutor reads the text, it also presents pronunciation strings, concept images for understanding and a knowledge graph for the word.

The 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 utilizing all ReadToMe tools in the stipulated class time. Further, the lesson plan aims to build confidence in the effectiveness of the technology in classrooms.

The framework allows the learner to record their reading for analyzing their proficiency. This can be supplemented with other assessments to gauge comprehension. The automated Assessor examines the learner's recording by comparing it to a normative benchmark and provides insights to the learner/teacher at a word, sentence and passage level. The benchmark can be based on any of the standard voices available commercially. The framework envisions an automated (assisted/self-learning) Remediator to feed the tutor appropriate adjustments to improve a specific learner's reading and comprehension proficiency based on the analysis.

This study reports results from the first phase of this implementation. The phased implementation reflects our strategy to address available infrastructure, change management, and effect a significant improvement in literacy levels speedily. Our hypothesis was that just introducing a multi-sensory learning experience with the teacher having the time to focus on learners will result in significant improvement in literacy levels. Besides, personalized guidance is moot when the learners in these schools barely have shared access to a computer for 30 min a week!

With the success at scale that this intervention has achieved, we now have the data, credibility to roll out subsequent phases including personalized guidance. Subsequent phases are being gradually rolled out and will report on them separately in the future.

3.2. 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 [Schramma, 2016]. 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 et al., 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 & 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 starting to exhibit any fluency in visual word recognition. Blomert & 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.

We view MSLE as the first step in a broader framework of AI-based approaches to improve learning outcomes. After MSLE is adopted, more methodical attempts can be made to see if there is further specialization within humans in their ability to acquire specific types of intelligence as a function of word types to different content forms – text, graphs, audio, to different types of content like language, numbers, pictures, etc.

While MSLE holds much promise to accelerate reading and comprehension 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.

3.3. The RightToRead intervention

In 2013, the RightToRead intervention was launched 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 5000 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 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. <u>No additional or new study material is introduced to the students</u>.
- 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.

3.4. Illustrative system interfaces

Figs. 2–6 are illustrative visuals from ReadToMe use in class room settings.

Fig. 2 is a view of ReadToMe in use in a class room in a 1:n (students) setting. RTM takes over the reading task from the teacher who is now free to focus on whether the students are able to read or not. The teacher can control how RTM interacts with the student by configuring many aspects of its interaction with the students.

As shown in Fig. 1, the RTM AI Engine trains itself on the text book and the content becomes MSLE enabled. RTM can then read the text book, and interact with the learner leveraging its many reading and comprehension features. It automatically maps the words to various

Fig. 2. Rtm in a physical classroom.

internal learning mechanisms like pictures and senses. The automatically created MSLE content can be adjusted further by a human expert using a configurator as shown in Fig. 3.

Fig. 3 outlines the major features of ReadToMe which have been implemented in phase I of the intervention.

A major aspect of learning to read is to be able to syllabify a word and to pronounce it properly. As shown in Fig. 5, RTM enables the student to internalize and practice the syllabic and phonemic structure of the word.

Fig. 6 is a screen shot from the automated Assessor as shown in Fig. 1, which evaluates the learner's reading and compares it with a normative benchmark. A number of aspects of the learner's reading are evaluated including reading speed, pronunciation including measuring pauses and punctuation. The normative benchmark is how the sentence/word is spoken by a native speaker from a specific geography, say the U.S. In addition to the immediate feedback to the learner, the automated assessor is designed to learn from cross-sectional data to inform the teacher on systematic patterns in reading difficulties across different sub-groups of the learning population.

The automated Assessor and Remediator were not part of phase I implementation and will be part of phase II of this intervention.

3.5. Implementation model and process

The RightToRead implementation model relies on leveraging existing infrastructure, personnel, and operating processes (Fig. 7). Strong partner support and collaboration was vital for the implementation and smooth running of RightToRead.

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

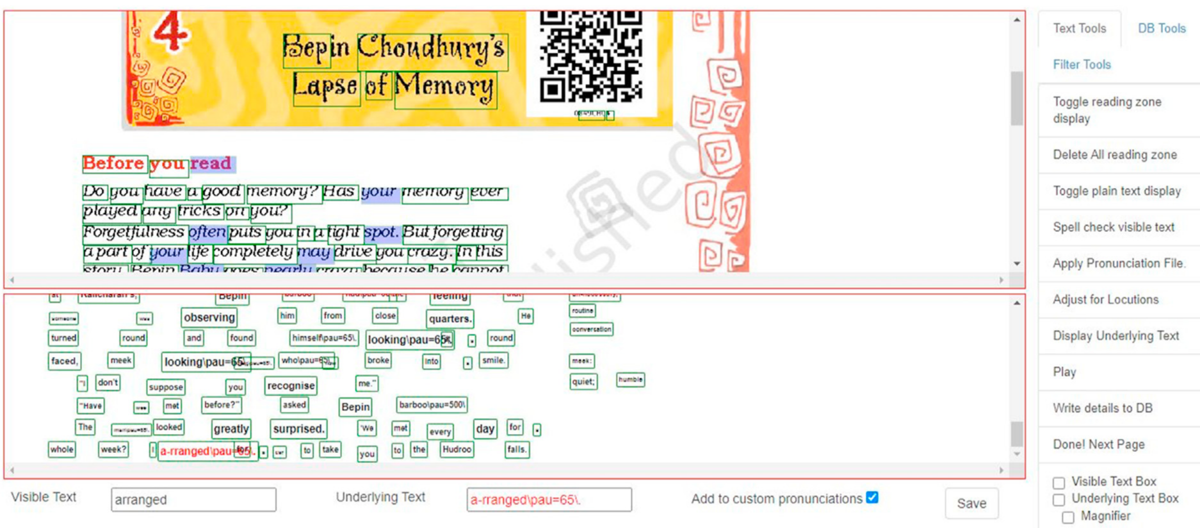


Fig. 3. Configuring RTM MSLE content.

