Enhancing ESL Education in India with a Reading Tutor that Listens

Frederick Weber
Adjunct Sr. Research Scientist
Earth Institute, Columbia University
New York
+1-917-470 9211
fw2174@columbia.edu

Kalika Bali
Researcher
Microsoft Research Labs India
Bangalore, India
+91080-66586218
kalikab@microsoft.com

ABSTRACT
We report results of a 2 ½-month pilot study of Project Listen's PC-based Reading Tutor program for enhancing English education in India. Our focus was on low-income elementary school students, a population that has little or no exposure to English outside of school. The students showed measurable improvement on quantitative tests of reading fluency while using the tutor. Post-pilot interviews explored the students' experience of the reading tutor. Further, a survey of educational programs gives a picture of the wide range of institutions providing training in English in and around Bangalore to low-income populations. Each has associated infrastructure, personnel, and curricular constraints that would be faced by interventions like the reading tutor, even if it can be shown to be effective. The perceived advantages of literacy software and associated measures of success also vary by program.

Categories and Subject Descriptors
K.3.1 Computer Uses in Education

General Terms
Measurement, Experimentation, Human Factors.

Keywords
speech recognition, ESL, educational technology, literacy.

1. INTRODUCTION
English in post-independence India has retained its importance as a means for upward mobility, despite the linguistic diversity of the country. It is the secondary official language of the country, after the primary official language Hindi, and is so recognized in the Indian Constitution, ahead of the remaining 22 other official languages. Especially outside the Hindi-speaking north, English is the language of higher education, business management, and the professional classes.

The primary motivation to learn English as a second language is instrumental [1, 2], as it is widely perceived to be significant in enhancing social status. It also serves as a gateway to a wide range of employment opportunities, from working in the stores of the upscale shopping malls to serving to differentiate between applicants for basic entry-level employment in business. It is much easier to teach basic job skills than to teach a new employee English and a good command of English is automatically associated with competence, good education, and intelligence. As the principal of one school put it, "With these 26 letters, we can rule the world". This puts the large majority of children going to schools where instruction is in their local language (Tamil, Kannada, Telegu, etc.) at a distinct disadvantage.

Given the stakes, it is not surprising that many aspirational parents will make extensive sacrifices to obtain an English education for their children in private schools. However, a lack of trained teachers, especially for low-resource schools catering to the lower socio-economic strata of the society, makes it difficult to provide consistent quality of English teaching in India. Large class sizes and a focus on completing the prescribed syllabus mean that individual attention and practice in English classes are minimal in most schools. A strong emphasis on the end of the year examination leads instructors to target written skills at the cost of comprehension and spoken skills [3]. However, it is precisely the skills of comprehension of and self-expression in spoken English that confer the benefits in social status and upward mobility.

There is significant potential for introducing technology to enhance the reach and effectiveness of English language teaching in the resource-limited Indian context. A computer-based reading tutor can address all of the problems mentioned above: it provides extensive individual practice of spoken skills, it is consistent with every student, and it can assist and test comprehension and spoken fluency. We present here results from a recently concluded pilot program to evaluate a speech-recognition based reading tutor

ACM DEV'10, December 17-18, 2010, London, United Kingdom.

built by Carnegie Mellon University's Project Listen [4] in three educational programs in and around Bangalore, Karnataka in southern India. Our goals are to establish if the reading tutor is effective for Indian students from disadvantaged backgrounds.

However, effectiveness is a necessary but insufficient condition for overall success—the software must be deployable across the target population, Hence, we also discuss the constraints faced when deploying any such technology across a range of Indian educational institutions, based on a survey of eight such programs in the Bangalore area. This sampling has been chosen to provide case studies that represent the range of potential programs—formal vs. informal, high or low resource, public or private—available in the city. We restrict ourselves to programs serving the lower-income demographics, as this population is will have the hardest time finding high quality English training. We believe that the proportions of programs in these categories will vary across states and regions, but the problems faced will be typical across much of India.

Since low-resource schools and programs face a shortage of fluent, trained instructors, but often have a small cluster of PCs, we believe that a technological solution is more appropriate, sustainable, and scalable than a program involving additional personnel or substantially more time from existing personnel. Accordingly, we have focused on whether the reading tutor itself can be an effective intervention with existing staff and resources, rather than comparing its effectiveness to alternate non-technological approaches such as guided reading with English-fluent volunteers. However, one study found the reading tutor to be nearly as effective as one-on-one reading practice with a teacher [5].

Earlier studies on the effectiveness of technological interventions in the Indian context for teaching in general, and of English as a Second Language (ESL), in particular, have reported mixed results. In a two year long study of nearly 10,000 children in schools in Western India, [6] show that children who lagged behind in the class had higher gains on teacher-implemented non-ICT interventions. The stronger children, on the other hand, benefitted more from the self-paced ICT intervention monitored by external assistants. In contrast, in another study similar in scale [7] claim that while both methods prove effective, teachers in the normal course do not have the time, and sometimes the inclination, to pay special attention to students who are left behind.

