

**Tutoring Bilingual Students
With an Automated Reading Tutor That Listens:
Results of a Two-Month Pilot Study**

Robert Poulsen

CSC 696 Masters Thesis

DePaul University

June 16, 2004

Abstract

A two-month pilot study comprised of 34 second through fourth grade Hispanic students from four bilingual education classrooms was conducted to compare the efficacy of the 2004 version of the Project Listen Reading Tutor against the standard practice of sustained silent reading (SSR). The Reading Tutor uses automated speech recognition to “listen” to children read aloud. It provides both spoken and graphical feedback in order to assist the children with the oral reading task. Prior research with this software has demonstrated its efficacy within populations of native English speakers. This study was undertaken to obtain some initial indication as to whether the tutor would also be effective within a population of English language learners.

The study employed a crossover design where each participant spent one month in each of the treatment conditions. The experimental treatment consisted of 25 minutes per day using the Reading Tutor within a small pullout lab setting. Control treatment consisted of the students who remained in the classroom where they participated in established reading instruction activities. Dependent variables consisted of the school districts curriculum based measures for fluency, sight word recognition and comprehension.

The Reading Tutor group out gained the control group in every measure during both halves of the crossover experiment. Within subject results from a paired T-Test indicate these gains were significant for two fluency measures ($p < .001$) and one sight word measure ($p = .056$). Effect sizes were 0.55 for timed sight words, a robust 1.16 for total fluency and an even larger 1.27 for fluency controlled for word accuracy. That such dramatic effects were observed during a one-month treatment, suggests this type of technology may have much to offer English language learners.

Introduction and Motivation

Of the many challenges facing public schools today, one clear area of concern is how to meet the growing demand to educate our countries rising population of limited English proficient (LEP) students. Slavin & Cheung (2003) identify reading instruction for English language learners¹ (ELL) as “one of the most important issues in all of educational policy and practice”. A surprising 20% of all U.S. students come from homes where English is not the primary language spoken (Van Hook & Fix, 2000). This population continues to grow at an overwhelming pace. From 1991-92 through 2001-02

¹ Throughout the body of research, the terms limited English Proficient and English language learner are used somewhat interchangeably. Following the convention of (August & Hakuta, 1997) I will adopt the term English language learner or ELL whenever possible as it carries a more positive tone.

ELL enrollments rose 95% compared to a 12% growth in total K-12 enrollment. This signifies an ELL growth rate of nearly eight times that of the general student body (Padolsky, 2002). While this population represents a wide range of language groups, roughly $\frac{3}{4}$ of all English language learners come from Spanish speaking homes (Moss & Puma, 1995; Ruiz de Velasco & Fix, 2000; Zehler *et al.*, 2003).

Not only is this population growing, but also their achievement levels continue to lag far behind their native English-speaking peers. Third grade ELL students rank in the 30th percentile for reading, with 16% of this group receiving a grade of unsatisfactory in reading compared to only 7% of native English speakers (Moss & Puma, 1995). District coordinators report that 76% of third grade English language learners were either below or well below grade level in reading (Zehler *et al.*, 2003). The National Center for Education Statistics find that a mere 7% of LEP fourth graders were at or above the Proficient level and only 28% reached the Basic level for reading achievement within the 9 major urban school districts sampled as part of their Trial Urban District Assessment (NCES, 2003).

Research overwhelmingly indicates that current educational practices are not meeting the needs of this population. Guerrero & Sloan (2001) cite a large body of research indicating lower achievement levels for minority-language children (predominantly Spanish speaking) and identify this group as having increased risk of poor literacy in both their native and second languages (see, Arias, 1986; Congressional Budget Office, 1987; De La Rosa & Maw, 1990; Durgunoglu, 1998; Haycock & Navarro, 1988; Kao & Tienda, 1995; Orfield, 1986; Verhoeven & Aarts, 1998).

Under the present political environment, this issue is becoming even more critical. The No Child Left Behind Act of 2001 has placed demands on school districts to meet Adequate Yearly Progress goals for all subgroups including English language learners. In fact, Title III of this legislation (Language Instruction For Limited English Proficient And Immigrant Students) is devoted specifically to the need to raise the achievement levels of our nations English language learners. Those schools that fail to meet AYP goals for this or any other subgroup over three consecutive years will face a myriad of sanctions (NCLB, 2001). Clearly, helping English language learners close the gap on their native English speaking peers will need to be a priority if our schools are to meet the challenges set out by this legislation. While NCLB places strong demands on performance and accountability, it does not legislate the methods that schools must use.

In terms of methodology, a great deal of research has centered on the language of instruction for English language learners. This research can be divided into two broad categories; one supporting the use of native language instruction initially and then transitioning students to English and the other supporting instruction based in English-only immersion. Although there are extensive findings on both sides of this paradigm (see August & Hakuta, 1997; Chu-Chang, 1981; Seder, 1998; Slavin & Cheung, 2004; Slavin & Cheung, 2003; Thomas & Collier, 2001), one conclusion that seems common among research reviewers is that instructional practices may in fact have greater bearing on achievement than the language of instruction (August & Hakuta, 1997; Slavin & Cheung, 2003). August (2003) specifically identifies “a desperate need for more theoretically-driven research that employs quasi-experimental designs and high quality

assessments to examine the effectiveness of instructional practices designed to bolster the literacy of English language learners”.

Technologically based reading interventions are specifically cited as an area where future research is needed. August (2003) identifies the use of technology to support ELL literacy education as one of two areas particularly worthy of research efforts, noting its ability to both teach and assess component literacy skills. The National Reading Panel identifies the use of speech recognition technology in reading instruction as an area in need of further research (NRP, 2000). The use of technology and electronic texts has been observed as an important component in K-8 grade ESL classrooms (Meskill, Mossop & Bates, 1999). Finally, ESL teachers report that their students are both highly motivated by the use of computers and that they perceive higher social status with the mastery of computer skills (Meskill & Mossop, 2000, cited in August, 2000).