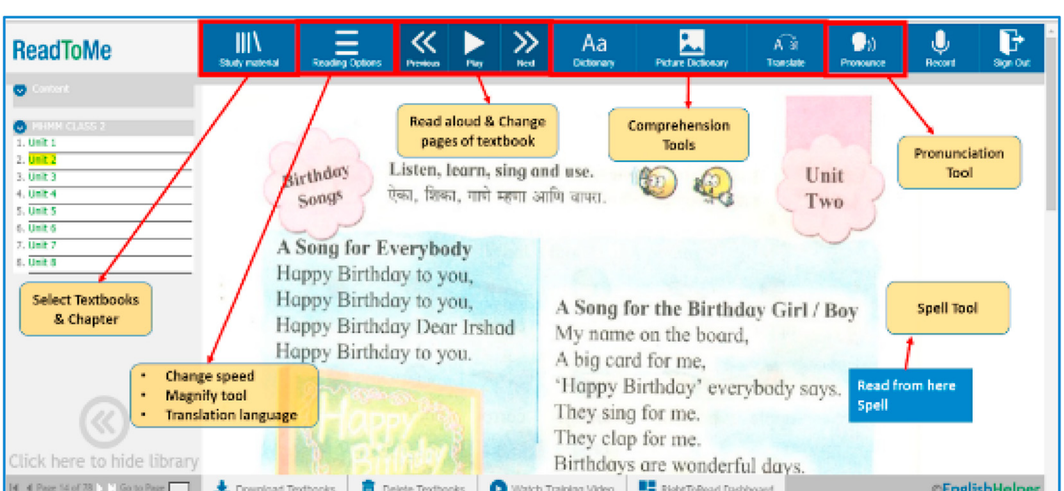


Fig. 4. Major features.

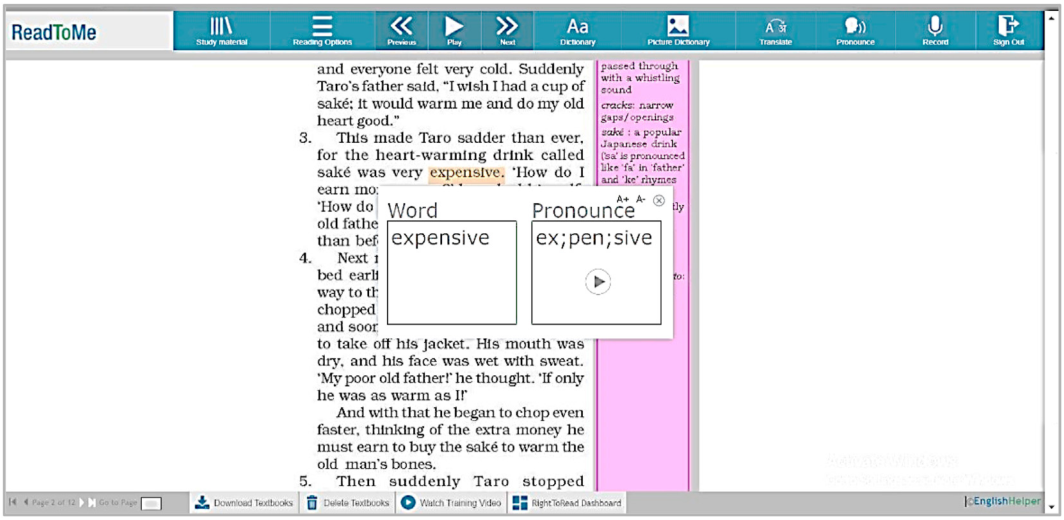


Fig. 5. Syllabification and pronunciation.

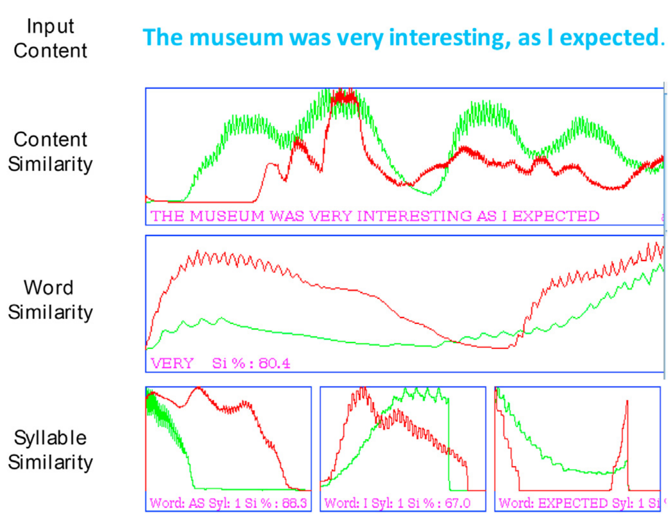


Fig. 6. Personalized reading improvement.

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).

3.6. 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 classroom use of ReadToMe was suggested to be at least 3 classes per

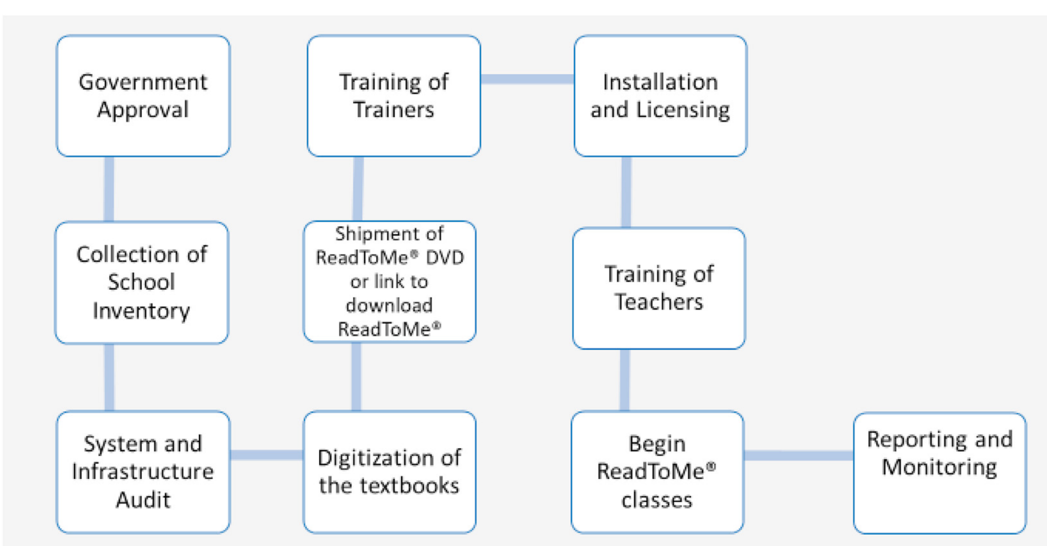


Fig. 7. The implementation process.

grade per week and a total of 75 classes per grade in the academic year. This was based on earlier studies of efficacy from RTM implementations. Our 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.