In this study, we attempt to evaluate infrastructural constraints, in terms of both technological as well as personnel support, along with the attitudes towards interventions that may be detrimental to successful implementation of a program. As the program itself was narrowly focused on one skill, viz., reading fluency in English, we were able to concentrate on the fundamental constraints rather than complex pedagogical issues.

This study is similar in construction to a study deploying Project Listen technology in Ghana [8]. While the success of the Ghanaian study was encouraging, it by no means ensured that the program would be effective in a very different cultural context. One important difference is that the medium of instruction in Ghana is English—in India many students attend school in their vernacular language through 10th standard (grade). This study deployed the reading tutor with students from both English-medium and Kannada-medium schools. Also, the longer duration study in Ghana relied on bussing the students to a central location to use the computers there; we focus here on what can be done with minimal intervention in the existing school or program.

A similar investigation into the efficacy of the reading tutor for ESL students suggests positive gains for some of the students in low-resource schools in Canada [9]. However, in the Canadian study English is widely spoken and encountered outside of school. For most students in our study, any exposure to English is mainly confined to the school environment, and even there, they may not have regular access to any native/fluent English speaker. We wanted to confirm that the reading tutor would still be accepted and effective given these constraints.

The Millee project [10] deployed cell-phone based games in rural India to develop literacy skills in English. However, the games focus on either listening comprehension, vocabulary building, or spelling skills. Similarly, the technologies in studies [6, 7] do not test spoken English fluency. To our knowledge this paper is the first deployment of a speech recognition-based "listening" technology in India with a focus on developing spoken English skills.

2. THE READING TUTOR

Project Listen's reading tutor is PC-based software that uses story reading to enhance children's English reading fluency providing reading practice targeted to each student's individual reading level.

The reading tutor assigns each student an ID. Upon logging in, the student is given a choice of stories appropriate to their reading level. They are then shown a sentence at a time from their chosen story, which they read into a headset microphone. The program "listens" by using semi-constrained speech recognition, based on the Sphinx recognizer [11], to identify what the student said and track where they are in the sentence. If the student gets stuck, or makes a serious error in pronunciation, the program can prompt with word-specific help (at the point where the student is stuck) or have a prerecorded narration (from a US-accented native speaker) play the whole sentence back. Students can also click on words they do not recognize for pronunciation and definition help. Students are expected to use the program for 30 minutes a day.

The reading tutor keeps a record of all interactions with each student in a central database, so the tutor and the teacher can monitor the improvements of each student separately. By tracking speaking rate, the program can assess when a student is ready to advance to a new level of stories. It starts with simple games to ensure that the students have mastered letter-to-sound rules, then proceeds through advancing levels of story complexity (through 7^{th} or 8^{th} grade US levels). Note that the reading tutor does not test for reading comprehension—its perception of proficiency is based on reading rate alone.

The reading tutor can address several problems facing an Indian elementary or secondary school English program. It supports individual English reading practice in a forgiving environment. The only alternative is when the student is called upon to read in class—rarely for longer than a few minutes, not every day, and with the whole class watching. The narration, help, and instructions are provided by native English speaking voices, which may be the child's only exposure to fluent, native English speech. And the focus on stories and some games makes the practice of reading more enjoyable.

The software requires 1 GHz processor speed and 1 GB RAM to run smoothly, and exclusive use of the PC by a single student for their 30-min daily sessions. These specifications can be significant constraints for Indian educational institutions.

3. ENGLISH EDUCATION PROGRAMS

There are a wide range of educational institutions serving the area around Bangalore. For the purpose of the interviews, we divided these programs into the taxonomy below. Two schools from each category were selected for detailed interviews with the heads of the schools as well as teachers wherever applicable.

Government schools: Most of the population is served by government schools, with Kannada, the local language, as the medium of instruction. English is introduced in 3rd standard (grade) as a second language, though now this is being extended to oral instruction in lower standards. These schools typically work with severe resource constraints in infrastructure and personnel. They are also the most highly regulated segment of the educational spectrum, with most curricular decisions controlled by the state board of education.

Informal programs: To compensate for the limitations of government school education, many informal programs have been founded around the city by various groups. These supplement the instruction the children receive at their regular school, and often offer a meal to the children as well. The programs offered vary considerably, but in terms of infrastructure these are the most limited operations in our list. Many have some focus on English literacy.

Low-resource private schools: Those that can afford to typically send their children to private schools. The least expensive of these charge tuition of a few hundred rupees a month. They operate in either Kannada or English medium, and have somewhat better resources than the government schools. They have considerably more autonomy in how they deploy their funds, though they are required to follow one of two government-specified syllabi.

Sponsored private schools: These are a special case of private schools that target underprivileged children. No fees are charged, and the costs of this education are borne by a single benefactor or a variety of sponsors. As these are focused on lifting children out of poverty, they are typically English medium. They can have extensive resources available, but the children may still face significant challenges in achieving the aspirations that the school, parents, and the children themselves have for their future.