In response to this driving need for research to identify better tools and methods to help English language learners, this research seeks to provide some initial findings on the efficacy of one particular computer-based tool for reading instruction, the Project Listen Reading Tutor.

Background

The Project Listen Reading Tutor has been an ongoing area of research at Carnegie Mellon University since 1990. Its development has been research based and has centered on modeling expert teachers (Mostow & Aist, 2001; Mostow *et al.*, 2003a). Since its inception, the research team at the University has compiled an extensive body of research indicating the technology to be an effective tool for literacy instruction within various populations of native English speaking children (Aist, 2002; Aist *et al.*, 2001;

Aist & Mostow, 1997; Mostow *et al.*, 2003a; Mostow & Aist, 2001; Mostow & Beck, 2003). While beyond the scope of this paper, a complete description of the research basis and findings of the Project Listen team can be found at <http://www-2.cs.cmu.edu/~listen/research.html>.

The following briefly summarizes some of the major findings that relate to this paper.

- A 1994 within subject experiment of 12 low-reading second graders on a prototype Reading Coach showed that children were able to read and comprehend material six months more advanced with the coach than they could independently (Mostow, Roth, Hauptmann & Kane, 1994).
- In a second 1994 within subject experiment of 34 second-graders, subjects averaged 20% higher comprehension scores on a third grade reading passage with the assistance of the Reading Coach than they did without it (Mostow & Aist, 2001).
- A 1996-97 pilot study of six bottom performing third grade children who pre-tested at approximately 3 years below grade level identified an average two-year gain in reading level pre to post test while using the Reading Tutor during the eight-month study as measured by school administered reading inventories (Aist & Mostow, 1997).
- A 1998 within classroom controlled study of 72 second, fourth and fifth graders compared the Reading Tutor to regular instruction and commercial reading software over a four-month study. The Reading Tutor group significantly outgained the regular instruction control group in Passage Comprehension as measured by the Woodcock Reading Mastery Test (WRMT). No significant differences were recorded between groups for Word Attack, Word Identification or oral reading fluency (Mostow *et al.*, 2003b).
- A 1999-2000 between classrooms controlled study of 131 second and third graders from 12 classrooms compared daily usage of the Reading Tutor to daily human tutoring by a certified teacher and to a regular instruction control group within each classroom. Results from the yearlong study indicate that children assigned to both the human-tutor and Reading Tutor conditions significantly outgained control in word comprehension and suggestively in passage comprehension. The human tutored group significantly outgained the Reading Tutor group in Word Attack only. No significant difference in gains for Fluency and Word Identification were recorded (Mostow *et al.*, 2003a).
- A 2000-2001 study of 178 children from grades 1 through 4 at two schools compared 20-minute daily treatments of the Reading Tutor to the standard practice of 20 minutes sustained silent reading over a 7-month study. Reading Tutor group significantly outgained statistically matched SSR group in word identification, word comprehension, passage comprehension, fluency, phonemic awareness, rapid letter naming and spelling using normed WRMT. Most of the significant gains were observed in grade 1.

- A 2002 pilot study of 35 Canadian ESL students ranging from first to sixth grade investigated the usability of the Reading Tutor for English language learners. Participants represented three different native languages: Tamil, Mandarin and Cantonese. Results indicated that roughly 86% of participants were able to effectively interact with the Reading Tutor. However, questions were raised as to whether the ESL students would be able to benefit from the Reading Tutor in its current form (Li, 2002).

The primary goal of this research was then to build on to the results presented by Li (2002) and identify if English language learners would demonstrate measurable gains in reading skills as a result of reading instruction that includes regular use of the Project Listen Reading Tutor.

Reading Tutor Description

The Project Listen name is based on the acronym “Literacy Innovation that Speech Technology ENables”. Central to the pedagogy of this tutor is its implementation of the Sphinx II speech recognition engine. This technology enables the Reading Tutor to analyze children’s oral reading, track their place within the context of a story and provide feedback to children both preemptively and in response to difficulties they encounter during the oral reading task (Mostow & Aist, 2001). The software is implemented on standard Windows computers and utilizes inexpensive headphones with a noise-canceling microphone.

Reading Activities

All Reading Tutor sessions begin with the student logging in to the system by selecting their name and birth month from talking menus. When students log in for the first time, they are presented with an initial reading activity that also serves as a basic tutorial. Students are walked through the simple controls for the tutor via a story

featuring a mouse named Kyle, who in the context of the story is also learning to use the Reading Tutor. This tutorial focuses on learning navigation controls, understanding when the student is expected to read aloud and how to get help from the tutor on difficult words. Two other “tutorial stories” are presented to the students at later times, one on how to use the keyboard and the other on how to write and narrate a story within the Reading Tutor environment.

After completing the initial tutorial, students begin taking turns with the Reading Tutor to select the next story to read. This alternating choice approach was first implemented in the 1999 version of the Reading Tutor in order to address a pattern where students were repeatedly selecting the same easy stories (Aist, 2000; Aist & Mostow, 2000; Mostow *et al.*, 2003a). When it is the student’s turn to pick a story, the tutor suggests an appropriate level and the student is free to choose any story at that level or select a story from any other level. The number of times a student has read a particular story is displayed alongside each story title in the menu. When it is the tutor’s turn, a previously unread story is selected at the students current recommended reading level. The recommended reading level is continuously assessed and adjusted by the tutor based on the student’s oral reading rate (Aist 2000; Mostow & Aist, 2001; Mostow *et al.*, 2003a; Jia, Beck & Mostow, 2002).