4. Impact assessment: the initial RCT study

This large-scale rollout of RightToRead covered over 1M students in grades 3 to 8 and touched 15,000 teachers in 5000 schools across 8 States – Punjab, Delhi, Gujarat, Maharashtra, West Bengal, Tamil Nadu, Telangana and Karnataka. Fig. 8 shows the students, teachers and schools that were part of the rollout.

RightToRead was implemented in the states of Maharashtra and West Bengal in partnership with SchoolNet (formerly IL&FS Education). In the remaining six states, the American India Foundation (AIF) were the implementation partner. Gray Matters India (GMI) were engaged 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 were selected. Baseline and End line assessments were administered in these schools. Assessments were conducted in these states. In addition, the assessments followed a standardized rubric across all states and grades assessed. Details of the assessment instrument and rubrics, with illustrative examples, are presented in Appendix A. Further state wise details of the assessment results are reported in RightToRead (2017).

4.1. 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 shown below in Fig. 9. The timeline for the assessments was dependent upon the academic calendar period stipulated by each state.

4.2. Assessment sample

More than 33,000 students were assessed across the four states (see Fig. 10) and between Treatment and Control groups.

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.

As shown in Fig. 11, 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%).

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

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**

Fig. 8. RightToRead coverage 2016-17.

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

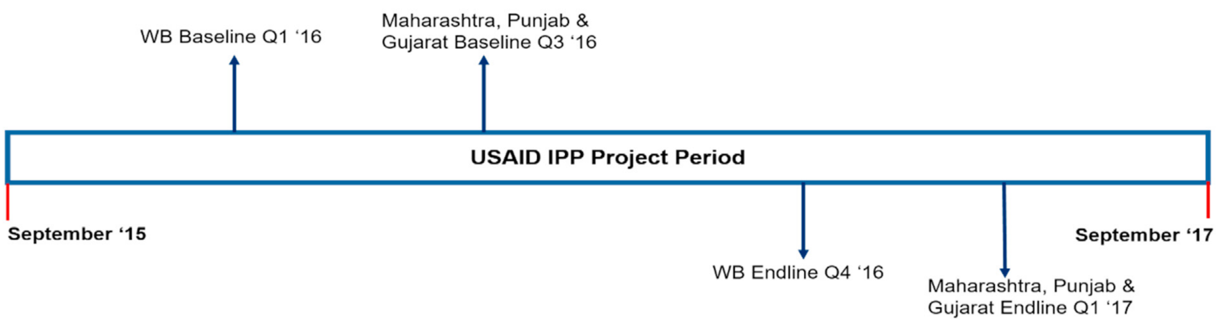


Fig. 9. Assessment schedule (Q = calendar quarter).

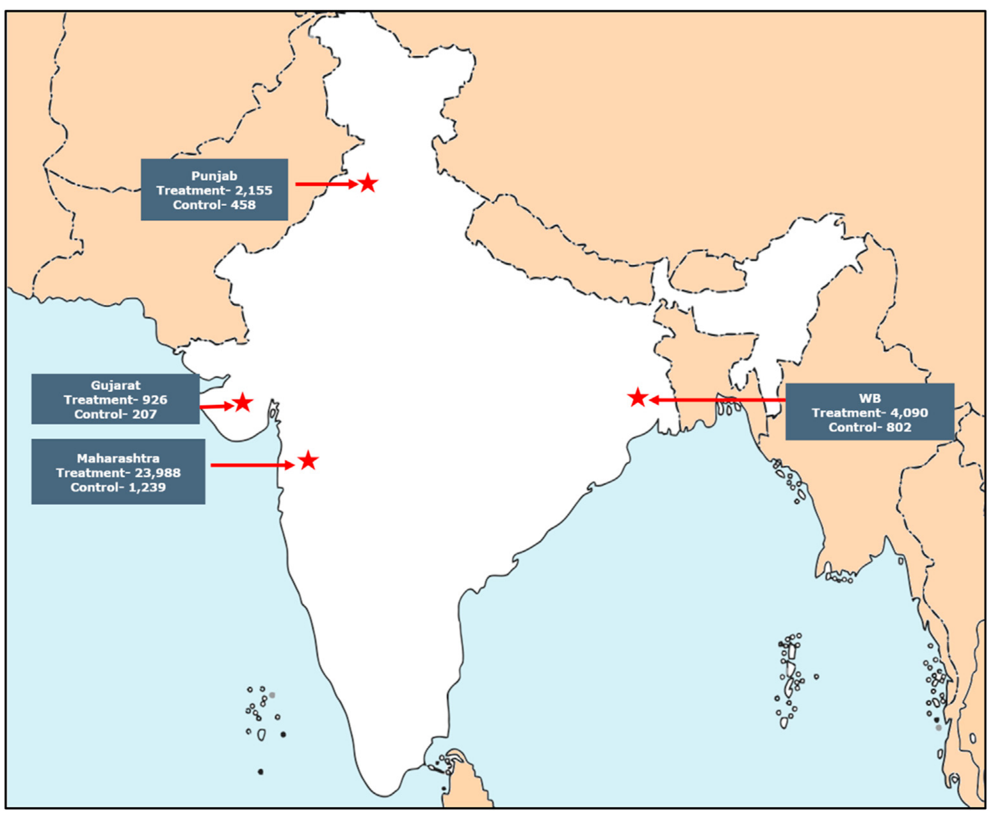


Fig. 10. Sample distribution.