Of the eight schools surveyed, three educational programs in Bangalore were chosen to participate in the pilot: an informal afterschool program (S1); a low-cost English medium private school (S2), and a donor-supported English medium private school (S3). Although including a government school would have been desirable, lack of computing infrastructure and a long approval process prevented us from doing so. However, all the participants at S1 attended one of the government schools surveyed.

S1 is an afterschool program in an urban area, with 8-12 students at each standard, who attend between 4-6 pm, after their classes at the local government school. The medium of instruction in the government school is Kannada, the local language. English is introduced in 3rd standard as a second language. As a result, the 9th standard students at S1 have English language skills comparable to a 3rd or a 4th standard student in an English medium school. S1 had no existing computing resources—three laptops were lent to them to allow them to participate in the pilot. However, it was also the most flexible of the three programs in terms of curriculum, and found it easy to make room for the use of the reading tutor. The pilot was conducted in S1 with a group of 9 students from 9th std that ranged in age from 14 to 18 (there is no age restriction in government schools).

S2 is a school on the outskirts of the city, in an urbanizing area where many of the families are still farming. It follows the state government-specified syllabus with English as the medium of instruction. The school has 5 PCs that they use for computer-based instruction. Though these initially had inadequate memory to run the reading tutor, the principal had the machines upgraded within a week. One PC was also being used for administrative work; to ease the load on the school we loaned them one laptop for the pilot. 22 students from the 4th standard participated in the study.

S3 is also on the city's outskirts, and serves the very poor families (< 3500 Rs/month family income) in a 15 km radius around the school. It is a private sponsored school targeted at lifting their students out of poverty. S3 was the only pilot participant whose existing computing resources were ready to run the pilot. Their junior computing lab has 16 modern PCs on an intranet, allowing them to take advantage of the networked aspects of the reading tutor. In The school chose 30 students from the 3rd standard, mostly those in need of remedial support, to participate in the program.

4. SCHOOL SURVEY

For each of the eight schools we visited, we conducted a semi-structured interview with the principal and available staff. In all cases the principal or a senior staff member had a good command of English, though in one or two instances we used a native Kannada or Hindi speaker to get additional information. We then toured the computing facilities, if any, to determine available computing infrastructure at each school.

Our survey represents a snapshot of various school programs around Bangalore. We have chosen these 8 schools to try to cover the range of options in the city available to children in the lower-income demographics. There are many types of schools we have not considered, such as special central government schools (that require an entry test), and schools run by the state for the children of government employees—armed forces or Indian Administrative Service personnel, university professors, etc. There are also a wide range of private schools; we have focused on the most affordable. We believe that our cases are representative of the choices of the children most in need of English training support.

The discussions focused on four basic themes: the demographic groups served by the program, the importance of English literacy in the curriculum, computing and personnel infrastructure, and the attitudes toward technological interventions, including how to measure their success. The following sections summarize the results by theme, including anonymized comments from the principals and staff interviewed to illustrate some conclusions.

4.1 Demographic Served

The least privileged groups, such as the children of rag-pickers or

migrant labor, can only attend government schools, and may not graduate. Since these are the highest-risk populations, they tend to be the focus of the sponsored private schools and informal programs as well. S4 is a case in point, a residential boarding school that accepts 24 of the poorest children in its area at the age of four. The students stay in an immersive English-speaking environment except for 4 weeks of annual vacation until graduation at 10th standard. S3 targets the general population of the slums around it. It is non-residential, uses buses to extend its reach, and sets a cap of 3500 rupees annual family salary to gain admission. The two informal programs (S1 and S5) serve the children of their local government schools, who come from slightly more well-to-do families; their parents might be auto rickshaw or truck drivers, domestic servants, or construction workers. Though free, the programs are voluntary, and thus tend to select for the children whose parents have some aspirations for them. The low-resource private school, S2, serves the lower middle class in its urbanizing neighborhood on the outskirts of Bangalore. The parents are largely from agricultural backgrounds, and seek to get an English education for their children to move them into alternative career paths. S6 is a low-cost English medium private school in urban Bangalore, serving the lower middle class in its neighborhood. These two private schools were the only programs reporting significant parental involvement in the child's education. Class sizes seemed driven by available space; one government school (S7) had 25 students/class, the other (S8) 40-50 (and up to 80 in 8th std). This pattern was matched in the private schools—S6 has 40-45 students per class, while S2 has 25-30. S4 has around 20 children/class in lower stds and between 12 and 15 in the upper grades (6th-10th). The informal programs are smaller, with between 8 and 15 children in each class. All private schools (sponsored and fee-charging) had English as the medium of instruction; the government schools were Kannada medium, while the informal programs used a mix of both the languages.

4.2 English Literacy Program Support

At every level, the importance of knowing English for better future prospects was readily acknowledged.