Figure 1 shows a typical screen during the assisted oral reading task. Sentences (or phrases) are displayed one at a time for the student to read. Words that have been accepted by the tutor are highlighted as the student reads them. The Reading Tutor provides assistance whenever it detects a long pause, a skipped word, a seriously misread word or preemptively for difficult words (Mostow & Aist, 2001; Mostow *et al.*, 2003a).

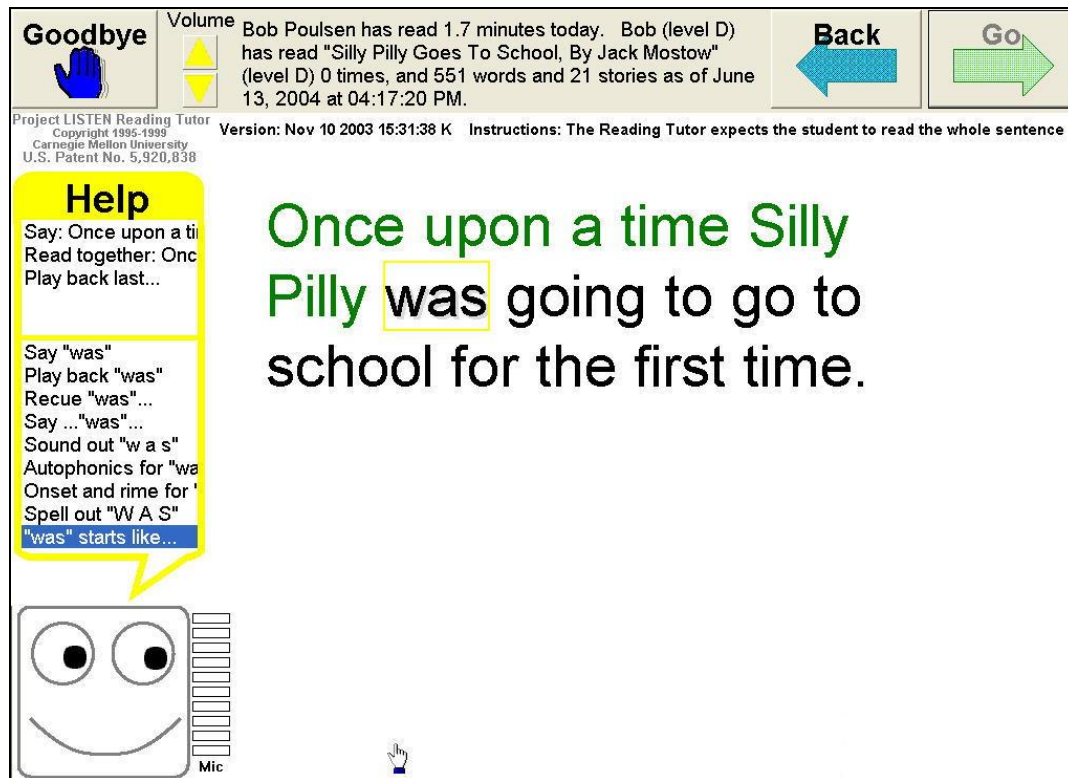


Figure 1: Screen capture from Project Listen

Students can also request help from the tutor by simply clicking on a word. The Reading Tutor provides assistance in one of the following forms:

- speaks the whole word aloud
- re-cues the word by rereading the sentence leading up to (but not including) the word
- decomposes the word into syllables, onset and rhyme, or phonemes (speaking each component while highlighting it)
- displays and reads a different word with the same onset or rhyme (ie. for “lump” it might display “jump” while saying “rhymes with jump”)
- display a picture (ie. display an apple for the word apple)
- play a sound effect (ie a roar for the word lion)

The last two interventions are only available for a small set of words and are therefore rarely used. When more than one of these interventions is appropriate, the tutor chooses randomly between them (Aist, 2002; Mostow *et al.*, 2003a).

To support comprehension, the Reading Tutor also provides whole-sentence help in several forms. It either reads the sentence fluently or it reads the sentence one word at a time while asking the student to read along. Words in the sentence are highlighted as the tutor reads them. This type of support is provided whenever the student requests it (by clicking on the screen below the sentence), when the student has difficulties on multiple content words, when the tutor detects long pauses between words or preemptively for sentences containing difficult words (Aist, 2002; Mostow *et al.*, 2003a). For most words and sentences, Reading Tutor assistance uses recorded human voices as opposed to computer-synthesized speech.

Experimental Design

This study was undertaken to present some initial indication as to the efficacy of the Reading Tutor compared to a control condition consisting primarily of sustained silent reading within a population of English language learners. The participants in the study were recruited from a Chicago area suburban school with a significant population of Hispanic bilingual students (36.4%). The school enrollment of 586 students is made up of roughly 35% white, 4% African American, 51% Hispanic and 10% Asian / Pacific Islander. 36% of the school's student body is designated as low-income (ISBE, 2003). All students enrolled in the bilingual program in grades two, three and four at the school were invited to participate via a consent agreement sent home with the children. The

consent form was distributed in Spanish as many of the children's parents have a very limited capacity to read in English.

The consent agreement indicated that all students who wished to participate would need to stay after school for one hour Monday through Thursday. This was because the school district had initially agreed to only allow the study outside of regular classroom hours. This requirement may have reduced the number of children who volunteered to participate, although no measure of this was collected. After the completion of pre-testing and assignment to groups, a busing issue caused the after school aspect of the treatment to be cancelled and the study treatments were rescheduled to occur Monday through Friday during the classrooms daily scheduled silent reading time.