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%

Fig. 11. Maximum state-wise sampling error.

between STAMP, us, 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.

4.3. 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 our teams 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 via 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 our 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.

4.4. 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. 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 us. This was checked 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.

5. Results

Analysis of the assessments show 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).

Outcome at an overall level is discussed first. For state-wise analysis, see RightToRead (2017). Throughout the analysis, two cohorts are reviewed – 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

The primary grades of 3–5 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. Fig. 12 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.

Both cohorts witnessed a 20% improvement of the Treatment group compared to that of the Control group (Fig. 13), with improvement across the grade-within-state cohorts 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.

5.1.1. 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. Fig. 14 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

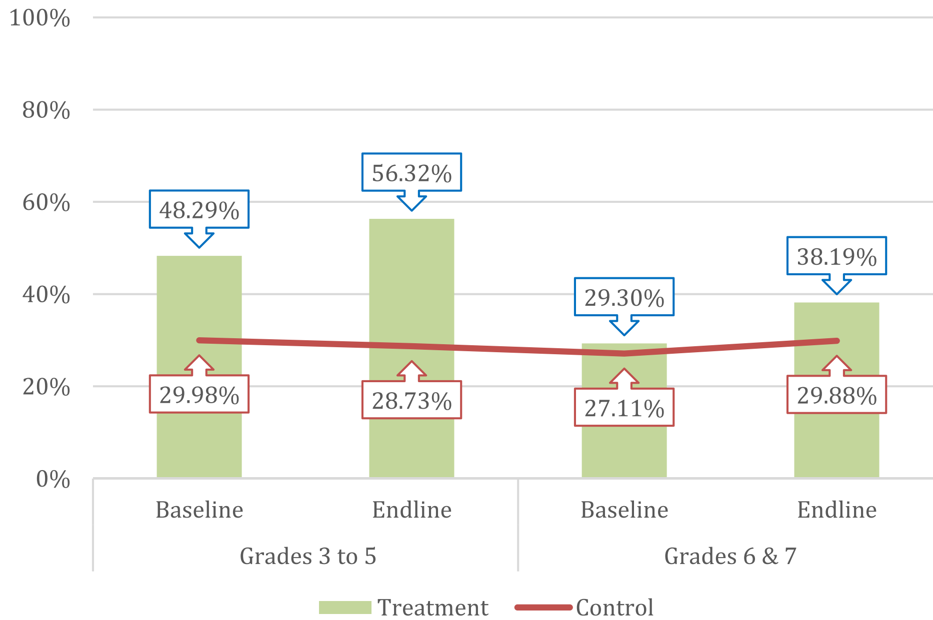


Fig. 12. Change in scores from Baseline to End line – Overall.

End line, in the Treatment group. In comparison, the top one-fourth of students' scores in the Control group has declined by 5 points.

In both grade cohorts, the Treatment group exhibits increase in the scores in every quartile (75th percentile, median 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.

5.2. 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), the Common European Framework of Reference (CEFR) was adopted 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

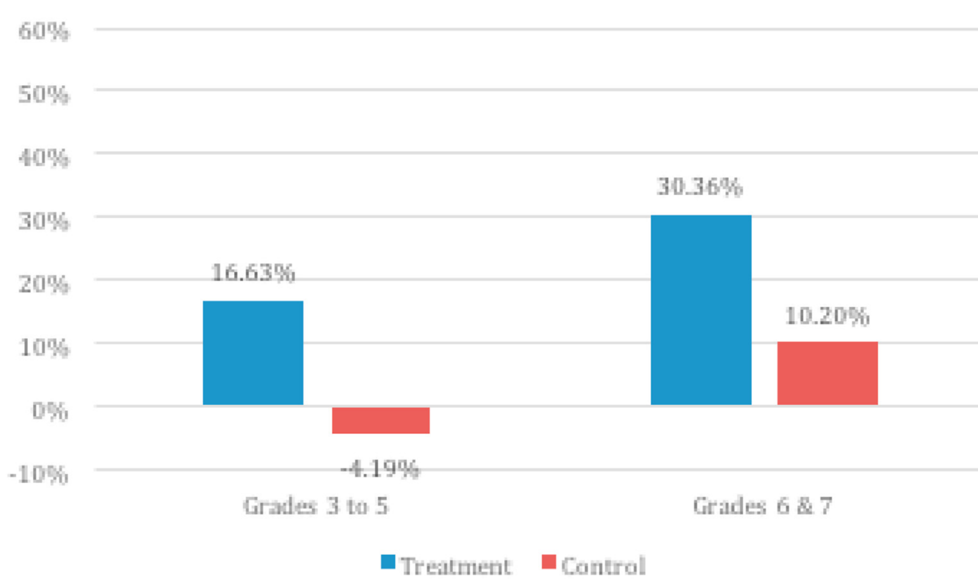


Fig. 13. Improvement of treatment and control – overall.

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%**

Fig. 14. Improvement in quartiles – overall.

4 Comparing the CEFR distribution of students in the Baseline and in the End line

5.2.1. 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 is shown in Fig. 15.

5.2.2. 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 Fig. 16.

5.2.3. 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.

5.2.4. 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

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

Fig. 15. Adapted CEFR for overall reading comprehension.

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.

Figs. 17 and 18 llustrate 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.

	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%

Fig. 16. The CEFR composition of assessment instruments.