> "I will go all out to see to it that the students learn proper English. That is their stepping stone to get into the job market"

The 4 private English medium schools (sponsored or fee-charging) have of course made it a central element of the curriculum, but the heads of the government schools and informal programs were also clear that a command of English would make a very significant difference in the career prospects of their students.

There were three main problems that each program faced to different degrees in providing effective English instruction: shortages of time and adequately trained personnel, a highly constrained curriculum, and the absence of spoken English outside the school.

> "Parents are not talking in English, neighbours are not talking, friends are not talking, it is only in the school premises that the child hears English"

The government schools were the most constrained, as they teach English as a second language, typically starting formal (written) instruction at 5th standard. Even so, there may not be an official English teacher as such until higher standards such as 8th. The English instruction until then is provided by the class teachers, who may have limited English skills themselves.

> "All the teachers in (the government school) are well qualified but they cannot teach English effectively as they were also taught in Kannada medium schools"

This can lead to English classes where students are tested on their ability to provide stock answers to predefined questions, with no real focus on comprehension or pronunciation. Despite this, the school is required by the board of education to adhere to a fixed curricular schedule which advances rapidly to very challenging levels of prose by 9th standard. Once behind, a student has little hope of ever catching up.

> "I see them struggling really hard [with English] when they come to 9th grade because they have no foundation in the early years"

From the standpoint of introducing a technology to enhance English training, there is no real flexibility in the rest of a government school's schedule or budget to accommodate remedial instruction on the premises. They also would have great difficulty in establishing a program outside of regular school hours; union rules do not allow them to keep teachers at the school late (even if they are willing), and schools could not risk the liability for teachers, students, and equipment in an off-hours program, once the rest of the staff have gone home. Finally, any technology intervention needs the intervention of the state school board; so long lead times are required to establish a pilot program.

The students can turn to supplemental training from afterschool programs to enhance the quality of their education. They typically provide 1 ½ to 2 hours of supplemental instruction between 4 and 6. One of the two programs we interviewed focused primarily on English instruction, but does offer classes on other subjects for specific groups of students. These programs are essentially volunteer-driven; thus both the teachers and the students tend to be committed, though ensuring quality of coverage for English instruction can be difficult. These programs have the greatest flexibility to adjust their curriculum to make room for a new intervention.

The four English-medium private schools necessarily have a much higher base standard for teachers' command of English than the government schools. They can still face a range of skill levels and a need to keep improving:

> "I insist that the teachers also converse with each other in English in the school, if they are not comfortable with the language how can they teach the children"

The overall curriculum is still set by the government (required for accreditation), but not to the level of weekly scheduling, which allows them some flexibility in setting intermediate goals. Most of these schools also have (or can provide) an open period where children can participate in additional training programs outside of the standard curriculum. However, within the classroom the teacher can still be hard-pressed for time to cover the curriculum, which limits opportunities for students to practice:

> "The English teacher goes to the class with a target, she has to complete a particular portion of the text-book, give (the students) notes, and make them go through the story in one week. She cannot make them all read in the class"

In terms of parental support for English practice, the sponsored private schools do not expect any, as most students are the first in their families to get an extensive education. The fee-paying private schools have some parental involvement in general, but only S6 reported roughly 20% of the students have parents with a command of English. S4's boarding school format was the only environment where English was routinely used outside of class.

4.3 Computing Infrastructure and Personnel

There were no real surprises in the available computing infrastructure to support a literacy intervention. The informal programs had either no PCs at all or a handful of older, donated machines. S1 was loaned 3 laptops for participation in the pilot work. The power was rather reliable between 4 and 6, so between line power and the laptop batteries was regular enough to present no difficulties for the pilot.

The government schools had either a set of PCs, some running Windows and some Linux, or a Linux server machine with several thin clients. In both cases, only one PC would have been capable of running the reading tutor immediately. They also were vulnerable to power cuts, without significant battery backup. A significant upgrade would be required for either of these schools to run the reading tutor for a class of students.

The low-resource private schools had more machines, but S2, which participated in the pilot, had to upgrade the memory on their machines and purchase headsets to accommodate the needs of the reading tutor program. They also had to work around power cuts, as they lost power for 1-2 hours a day, with no warning and no schedule, and did not have UPS backup. S6 had a larger number of PCs but would also have to do some upgrades. They were anticipating getting an internet connection in the near future.

The sponsored private schools had the largest number of machines— as reported earlier, S3 was the only pilot participant that was ready to go with no additional updating, with 16 PCs in their junior lab ready to go. They also had the only general internet connection, though it was somewhat unstable and could break down for days at a time. The power for the PCs was reliable, provided by a generator.

S4 had a large number of PCs, but they were also older Pentium 3s, which could not have supported the program. Because of their rural location, they also faced the longest power cuts, with power only available for a few hours a day. The cost of supplementing that shortfall with a diesel generator was prohibitive. This meant that the students requiring computer instruction (6^{th}-10^{th} standard) would adjust their schedule to use the computers when power was available. Any additional program would be able to operate only when the curricular constraints were satisfied.