It should be disclosed that the principle investigator had an established connection with the school site where this research was conducted having worked at the school as a reading tutor for both bilingual and mainstream children and with additional responsibilities involving PE class and playground supervision. As such, he also had an established rapport with the majority of the study participants.

Subjects

All students who replied to the consent letter were admitted into the study with the following exception. Nine students who signed up were also enrolled in a peer-tutoring program that was being offered concurrently at the school. Because of scheduling conflicts between the two programs, it was decided that those children enrolled in peer tutoring would not be eligible for this study. We also intended to screen

out those children whose English language skills were felt to be too limited to operate the Reading Tutor. However, based on the results of (Li, 2002) identifying very low English proficient students as still being able to operate the Reading Tutor and discussions with the teachers, it was decided that all of the remaining 34 applicants met this liberal standard for inclusion. It should be noted that students whose self-assessed English reading proficiency was extremely low may have “screened themselves out” by simply not volunteering. Participants came from four classrooms: one second-grade, one third-grade, one 3-4 multiage classroom and one 4-5 multiage classroom. Table 1 shows the distribution of participants by grade level and gender.

N	2nd grade	3rd grade	4th grade	total
male	3	6	8	17
female	8	4	5	17
total	11	10	13	34

Table 1: Participant Distribution by Grade Level and Gender

Treatment Model and Group Assignments

A crossover experiment was identified as the model for the study. Under this model students were randomly assigned to two groups, stratified by reading fluency pretest score (total words read) and by grade level. During the first month of the study, group one was assigned to the Reading Tutor intervention while group 2 was assigned to the control condition. During the second month, group treatments were reversed. Primary motivation for this design was to allow all volunteers the opportunity to participate in the reading tutor treatment condition for some amount of time. This was

considered important because the majority of the respondents who enrolled in the study did so specifically because of a desire to use the computer-based reading tutor.

Therefore, the design was established to meet this equal opportunity criterion.

Interventions

Reading Tutor interventions took place in a small pullout lab staffed by the principle investigator and equipped with ten PC's each running the 2004 version of the Project Listen Reading Tutor. Students attended the lab for 25 minutes each day over the four-week treatment. During each session, students independently worked on oral reading and writing tasks with the computer-based tutor as described in the background section of this paper. All reading tutor interventions were provided in English language only.

Control interventions took place in the regular classroom during the time that Reading Tutor treatment was provided in the pull out lab. Students who remained in the classroom included both students not enrolled in the study and study participants assigned to the control condition. During this time they participated in what the school defines as D.E.A.R. time (Drop Everything And Read). This time is somewhat unstructured, but is primarily designated as a time students spend engaged in Sustained Silent Reading (SSR) of self-selected materials. All four teachers reported 100% of this time was spent in some form of reading instruction and confirmed that SSR was the primary activity. Additional group interventions reported by teachers included some read alouds (teachers reading a story to the class) and partner reading (two students taking turns reading to each other).

Two teachers also reported that some writing activities took place in the form of time spent writing in their daily journals.

The second and third grade teachers reported working individually with approximately 1 to 3 students per day in guided reading activities during this time. The second grade teacher indicated this individual instruction to include word recognition, phonics and decoding skills, “mainly in Spanish, but some English”. The third grade teacher indicated this time was spent individually reading with each student. She further indicated that this instruction took place in both Spanish and English depending on the individual needs of the student. No quantitative data about the exposure and specific nature of this individual attention provided to control subjects was available.

All teachers reported that both Spanish and English reading instruction was included during D.E.A.R. time. Students were free to pick materials in either language as part of their silent reading activity. The second grade teacher reported students were more likely to select Spanish material than English, the third grade teacher reported that the children were equally likely to read in Spanish or English and the fourth grade teachers reported that English materials were more commonly chosen.

Other Reading Instruction

The schools principal reports that between 120 and 150 minutes per day is devoted to direct reading instruction in the classroom. This figure was confirmed by feedback obtained from teachers and appears to have been consistent across all classrooms involved in the study. The language used for reading instruction is much more difficult to characterize. In general terms, the ratio of Spanish instruction to

English instruction was higher for the younger grades. Second grade instruction was reported to be primarily in Spanish with some English, third grade instruction was reported to be 50 / 50 and fourth grade instruction was reported to be primarily English. However, this was reported to be a broad characterization and all teachers reported that language of instruction was determined more by individual need than by this sweeping guideline.

Measures

Assessments used were selected reading components from the school districts Curriculum-Based Measures (CBM's). The particular measures selected constitute an informal reading inventory used throughout the school district and included measures for fluency, sight word recognition and comprehension. All reading passages selected were adapted from the basal reading curriculum used in the school. All assessments for a given grade level used the same reading passages and word lists.

The comprehension measure was made up of a single, group administered cloze test where students were asked to identify ten missing words from a reading passage at their current grade level. Second grade passages provided three multiple-choice options for each word. Third and fourth grade passages required students to fill in the blank.

The fluency measure included two subcomponents, one a measure of total words read in one minute (referred to as fluency: total words) and the other a measure of words read correctly in one minute (referred to as fluency: read correct). Reading passages were selected at one year below grade level. This selection is supported by Mirkin &

Deno (1979) who find that passage difficulties at the independent or instructional level are more sensitive to growth than passages at the frustration level.

The sight word identification measure also included two subcomponents, one a measure of the number of words identified automatically (words identified within a 2 second time limit) and the other a measure of total words decoded (words identified with no time limit). These measures are referred to throughout this paper as sight words timed and sight words untimed respectively. All students were tested on three levels of sight words ranging from two levels below grade level through their current grade level. Each level was made up of a set of 20 words (60 words in total) adapted from the Houghton Mifflin Informal Reading Inventory.