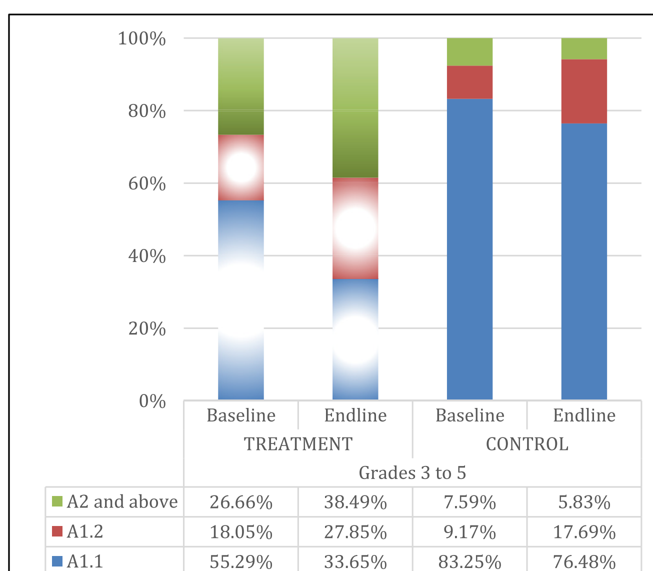


Fig. 17. Comparison of the CEFR achievement between Baseline and End line – Grades 3 to 5.

5.3. Ongoing impact

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.

The improvement of mean scores obtained by the students in 2017–18 over the Baseline scores, is presented in Fig. 19 below.

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. Fig. 20 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.

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.

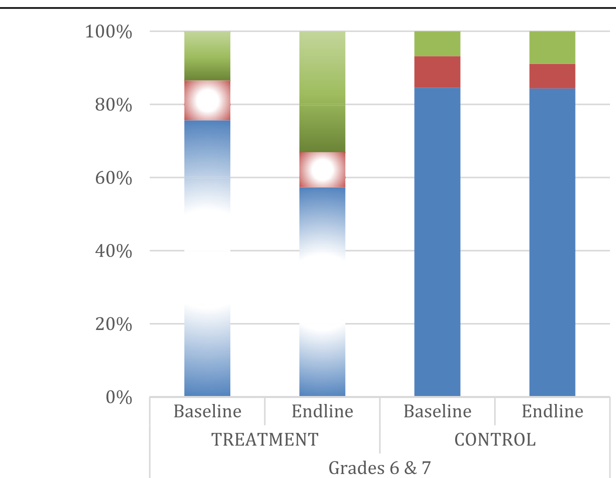
5.4. Relationship between ReadToMe sessions and improvement

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. Randomized sample assessments were conducted 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.

We recommend 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.

Fig. 21 below shows the mean number (and percent) of ReadToMe sessions that were conducted in each region in the academic year.

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.



	Baseline (TREATMENT)	Endline (TREATMENT)	Baseline (CONTROL)	Endline (CONTROL)
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%

Fig. 18. Comparison of the CEFR Achievement between Baseline and End line – Grades 6 & 7.

Presented below (Fig. 22) is the total (End line) count of students in each region.

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. Fig. 23 below presents the mean Baseline scores in each region.

The figure below (Fig. 24) 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.

It can be observed that with higher ReadToMe usage (bars), significantly higher gains in improvement (line) can be achieved in learning outcomes. There is clear evidence of a positive relationship between ReadToMe usage and improvement in learning outcomes, across grades and geographies.

6. Conclusion and future directions

We have reported the impact 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 8 million students besides several countries. Besides India, the intervention has also been successfully rolled out in Sri Lanka, Bangladesh, Nepal, Vietnam, Sierra Leone, Nicaragua, Honduras, Guatemala and Mongolia. By the end of the 2020 academic year (Mar/Apr 2021), it is expected be rolled out to approximately 20 million students in k-12 schools in India, all 10,194 schools in Sri Lanka constituting approximately 4.2 million students. Results in all the countries have been positive so far and we expect all countries to scale as the initial pilots prove efficacy to the authorities.

In India, the intervention both through a rigorous RCT based assessment and subsequent ongoing studies, demonstrated the sustained efficacy of multi-sensory engagement to effectively optimize the cognitive load on students. Improvement was observed across all levels consistently compared with the Control group, indicating that the intervention positively impacts learning across grades (ages) and for students at all levels of English proficiency. Further normalizing the assessment instrument and results according to the CEFR in order to enable comparison across geographies, the Treatment group showed reading improvement across all levels. Assessment in other countries are in various stages reflecting the stage of implementation. All of them report highly positive outcomes so far. In Sri Lanka, the results from an initial pilot at 51 schools led to an expansion to 384 schools and the outcome from the expanded pilot has led to an approval to deploy the intervention to all 10,194 schools in the country. Similarly the initial implementation in 2 schools in Sierra Leone has expanded to 6 schools in Oct 2020. For details, see Appendix B.

The intervention and ongoing assessments clearly demonstrate the power of learning science enabled technology to beneficially impact learning outcomes and specifically language acquisition and comprehension. This has broad implications for learning any subject.

The RightToRead program has demonstrated that it is possible to massively scale impactful educational technology to impact entire populations. It is significant to note that the technology does not introduce any new pedagogical content. Of note also is that it has been successfully scaled in environments with weak infrastructure and significant change management challenges.

We envision 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 learning outcomes in schools. While the RightToRead program is focused on English language reading and comprehension, the technology has much broader implications for reading and comprehension of all subjects. Besides, the program provides a potential template for how to roll out educational technology in large scale.