One aspect of the infrastructure that is not often emphasized is the instability of the typical computer in Indian schools. Antivirus software is considered something of a luxury, and is rarely installed. Even when present, the absence of internet connections in most schools means that definitions are rarely up-to-date. The systems also tend to accumulate free and poorly tested software. This contributed to the instability of the reading tutor on several systems, for example when the audio drivers stopped working.

In terms of access to computing resources outside of school, only the fee-based private schools reported significant numbers of families with home PCs, 25% at S2 and 40% at S6. The use of cyber cafes varied considerably; some reported frequent use

> "Many times we give them assignment and they go to the internet [at cyber cafes] for information,"

while other programs reported little use, or expressed concern about younger students going there:

> "Parents might be less willing to send younger kids, especially girls, to cyber cafes"

4.4 Attitudes towards Technology Interventions

The perceived value of the reading tutor program was driven largely by priorities only partially related to English reading fluency. Assuming that a tangible benefit to using the program could be shown, the heads of schools were also interested in fostering an enthusiasm for reading and learning in English—making the process fun. As one said, "stories are the best way to teach anything."

Introducing any new component in the curriculum represents additional overhead for the teachers. We anticipated more resistance on this score, but perhaps because the reading tutor is designed to be largely self-directed, this did not come up very much. The teachers in general do not have a great deal of computer skills, and tend to rely upon a resource person, sometimes off site, to take care of any problems. Problems raised tended to focus on time, hardware infrastructure, or personnel simply for monitoring the children while they use the software.

Some of the principals were ready to counter a claim that the software could replace a teacher. Though we made no such claim, it would seem that they are alert to overblown claims made by previous programs or proposals. In general, all the schools were open to using such software as an enhancement to classroom teaching. One advantage is consistency of instruction:

> "I would use (technology) to bridge the gap between the really good teachers and those who are not so good"

Limited computing resources could be used to focus on remedial support for children that are falling behind.

> "[We could] use it for the kids who need remedial support in reading fluency in grades 3-5."

When we asked how the schools would measure success, all the schools agreed the ultimate test would be whether the students did better on classroom exams and (more subjectively) on the student's command of and confidence in read or spoken English in the classroom. Some indicated that they would seek positive feedback from the parents as well on the child's improvement in English language skills. Without such measurable progress, it is doubtful that a technological intervention would long survive the departure of the research team that introduced it. The specific measures we use in our pilot study to determine progress—spelling and reading rate improvements—are certainly germane, but may not be enough. Though all the principals agreed that reading fluency was the first step towards better comprehension, in an ESL environment in particular, the teachers wanted a way of training and testing comprehension of the passages read

> "It really does not matter whether they have learnt ten new words or twenty, but if their reading fluency improves and if there is some way we can test their understanding of

what they have read, I would say it is a success."

5. PILOT STUDY

The pilot study focused on the qualitative and quantitative enhancement of reading skills through the implementation of a reading program using the Reading Tutor.

We hypothesize that all students will show improvement in their reading skills over the period of the pilot, due to regular classroom instruction and the use of the Reading Tutor. For the intervention to be considered successful, the students supplementing their classroom learning through the use of the Reading Tutor should show larger, quantifiable gains in reading skills that persist after the student stops using the program. However, we expect these gains will not be equal across all the students as they differ significantly in ages and initial skills; it may be that the Reading Tutor is best suited to certain subsets of students.

We were also interested in the ways the Reading Tutor may need modification to better integrate into the Indian cultural context. Specifically, we wished to assess the impact of the unfamiliar American accent of the narration on the children using the program, and to understand the importance, if any, of introducing story content from India. We accordingly added some content to the program to provide a contrast in both accent and milieu, using an Indian English speaker to narrate several children's stories with Indian themes and characters, from an Indian publisher, Pratham [12]. We then conducted post- pilot interviews of a sample of the children to establish the range of responses to these qualitative questions.

5.1 Participating Programs

None of the students in any of the three programs had exposure to English outside of the school environment. In several cases, they were the first generation literates in their family in any language.

We chose this range of student ages and programs to get a broad view of where the reading tutor could be effective in the Indian context. Table 1 summarizes the student population participating in the pilot.

Table 1. Distribution of participants in pilot.

School	Std	Group 1	Group 2
S1	9	3f/1m	1f/4m
S2	4	6f/5m	4f/7m
S3	3	9f/6m	6f/9m

5.2 Protocol

The study ran for roughly 2½ months, the length varying slightly by site. We divided each set of participants into two groups, trying to make each group equal in terms of existing English skills. Because skills varied widely within each site, this could only be an approximate match.

We then started one group of students with the reading tutor, using it daily for 30 minutes. The second group at each site had no additional intervention. After half of the available time had elapsed for the study, the groups would switch roles—the first would cease all activity with the reading tutor, and the second would take its place. As the reading tutor was used outside of regular class time, it was always in addition to the work that the children did normally in English class.