The CBM's were selected for several reasons. First, research has consistently supported the validity and reliability of these measures (Baker & Good, 1995; Deno, 2003; Fuchs & Deno, 1991). The characteristics of these measures make them particularly well suited to monitoring short-term progress compared to published norm-referenced tests that are geared toward measuring growth over long periods of time (Baker & Good, 1995). A primary use of the CBM identified by Deno (2003) is the formative evaluation model where regular fine-grained assessments are used to make judgments on the effectiveness of particular interventions. Finally, research has identified the reliability, validity and sensitivity of CBM assessments within populations of English language learners (Baker & Good, 1995; Deno, 2003). These characteristics combined with the nature of this short duration study made the established district CBM reading inventories a robust assessment choice.

Materials

Reading materials available to students during regular classroom instruction include both Spanish and English language materials. Teacher's characterization of this material indicated that it was very general in scope. They indicate that both fiction and non-fiction books were available of various levels and lengths. Teachers indicated that materials were separated by level and that students were individually directed toward material appropriate for their recommended reading level. Students were also allowed to use materials checked out from the schools library. The third grade teacher further indicated that materials were rotated and the selection would include some representation of a general theme being discussed in class. "Ocean books" was one of the themes mentioned during the course of this study.

The Reading Tutor contains hundreds of leveled stories for students to choose from representing a wide range of interests and styles. Materials range from introductory levels "K" (kindergarten) and "A" (1st Grade) through advanced material at level G (7th grade). Stories are drawn from a variety of sources including *Weekly Reader*, public domain web sources like www.gutenberg.net and stories written specifically for the Reading Tutor. Mostow *et al.* (2003a) presents a detailed description of the specific types of materials represented by each level.

Results

We now consider the results in an effort to address two questions. First we consider a within subject analysis to address the question: Did individual ESL students gain more in the month they used the Reading Tutor then in their month in the sustained

silent reading group? Second, more generally, we consider an analysis of covariance to identify what other variables may have significantly affected the observed gains.

Within Subject Analysis

Based on the crossover design, we first consider the within subject results based on a two tailed paired T-test. We analyze the gain for each student's score and compare the gain observed during control treatment versus Reading Tutor treatment. Mean gains for students during Reading Tutor group exceeded mean gains for the control group for every dependent measure. Results of the paired T-test indicate that these differences were significant for fluency: total read ($p < .001$), fluency: read correct ($p < .001$) and for timed sight words ($p = .056$). In order to present results comparable to other research on reading interventions, we follow the National Reading Panel standardized mean difference formula $(\text{treatMean} - \text{contMean}) / (0.5 * (\text{treatStdDev} + \text{contStdDev}))$ to calculate effect sizes (NRP, 2000). Effect sizes were robust for fluency at 1.16 for fluency: total read and 1.29 for fluency: read correct. Effect size for the timed sight word measure was also substantial at 0.58.

The sight word measure was subject to a ceiling effect with three students identifying all 60 sight words on the timed measure and seven students able to decode all 60 words for the untimed measure. It is not clear what impact this ceiling effect may or may not have had on the results reported here.

A summary of the data from the within subject analysis is presented in table 2.

Dependent Measure ¹ Treatment	Mean Gain	Mean SD	p ²	effect ³
Fluency: total read (n = 30) Reading Tutor SSR ⁴	17.1 4.1	11.1 11.3	0.000	1.16
Fluency: read correct (n=30) Reading Tutor SSR ⁴	17.7 4.6	9.8 10.5	0.000	1.29
Sight Word: timed (n = 30) Reading Tutor SSR ⁴	3.0 1.4	1.4 2.8	0.056	0.58
Sight Word: untimed (n = 30) Reading Tutor SSR ⁴	2.0 1.0	2.1 2.1	0.127	0.49
Comprehension (n = 25) Reading Tutor SSR ⁴	0.6 0.1	2.0 2.0	0.553	0.22

¹ Data based only on subjects who took all three tests for a given dependent measure (4 students missed post testing, 5 additional students missed crossover comprehension tests)

² Effect size based on NRP (2000) Standardized Mean Difference formula.

³ Based on a two tailed paired T-test

⁴ Sustained Silent Reading

Table 2: Within Subject Analysis

ANCOVA analysis

Second we consider an analysis of covariance to address what other variables may have influenced the observed gains. To facilitate this analysis, we first consider the data from both halves of the crossover experiment as a whole. We define a start and finish parameter for each dependent variable. We also define a treatment parameter for the overall dataset. For every student, we then added two rows to the dataset, one for their data during the control treatment and the other for their data during the Reading Tutor treatment. Start and finish parameters are set to their test scores at the beginning and end of the month for the treatment specified by the given row. Group number in this dataset

now represents the month that they participated in the Reading Tutor treatment and is used as a fixed variable to test for the significance of the order in which students received treatments.

We then performed an ANCOVA analysis for each dependent measure using finish score as the dependent variable and start score as a covariate. The model also included fixed variables for grade level, group and gender. The results again identified both fluency measures as significant ($p = .002$) in addition to sight word timed ($p = .050$). Surprisingly, this analysis also identified comprehension as being suggestive ($p = .089$) although with a minimal effect size of 0.19. Table 3 summarizes the general results from this dataset.