In the near future, the next level of the AI framework will be

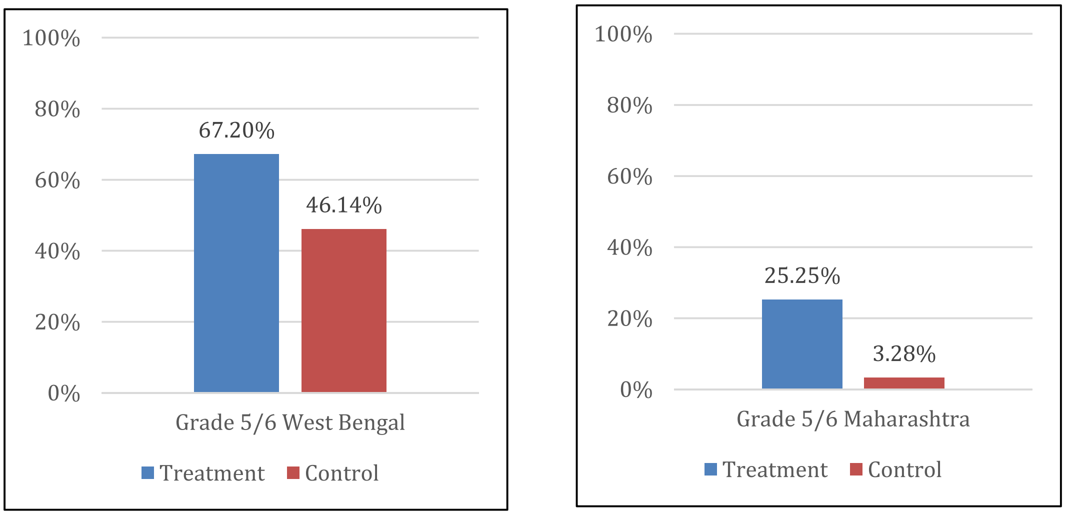


Fig. 19. Two-year improvement of Treatment and Control West Bengal and Maharashtra – Grades 5/6.

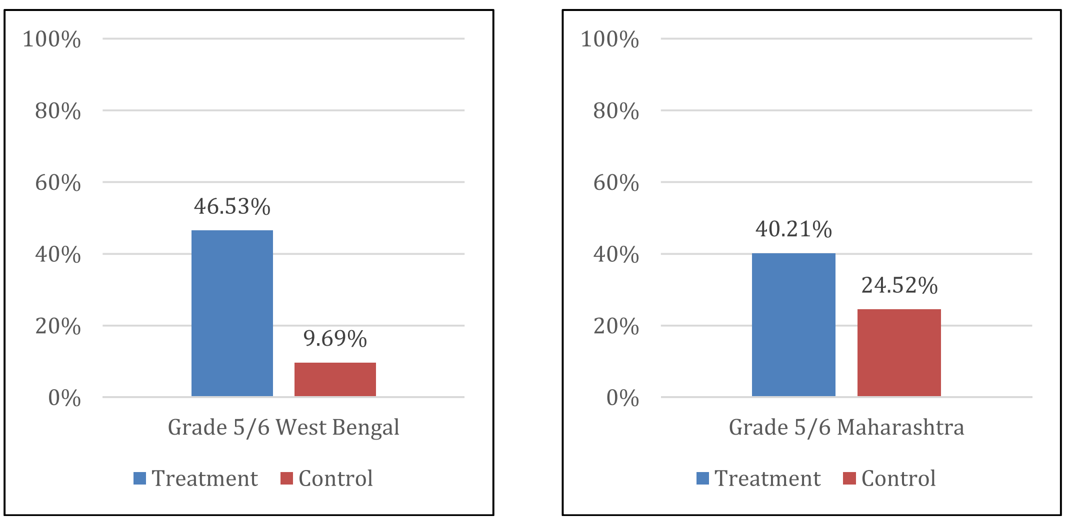


Fig. 20. Two-year change in Percentage of students in higher CEFR levels (A1.2 and higher): Treatment and Control – West Bengal and Maharashtra – Grades 5/6.

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

Fig. 21. Percent of ReadToMe sessions conducted in the academic year.

	Number of Students
Uttar Pradesh	249
Telangana	518
Delhi	1521

Fig. 22. Sample size 2017–18.

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

Fig. 23. Mean Baseline scores.

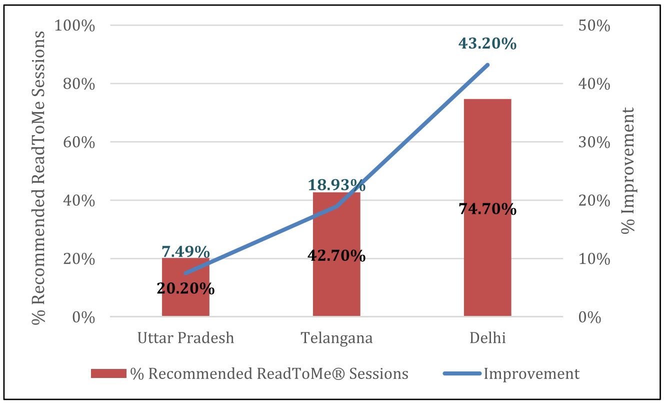


Fig. 24. Usage versus improvement – 2017–18.

implemented. This level introduces adaptive personalization at the learner level and also will provide insights at a learner group (class, school, district and so on) level which can inform curriculum choices. The growth in ownership of personal computers in the demographic serviced by the intervention accompanied by the acceptance of the intervention by teachers and students alike is enabling the next level roll out. Results from the next level implementation will be reported in the future.

Declaration of competing interest

None.

Acknowledgements

We acknowledge the helpful comments of M.S. Vijay Kumar, Associate Dean for Open Learning at MIT. We thank the USAID Grant – India Partnership Program APS# APS-386-13-000003 Cooperative Agreement No: AID-386-A-15-00019 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 SchoolNet for being the lead implementation partner. We also thank the American India Foundation (AIF) for being an implementation partner, Gray Matters India (GMI) for their instrument design and analysis and 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, SchoolNet, GMI or STAMP.

Appendix A. 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 B. Footprint outside India – Status as of Jan 2021

Asia

Other than in India, in Asia the RightToRead program has been adopted in Sri Lanka, Nepal, Bangladesh and Vietnam.

Sri Lanka

RightToRead was undertaken as a pilot in 51 schools on the island with 21,000 students using ReadToMe in the pilot phase of the program. The program met with high levels of adoption. Presented in the following table is usage of ReadToMe in comparison to usage of standard smart classes.

District	Schools	% Smart Class usage	% of English sessions conducted with ReadToMe
Kurunegala	8	38.13%	60.23%
Gampaha	9	21.94%	30.85%
Colombo	5	65.00%	71.78%
Polonnaruwa	12	50.83%	64.37%
Rathnapura	17	68.82%	86.07%
Total	**51**	**51.13%**	**65.76%**

Fig. B.1. Comparison of Smart Class Usage with RightToRead Usage – Sri Lanka

The success of the pilot encouraged a further 384 schools with approximately 150,000 students to contribute the hardware for the project. This was on their own initiative after seeing the success of the pilot program.