At the beginning, middle, and end of the pilot, all students underwent performance testing to quantify the improvements they made in English skills with and without the reading tutor.

5.3 Performance Tests

We used two test regimes to quantify progress of the students after using the software, following the approach used in previous Reading Tutor studies.

The first is a test of reading fluency, following the rubric of Curriculum-Based Measurement (CBM) [13]. We selected a story at each student's reading level that they had not seen before. Each child read from the passage for a minute in a private room, and the score was simply the number of words read correctly. "Correctly" meant without pronunciation errors or false starts, though inevitably there were grey areas. The CBM is a simple measure designed for ease of reliable use in varying school environments, and as such serves our needs better than more sophisticated testing regimes. However, it has been shown to correlate well with such measures of reading fluency.

The second test was of spelling, and relied upon the commercially developed Test of Written Spelling [14]. This involves presenting a series of spelling words of increasing difficulty to each child; the score is the number of words correct before five consecutive errors. This test was administered collectively, to groups of between 10 and 30 children.

5.4 Running the Pilot

The installation and startup went reasonably smoothly for all sites, as the students were able to acclimatize to the reading tutor quickly and with little instruction.

The enthusiasm of the support staff, though not directly related to the success of the pilot, did affect the smooth running of the program. Motivated support staff took a personal interest in the students' progress, kindling more enthusiasm in many students, and helping to ensure adherence to the daily use protocols. They also seemed to identify problems with the program or setup earlier, regardless of their own level of troubleshooting abilities, leading to prompt intervention.

There were some problems with the stability of the tutor, due to some bugs in the program itself and the inherent instability of the computing infrastructure and support in the schools. Antivirus software was often missing or out of date, and the systems tended to accumulate free and poorly tested software. Further, none of the schools had reliable access to trained IT support. These issues contributed to the instability of the reading tutor on several systems, particularly those used for multiple purposes.

Stability problems affected S3 the most, probably because they had the largest number of PCs, the highest variety of users (there were many students and staff using these computers besides the pilot participants), and the most complex setup—running the reading tutor in client/server mode over an intranet.

6. RESULTS

The pre-test results for the fluency and spelling tests are summarized in Fig. 1. The differences between the programs are much smaller than the variation between students in the same program. The scores for S1 are close to those of the other schools,

reflecting that these children start learning English as a second language several grades later in their Kannada medium government school, than the two English medium schools. Note that the raw scores differ in both tests by an order of magnitude between the strongest and weakest students. The improvements in mean scores on the fluency test in the first and second half of the pilot are collected in Table 2.

Both test and control groups received the same English instruction in class that was otherwise conducted at their respective programs. Group 1 and 2 are separately shown (except for S1 where there aren't enough statistics), and the combined results for each site are displayed above the component scores. The final column indicates the significance of any positive effect of using the reading tutor, given as the *p*-value of the corresponding pairwise t-test. We used a pairwise comparison, comparing each student to his or her own performance, rather than contrasting the scores from groups 1 and 2. This was because of the wide spread in starting values across the students. Two S2 students were excluded from these calculations; we shall return to their results in section 8.

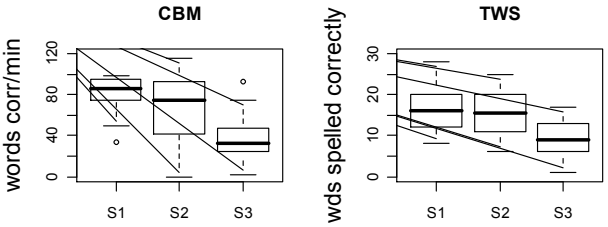

Figure 1: Pre-test scores on fluency and spelling tests by program.

The S2 students showed the clearest gains in fluency with the Reading Tutor; the S1 site's results were consistent with S2's, but their significance was not as strong. The combined fluency results from the two sites show a p-value of 0.0048. The combined S3 results do not show an advantage with the use of the Reading Tutor. Closer examination shows that group 1's pairwise t-test results show a positive result with a p-value of 0.016; the group 2 results are the ones showing less fluency gains with the Reading Tutor than without. We will return to these results later. However, if S3's fluency results are simply added to the results from the other sites, the p-value is still significant, 0.024.

Table 2. Fluency (CBM) test results. Change in fluency is in correct words in 1 minute.

School	Grp	Δ fluency with RT	Δ fluency no RT	p-value
S1	comb	9.3 (8.0)	4.7(9.5)	0.150
S2	1	12.4 (6.5)	4.7 (4.2)	0.028
	2	14.1 (7.2)	8.3 (8.7)	0.224
	comb	13.3 (6.7)	6.5 (6.9)	0.018
S3	1	12.9 (9.6)	1.2 (13.6)	0.016
	2	7.9 (10.0)	13.4 (13.2)	0.218
	comb	10.4 (9.9)	7.3 (14.6)	0.366

Table 3 summarizes the results of the spelling tests for the three sites, showing the average improvement when using the Reading Tutor vs regular school training. Results by group are not shown because we did not find any meaningful difference in spelling performance in any group or site. This may be simply because of the brevity of the pilot; a longer study may have led to a stronger effect.