Dependent Measure Treatment	Pre Test	Post Test	Pre to Post Gain ¹		P ²	effect ³
	Mean	Mean	Mean	Std Dev		
Fluency: total read Reading Tutor SSR ⁴	86.7	103.5	16.8	10.7	0.002	1.16
	92.7	94.9	4.2	11.1		
Fluency: read correct Reading Tutor SSR ⁴	81.6	98.7	17.3	9.5	0.002	1.27
	87.0	89.8	4.7	10.3		
Sight Word: timed Reading Tutor SSR ⁴	50.1	53.1	3.0	2.6	0.050	0.55
	50.7	52.5	1.5	2.8		
Sight Word: untimed Reading Tutor SSR ⁴	53.6	55.5	1.9	2.3	0.236	0.40
	53.8	55.2	1.0	2.1		
Comprehension Reading Tutor SSR ⁴	6.2	6.6	0.5	1.9	0.089	0.19
	6.2	6.5	0.2	1.9		

¹ Pre to Post test mean gain is the average of each subjects gain, not the difference between pretest mean and post test mean.

² Based on ANCOVA analysis of treatment effect on post test score with pretest score as a covariate. Model also included treatment month, grade level and gender as additional sources of variance.

³ Effect size based on NRP (2000) Standardized Mean Difference formula.

⁴ Sustained Silent Reading

Table 3: Summary of two-month combined results over all data (N = 68)
Each student's results are included from both of their treatments

The ANCOVA analysis did not show any other individual factors besides treatment as significant for any measure. However, there were several significant interactions for sight word and comprehension measures. For sight words timed, two interactions were significant: treatment * gender ($p = .004$) and treatment * grade ($p = .004$). Treatment * grade was also significant for sight words untimed ($p = .008$). Comprehension scores showed several interactions as significant: treatment * gender ($p = .060$), gender * grade ($p = .040$) and treatment * group ($p = .054$). In addition, the three level interaction, treatment * grade * group, was significant for sight words timed ($p = .047$), sight words untimed ($p = .022$) and comprehension ($p = .000$).

The meaning of these significant interactions is not clear. They may simply be a result of over fitting the data. Then again, it is interesting to note that grade is a component in six of the nine interactions in addition to being somewhat suggestive ($p = .177$) for the sight words timed measure. Mostow et al. (2002) in their comparison of the Reading Tutor to SSR with 178 first through fourth grade native speaking students, found that students from different grades showed dramatically different gains during the seven-month study. The results from the ANCOVA here seem to suggest that additional statistical analysis is required. In particular, data should be considered separately within each grade to see if the efficacy of the tutor was consistent at different grade levels, or if the tutor was more efficacious for some grades over others.

Discussion

The purpose of this study was to investigate the efficacy of the Project Listen Reading Tutor within a population of English language learners. Mostow & Beck (2003) define efficacy as the gain achieved based on a specified amount of usage and further define effectiveness as the increase in gains produced by a particular intervention when compared to what the intervention replaces. The simple formula they present to relate these terms is $effectiveness = efficacy * usage$. Based on these definitions, we now discuss the results of this study in terms of the observed usage and effectiveness, noting that the consistency of the observed usage would make any attempt to separate efficacy from effectiveness a tenuous prospect at best. We then conclude the discussion by examining the limitations of this study and potential directions of future research.

Usage

Usage data recorded by the Reading Tutor indicates that participants spent 323 minutes on average engaged in Reading Tutor activities over the course of a four-week intervention. This exposure was relatively consistent with a standard deviation of only 45.86 minutes or 15%. It is also interesting to note that the lab was only available for 450 minutes during each treatment period (18 actual school days * 25 minutes per day available to each class), indicating a very high overall utilization of better than 70%. This value is even more impressive when considering the context of this study during the last two months of the school year. This is a very chaotic period and there were many end-of-year assemblies, field trips and other distractions.

This result supports the observation that students were highly motivated to work on the Reading Tutor. Teachers reported that students regularly reminded them when it was time to use the lab and looked forward to their scheduled time throughout the duration of the study. Teachers also reported a belief that the Reading Tutor was beneficial to their students. The utilization data supports this perceived value, especially in the context of the work of Mostow& Beck (2003) identifying teacher as the largest influence in predicting Reading Tutor usage.

As a final usage note, we also briefly consider the usability of the Reading Tutor for ELL's. The results of this study confirm the findings of Li (2002) that children with very low English proficiency are able to interact effectively with the tutor. All participants in this study were able to operate the Reading Tutor with very minimal support. Participants were assisted in the registration process and then completed the initial tutorial reading activity independently. Based on the direct observations of the principle investigator while supervising tutoring sessions, it is theorized that the children were empowered by their ability to work within the Reading Tutor environment independently and this may have accounted for some of the observed enthusiasm and consequently also been a factor in the positive outcomes. The only area where additional direction was consistently requested was in writing tasks, and even this was infrequent and more commonly related to writing skills than operating the tutor.

Effectiveness

Clearly, the results indicate the Reading Tutor was effective, primarily in terms of raising fluency. However, ultimately the object of the exercise in literacy education is

comprehension. Many studies indicate a direct link between fluency and comprehension based on the ability of the fluent reader to redirect attention from decoding to comprehension (Denton *et al.*, 2004; NRP, 2000; Parker, Hasbrouck & Lara-Alecio, 2001; Snow, Burns & Griffin, 1998). Fuchs & Deno (1991) have specifically demonstrated a strong correlation between CBM oral fluency measures and reading comprehension measures from the Woodcock Reading Mastery Test.

Fluency has also been identified as a critical component of comprehension in research specific to English language learners (August, 2003; Parker *et al.*, 2001). While the results of this study do not show a significant difference in comprehension scores between treatment and control conditions, we do note that treatment groups did consistently out gain controls in the comprehension measure. Further, the magnitude of the gains observed in fluency measures give a good basis to believe the tutor can effectively support ELL's comprehension growth as well.

In order to more fully examine the effectiveness of the Reading Tutor, we now address two critical questions: what needs of ESL students did the Reading Tutor effectively meet and what can be done to make the Reading Tutor more effective for ESL students.

What does the Reading Tutor have to offer English language learners?