	Count
Districts	6
Schools	435
Students	273,386
Teachers	600
Grades	3 - 8

Fig. B.2. RightToRead Footprint – Sri Lanka, 2019–2020

Encouraged by the adoption, the Government of Sri Lanka approved the deployment of the program to all 10,194 schools (4.2 million students) in the country.

Nepal

RightToRead was launched in Nepal in May 2019. The pilot in 48 government schools of 4 provinces across 8 districts was initiated in partnership with NELTA (Nepal English Language Teacher's Association). The program covered around 5400 students across the nation and was the first AI technology-based learning software introduced in the country.

The pilot was for a duration of one year (Academic Year 2019–2020). The program was monitored through a virtual tracking system providing real-time usage data.

	Count
Districts	8
Schools	48
Students	5,400
Teachers	100
Grades	3 - 5

Fig. B.3. RightToRead Footprint – Nepal, 2019–2020

Bangladesh

RightToRead was launched in Bangladesh in October 2016. The program was implemented in two schools. Teacher feedback has been very encouraging: teachers report increased attendance during and high enthusiasm in the RightToRead sessions. They cite improved pronunciation of students and teachers alike, resulting from the use of ReadToMe.

	Count
Districts	2
Schools	2
Students	171
Teachers	4
Grades	3 - 5

Fig. B.4. RightToRead Footprint – Bangladesh, 2019–2020

Vietnam

In 2018, RightToRead went live in Koto Vocational Training Centre, Hanoi, Vietnam which provides training to young individuals in hospitality and English language.

It was a pilot for 3 months (November 2018 to January 2019) where ReadToMe was integrated in their English language classes. A total of 45 young individuals were enrolled in the program where ReadToMe was integrated in their English language syllabus. ReadToMe was extensively used to practise reading and pronunciation of simple words and sentences.

Teachers report that students had a positive experience with the program. Students enjoyed it, they exhibited progress in in-house tests and evidenced improved vocabulary.

	Count
Districts	1
Schools	1
Students	45
Teachers	3

Fig. B.5. RightToRead Footprint – Vietnam, 2018–2019

Africa

Sierra Leone

RightToRead was launched in Sierra Leone in 2015 in partnership with Love Sierra Leone (LSL). By September 2018, the program expanded to 2 schools educating almost 350 girls.

Year-on-year results indicate constant improvement in overall aspects of the English language skills and interpersonal skills of students. Assessments conducted in August 2019 showed an average improvement of around 36% in the first year of ReadToMe Usage.

Teachers have repeatedly expressed that ReadToMe has been helping the students pronounce English words correctly, build confidence in public speaking and improve language skills.

A number of schools have expressed interest in implementing the program. In October 2020, RightToRead was launched at 6 more schools with over 2000 Junior Secondary students in total.

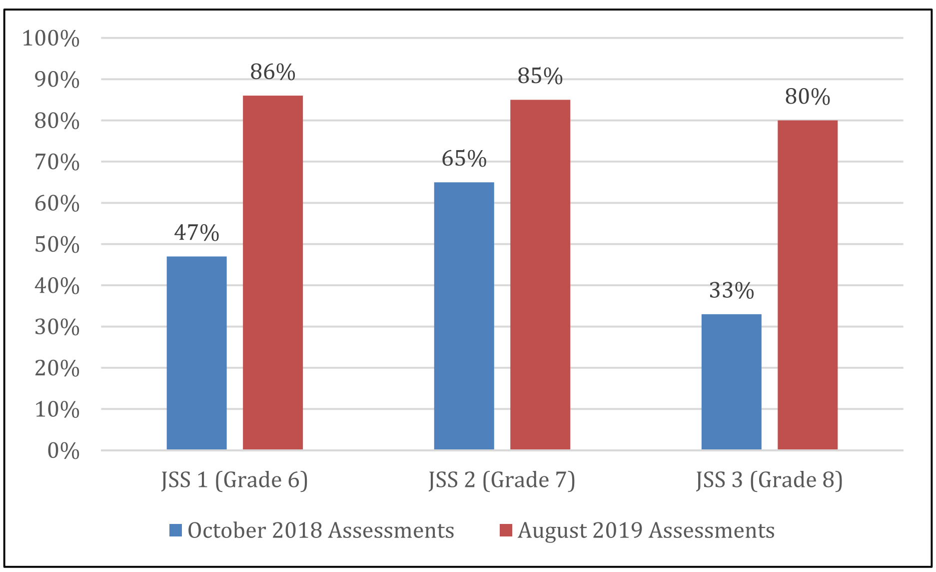


Fig. B.6. Improvement in Outcomes – Sierra Leone, 2018–2019

	Count
Districts	1
Schools	7
Students	2,450
Teachers	14
Grades	6 - 8

Fig. B.7. RightToRead Footprint – Sierra Leone, 2020–2021

Latin America

Nicaragua

The RightToRead program was initiated in Central America in 2017 in a subsidized school called Sagrada Familia in Ticuantepe, Nicaragua. Students took extra classes/hours to learn English using the program.

In 2018, the school incorporated English classes in their regular schedule using the EnglishHelper learning content. In 2019, the school acquired

English books to use with the ReadToMe software. Results showed a 20–40% increase in their grades.

In 2019, 2 other schools in Nicaragua (Chinandega and Granada) started using ReadToMe as English reinforcement classes, in addition to the regular English classes.

	School 1
Grades Assessed	2 - 7
Number of Students	135
Average Increase in Grades	1.43x
Minimum Increase	1.17x
Maximum Increase	2.5x
Initial Weighted Average	28%
Final Weighted Average	38%
Improvement in Weighted Average	*35.71%*

Fig. B.8. Improvement in Outcomes – Nicaragua, 2018–2019

	Count
Districts	3
Schools	3
Students	240
Teachers	6
Grades	1 – 10

Fig. B.9. RightToRead Footprint – Nicaragua, 2019–2020

Honduras

The RightToRead program in Honduras was initiated in 2017 in 3 schools sponsored by a private group, in the regions of Tela and Trujillo. They enabled their own textbooks on ReadToMe.