Table 3. Spelling (TWS) test results. Only combined results are shown.

School	Δ TWS with RT	Δ TWS no RT	p-value
S1	1.22 (2.2)	0.56 (3.3)	0.71
S2	0.05 (2.6)	1.10 (3.3)	0.36
S3	0.40 (2.2)	0.93 (3.6)	0.60

7. STUDENT INTERVIEWS

7.1 CBM Fluency Tests

The CBM test format is intended to be a coarse measure of fluency, and thus has some limitations. The most obvious is that the passages need to be chosen ahead of time, and occasionally students will not be able to cope with the simplest of the passages selected. There were several very low scores on fluency, but only one student was unable to read a single word from the text. This student was excluded from the results in Tables 2 and 3. However, after using the Reading Tutor for a month, he actually improved to 11 words correct on the simplest passage.

A more subtle limitation is the CBM test's vulnerability to a student changing strategy. The score is simply the number of words read correctly, i.e. total words read – number of mistakes. This means that students that just skip multisyllabic words will get much farther in the text. Since most students read all the short words ("the", "it", "said") correctly, this strategy will significantly inflate the student's fluency score.

We can now examine the other student excluded from the fluency results in Table 2. This student used the Reading Tutor in the first half of the pilot. At pre-test, she used the word-skipping strategy, simply reading the first syllable of any word she didn't know. After using the Reading Tutor for 4½ weeks, she read 23 *less* words than in her pre-test, because she was now trying to work through every word. Unfortunately, she switched back to the original strategy when she took the final test, and her score went back up. We excluded her because she is actually a success of the Reading Tutor, but the fluency test scores give exactly the opposite conclusion.

Figure 2 shows the number of words read total vs the number of mistakes for each child in S2 group 1, for the pre- and mid-tests. The expectation is that each child reads more words and makes a smaller percentage of errors, like the student marked C. The child that markedly changed her strategy is marked with an A. Most children do not show such a clear shift, but student B may show a smaller shift the other way. A potential refinement of the CBM test would be to use a quantitative measure like this to identify changes in student strategy, and to handle them differently.

7.2 General Observations

Our overall conclusion is that the Reading Tutor was successful at improving reading fluency, even over the relatively short duration of our pilot. As noted above, the strongest positive result is that

for the S2 school students, while the students of S3 showed no measurable aggregate gain, as group 2 had *less* improvement with the reading tutor than without. S1's results lacked sufficient statistics to make strong conclusions, though they were consistent with S2's.

There are a variety of reasons that could be behind the S3 fluency test results—clearly the reading tutor did not reach the second group of students. One possibility is that it was too difficult for this cohort. The S3 students were the youngest (3rd std) in our study, and were selected because they were in need of some remedial help. Despite our intention to balance the English skills of the groups, the second group in S3 had more difficulty than the first in advancing from letter-to-sound training to actual story reading. The second group also experienced the longest disruption (lasting about 5 days) where the program was not working properly for any of the students, which could have discouraged them. A longer study, earlier in the school year, would separate the reading tutor's effects from such confounds.

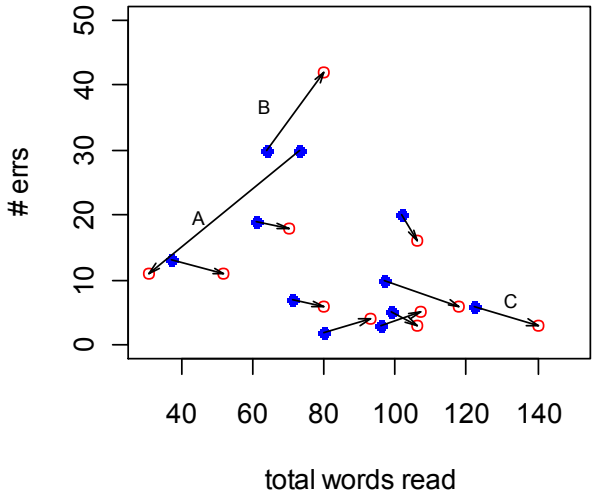

Figure 2: Total number of words read vs. number of mistakes on pre-pilot (dots) and mid-pilot (circles) fluency tests for S2 students using the Reading Tutor. Students A and B shift strategy between tests, C is the "normal" case.

We set the goal for the reading tutor to reach the largest population possible for enhanced English literacy. No matter which school we considered, the goal was acknowledged to be of pressing interest. Thus the question becomes practical, what would be required for the reading tutor to be successfully deployed in each school system in our taxonomy?

The government school system may be the most difficult place to introduce a computer-based literacy intervention. The time, personnel, and computing resources available are all highly constrained. In addition, the administrative process is the most involved, due to the need for state school board approval. The state board does have a very wide reach, however; if the technical and personnel problems could be solved, an intervention could be implemented on a large scale.