Foorman & Torgesen (2001) conclude that children who are most at risk for reading failure should be supported by the same instructional components as their higher performing peers, but that these components need to be emphasized in ways that make them more comprehensive, intensive and explicit and further that these characteristics be

supported by a base of small-group and one-on-one instruction. Research has shown that intensive reading interventions designed for native English speakers are also effective in bilingual settings when those interventions are carried out with fidelity and high levels of student engagement (Parker *et al.*, 2001). The Reading Tutor supports all of these general criteria.

Clearly, it can provide one-on-one assisted oral reading opportunities for at risk bilingual students. Additionally, the interventions used by the tutor support phonological awareness and provide explicit models of decoding strategies, both of which are critical to teaching English reading to ELL's (Denton et al., 2004). Further, Mostow & Beck (2003) cite the Reading Tutor's ability as an automated tutor to ensure treatment fidelity to an extent not possible by human interventions. Qualitative assessments also indicate the tutor is very effective in engaging and motivating students as discussed previously in the context of usage.

Slavin & Cheung (2003; 2004) in reviewing the body of research on language of instruction identify that there is strong evidence supporting paired bilingual strategies where students are taught reading in both English and their native language concurrently. In this context, the results of this study combined with the body of research detailed in Mostow *et al.* (2003a) indicate that the Reading Tutor may be a very effective tool in supporting English literacy for English language learners.

What can the Reading Tutor do better to support English language learners unique needs?

We have paid a great deal of attention to the reasons why interventions designed for native English speakers are relevant for English language learners. We now address some of the unique characteristics of this population in an attempt to identify areas for future improvements to the Reading Tutor that will bolster its effectiveness for ELL's.

Li (2002) in her pilot investigation of the usability and benefits of the Reading Tutor for ESL students identifies two major areas where the reading tutor could be modified specifically to better support the needs of ESL students. They are the need to better support the limited oral vocabulary and background knowledge of ELL's and the need to provide more culturally sensitive and content appropriate reading material targeted toward this population.

The Reading Tutor relies primarily on a background of word knowledge to build up vocabulary and consequently comprehension. The limited oral vocabulary of English language learners may account in part for the limited gains observed in comprehension measures during this study. Li (2002) specifically identifies the need for illustrations in the tutor to provide ELLs an alternative source to draw meaning from. The version of the Reading Tutor used in this study contained very limited illustrations, and these were primarily used in level K material. The addition of this content would not only benefit ELLs, but would provide additional context to all readers. Further, this type of modification would add to the look and feel of the tutor, only serving to aid the motivational component of the tutor.

The Reading Tutor may also be more effective for ELLs if its reading material included more culturally sensitive content. As cited earlier, educational statistics demonstrate that the roughly 76% of this nations English language learners come from

Hispanic cultures (Moss & Puma, 1995; Ruiz de Velasco & Fix, 2000; Zehler *et al.*, 2003) and therefore it may be possible to address a large majority the ESL population with a limited but focused amount of additional material. Quintana (2001) presents a unique methodology to inventory the reading preferences of Mexican immigrant students. Her research focuses on the reading preferences of sixth through ninth grade students, but can equally be applied to younger children. By using methods like these, new reading material in the tutor will benefit ELLs by providing them with content that supports their comprehension based on a better alignment with their particular background knowledge and interests.

Limitations and Directions for Future Research

Clearly the most significant limitation of this study is its size. No generalizable conclusions can be drawn from a single study of 34 students over a span of two months with only one month of experimental treatment. However, these results do indicate that there is a strong potential for this technology in literacy instruction for English language learners. As such, the most pressing direction for future research should be to scale up this work to larger populations using commercially norm-referenced measures over a much longer treatment periods. Future research should also address the following potential confounds that may exist in the results presented here.

First, as this sample population was drawn from students who volunteered to stay after school to participate in a computer based reading tutor program, the potential of a sample bias must be considered. It seems reasonable that our population may have included a disproportionate amount of students who were very eager and motivated to

become more fluent English readers. In discussing this concern with teachers, they indicated their belief that the students enrolled in the study were representative of the general attitudes and abilities of their students. However, the reliability of this qualitative measure is suspect and should not be relied upon. Future research should avoid such potential sample bias in order to confirm these results.

A more serious limitation that must be addressed in future research is the potential confounds from the variable nature of language of instruction. In this study, the SSR condition did not control for (or even record) language of instruction. It is therefore not clear how much of the observed gains may have been attributable to English only instruction in the treatment condition versus a mix of Spanish and English instruction during control treatment. This mix of language of instruction is common in bilingual instruction and may therefore be difficult to control for, but future research can put measures in place to record the balance of the language of instruction during control treatment and account for it in the analysis of results. Additionally, cleaner studies may be done in schools that employ English-only immersion methods, though it is not clear how those results could be applied to the more commonly used bilingual model.

Conclusion

The technology employed by the Reading Tutor clearly has the power to provide inexpensive one-on-one assisted oral reading opportunities within ESL classrooms. This research gives an initial indication that this practice may be significantly more effective in helping English language learners develop English literacy skills than the common practice of sustained silent reading. The technology seems to be both highly motivating

and effective in engaging ESL students in reading activities. All study participants were able to independently interact with the tutor indicating that the Reading Tutor is accessible to students with very limited English proficiency.

Future research needs to be done to validate these findings and should include controls for language of instruction. Additional research is also needed to identify areas where the Reading Tutor can be improved to specifically meet the needs of Bilingual students. This should include research into how to best provide support for comprehension in a population with a limited English oral vocabulary.