After excellent results in 2018, the schools continue to use the program and have incorporated new textbooks.

	School 1	School 2	School 3
Grades Assessed	1 - 9	1 – 9	1 - 9
Number of Students	221	205	246
Average Increase in Grades	2.44x	1.49x	2.40x
Minimum Increase	1.67x	1.03x	1.92x
Maximum Increase	3.12x	2.44x	3.22x
Initial Weighted Average	30%	37%	32%
Final Weighted Average	73%	53%	75%
Improvement in Weighted Average	*143.33%*	*43.24%*	*134.38%*

Fig. B.10. Improvement in Outcomes – Honduras, 2018

	Count
Districts	2
Schools	3
Students	782
Teachers	3
Grades	Kinder - 11

Fig. B.11. RightToRead Footprint – Honduras, 2019–2020

Kindergarten and Pre-school

The RightToRead program is also used in a network of around 15 kindergarten and pre-schools in public schools in the region of Tela, Honduras. This was implemented for the first time in 2018 after teachers took a mandatory evaluation test. After excellent results in 2018, the schools have continued using the program. In addition, RightToRead has implemented an adult English literacy program to enable more teachers to teach English at their own kindergartens, pre-schools and primary schools.

	Pre-school	Kindergarten
Number of Schools Assessed	14	10
Number of Students	103	47
Average Increase in Grades	4.2x	20.47x
Minimum Increase	1.07x	1x
Maximum Increase	10.54x	58.33x
Initial Weighted Average	24%	2%
Final Weighted Average	64%	24%
Improvement in Weighted Average	*166.67%*	*1100%*

Fig. B.12. Improvement in Outcomes – Honduras – Kindergarten and Pre-school, 2018

	Count
Districts	2
Schools	15
Students	238
Teachers	16
Grades	Kinder and Pre-school

Fig. B.13. RightToRead Footprint – Honduras – Kindergarten and Pre-school, 2018–2019

RightToRead, Guatemala

The RightToRead program was implemented in Guatemala in 2019 running pilots in 5 different schools: 2 private schools and 3 public schools, impacting more than 1300 students and 19 teachers. All grades from Pre-school to 11 implemented the program.

	Count
Districts	5
Schools	5
Students	1,354
Teachers	19
Grades	Pre-school - 11

Fig. B.14. RightToRead Footprint – Guatemala, 2019–2020

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.caeai.2021.100019.

References

Abdulrahaman, M. D., Farukb, N., Oloyedeb, A. A., Surajudeen-Bakindec, N. T., Olawoyinb, L. A., Mejabia, O. V., Imam-Fulanib, Y. O., Fahmd, A. O., & Azeeze, A. L. (2020). Multimedia tools in the teaching and learning processes: Asystematic review. *Heliyon, 6*, Article e05312, 2020.

Ahmadi, M. R. (2018). The use of technology in English language learning: A literature review. *International Journal of Research in English Education (IJREE), 3*, 2, 2018.

Azam, M., Chin, A., & Prakash, N. (2013). The returns to English-language skills in India. *Economic Development and Cultural Change, 61*(2), 335–367. January.

Banerjee, A. V., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "teaching at the right level" in India* (Vol. 16). Cambridge, Mass: MIT Department of Economics Working Paper. ZDB-ID 21886593, 08.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics, 122*(3), 1235–1264.

Blomert, L., & Froyen, D. (2010). Multi-sensory learning and learning to read. *International Journal of Psychophysiology, 77*, 195–204.

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.

Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR medical education, 5*(1) (2019), Article e13930.

Chapelle, C. A. (1997). CALL in the year 2000: Still in search of research paradigms. *Language, Learning and 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.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., et al. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split brain patients. *Brain, 123*, 291–307.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review, 101*(5), 1739–1774.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is the neocortex essentially multisensory? *Trends in Cognitive Sciences, 10*, 278–285.

Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language, Learning and Technology, 23*(1), 8–25. http://llt.msu.edu.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105.

Hwang, G. J. (2014). Definition, framework and research issues of smart learning environments - a context-aware ubiquitous learning perspective. *Smart Learning Environments, 1*(1), 4.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision challenges, roles and research issues of artificial intelligence in education. *Editorial. Computers and Education: Artificial Intelligence, 1*, 2020.

Iiyoshi, T., & Vijay Kumar, M. S. (2010). In Iiyoshi, & V. Kumar (Eds.), *Opening up education: The collective advancement of education through open technology, open content and open knowledge*. MIT Press, 2010.

Lai, C. (2013). A framework for developing self-directed technology use for language learning. *Language, Learning and Technology, 17*(172), 100–122.

Lai, C. (2017). *Autonomous language learning with technology beyond the classroom*. London & New York: Bloomsbury Academic.

Lai, C., & Gu, M. (2011). Self-regulated out-of-class language learning with technology. *Computer Assisted Language Learning, 24*(4), 317–335.

Liu, M., Moore, Z., Graham, L., & Lee, S. (2002). 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, 34*(3), 250–273.

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.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 317–334.

Panda, S. (2017). *There is a language to employability; it can make a difference to pay, even employment*. Opinion: Financial Express. Dec 25 https://www.financialexpress.com/opinion/there-is-a-language-to-employability-it-can-make-a-difference-to-pay-even-employment/988623/.

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.

ead: Large scale deployment of tech enabled English learning, report on learning outcomes, 2017. https://www.englishhelper.com/dashboard/righttoread-impact.

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

Schramma, E. (2016). *Multisensory learning and technology in the acquisition of English as a second language. Working paper*. EnglishHelper, Inc.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*, 411–417. https://doi.org/10.1016/j.tics.2008.07.006

Yang, J. (2013). Mobile assisted language learning: Review of the recent applications of emerging mobile technologies. *English Language Teaching, 6*(7).

Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal, 21*(1), 7–27. Retrieved April 18, 2021, from http://www.jstor.org/stable/24149478.