The informal programs have the greatest curricular flexibility, though they tend to serve the same populations as the government schools. They have the least time with their students, so they would need to be convinced that the technology was a net positive in terms of use of instruction time. They can also face the most stringent computing infrastructure limitations, as they are dependent on donations for equipment. However, if they see a real advantage to a technology they can be resourceful in finding the necessary components, and are "doers", willing to invest the time to understand how to do basic troubleshooting. They are unlikely to have on-site general computing expertise.

The low-resource private schools have some flexibility in scheduling for teachers and students. The schools we interviewed also have some computing infrastructure to work with, though not many machines compared to student population, nor are the machines state-of-the-art. Administratively, introducing the program required only approval of the principal (who also approved the upgrade of the RAM to make their PCs compatible with the reading tutor). The monitoring of the children while they used the program was done by the teacher also responsible for computing classes.

Finally, the sponsored private schools are easiest to work with. Both the schools surveyed have personnel that can supervise the children in the computer lab, and could make time in the day for them to use the program. The computing infrastructure is built by donations, and is the most modern of the schools we interviewed (S3) or in the process of being upgraded (S4). The administrators have autonomy on implementation of new programs. The only limitation is that there are few such institutions.

8. DISCUSSION

We have shown a positive benefit to using the reading tutor for children from Kannada and English medium schools from 4-6 weeks' use. The benefit was not uniform across programs, and a longer study would help to clarify the effects of factors like initial English reading skills or age on fluency improvements. Likewise, more data is required to establish a benefit on spelling skills with use of the reading tutor. We can already say that the reading tutor is a promising intervention for English education in the Indian context, for English and vernacular medium schools.

We believe that PC-based English literacy software could make a significant difference in the lives of many children in Indian urban schools who have very little exposure to English other than formal classroom instruction. Many of the schools are willing to introduce such software on top of their curriculum if they see it bring value in terms of better performance. In particular, both S1 and S2 expressed a desire to continue using the reading tutor in the following school year, thus voting with their feet. However, for this to be a success a number of technical constraints need to be worked around, in terms of footprint, OS flexibility, robustness and maximal use of PCs by the students/day. Besides these, adding the capacity to train, track, and test for comprehension of the passages is important in the Indian context.

The ultimate goals for such an intervention can only remain modest. Success from the technologist's standpoint is when we have significantly and enjoyably enhanced English fluency, spelling, and comprehension, for a large population of otherwise underserved students. .

From the student's perspective, the aspirations are higher. One graduate was recently asked by his former principal how he landed a job for which he had limited experience. His reply was simple: "My English got me this job".

9. ACKNOWLEDGMENTS

We would like to thank the principals, teachers, and students of all the schools for their support and participation. Jack Mostow and the Project Listen team provided essential support and valuable advice throughout the work. FW thanks MSR India for their support during this work.

10. REFERENCES

[1] Sheorey, Ravi, Learning and Teaching English in India, Sage Publications, 2006

[2] Agnihotri, R., and Khanna, A., English Language Teaching in India; Issues and Innovations, Sage Publications, 1995.

[3] Ramanathan, H., "Testing of English in India: A developing concept," in Language Testing, 25(1). 2008. 111-126.

[4] Mostow, J. and G. Aist. "Evaluating tutors that listen: An overview of Project LISTEN." In K. Forbus and P. Feltovich, Editors, Smart Machines in Education, 169-234. MIT/AAAI Press: Menlo Park, CA, 2001.

[5] Mostow, J., et *al.,* "Evaluation of an automated Reading Tutor that listens: Comparison to human tutoring and classroom instruction", Journal of Educational Computing Research, *29*(1), 61-117.

[6] He, F., Linden, L., and Macleod, M., "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program", Working Paper Columbia University Department of Economics. 2008.

[7] Bannerjee, A., Cole, Duflo, S. E., and L. Linden, "Remedying Education: Evidence from Two Randomized Experiments in India", Quarterly Journal of Economics, 122(3).2007. 1235-1264.

[8] Mills-Tettey, A., et al., "Improving Child Literacy in Africa: Experiments with an Automated Reading Tutor." 3rd ICTD Conference, (2009) Doha, Qatar

[9] Reeder, K., Shapiro, J., and Wakefield, J. "The effectiveness of speech recognition technology in promoting reading proficiency and attitudes for Canadian immigrant children," 15th European Conference on Reading, (2007) Humboldt University, Berlin.

[10] Kumar, A. et al.,, "An Exploratory Study of Unsupervised Mobile Learning in Rural India," CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

[11] CMU Sphinx: http://cmusphinx.sourceforge.net/

[12] Pratham Books. http://prathambooks.org/

[13] Deno, S. L. "Curriculum-based measurement: The emerging alternative," in Exceptional Children, 52(3):219–232, 1985.

[14] Larsen, S., Hammill,D., and Moats, L. Test of Written Spelling—Fourth Edition, Pro-Ed, Austin, Texas, 2009.