Appendix A: Additional Results (Group Mean Charts)

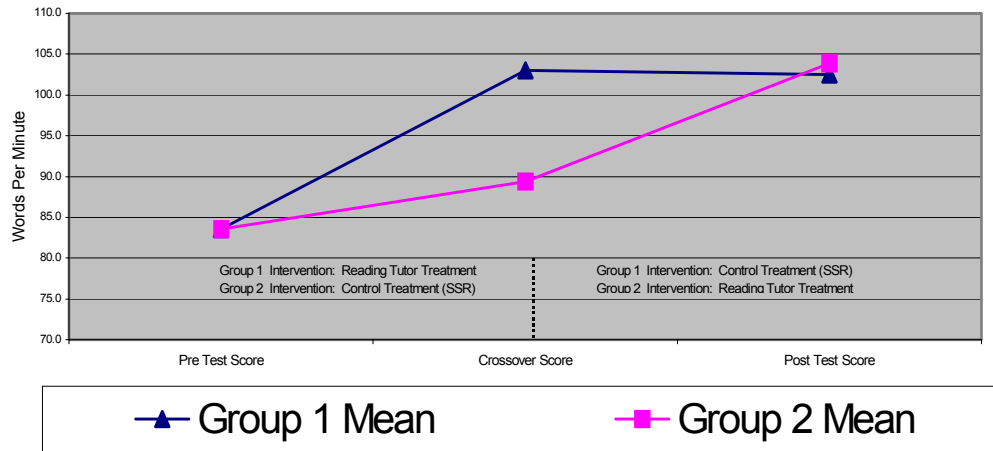


Figure 2: Fluency results for total words read in one minute

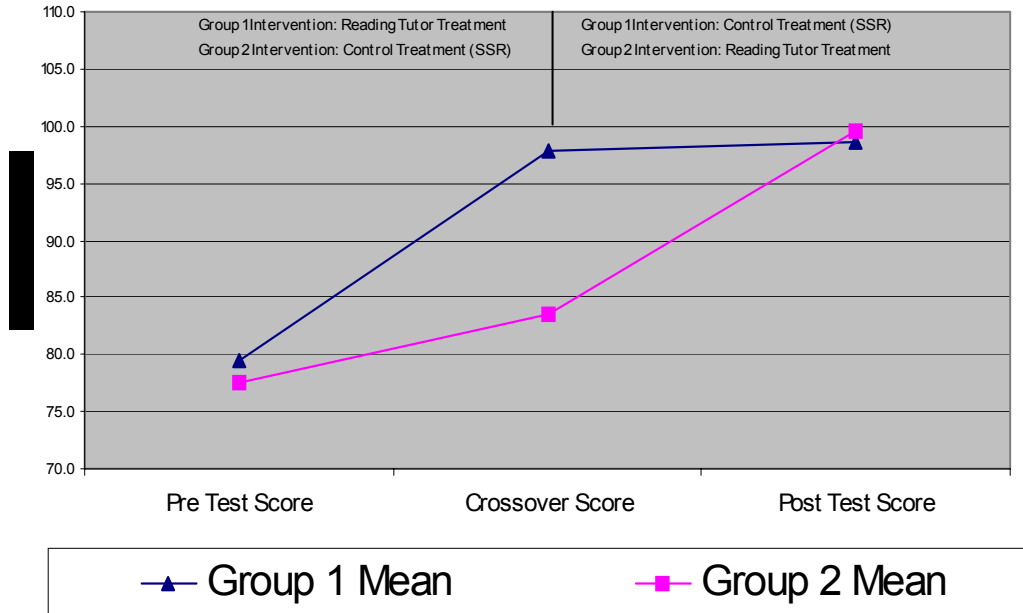


Figure 3: Fluency results for words read correctly in one minute

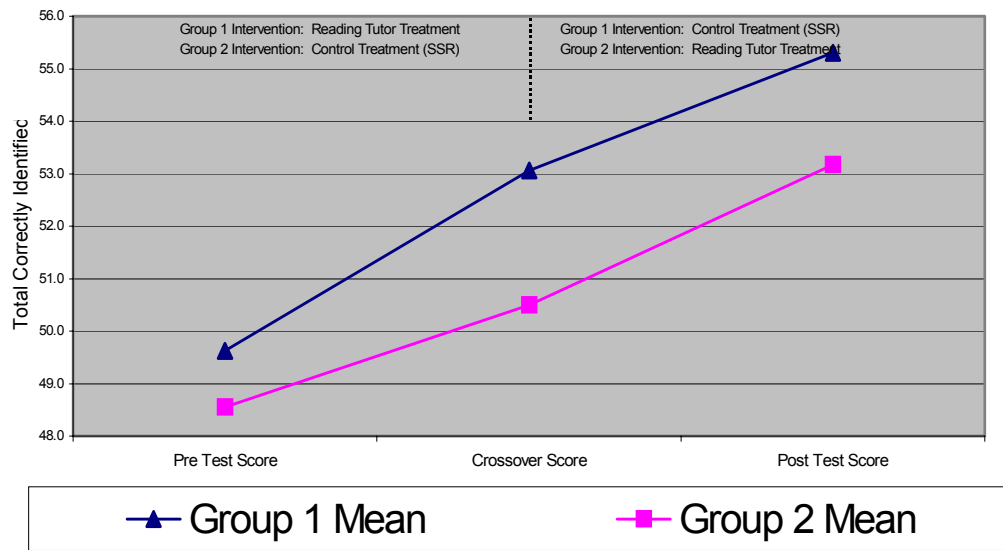


Figure 4: Timed Sight Word results

Appendix B: Study Summary Based on National Reading Panel Coding Scheme (NRP, 2000)

States or countries represented in sample	Illinois: Suburban Chicago community
Number of different schools represented in sample	1: (Eugene Field Elementary School)
Number of different classrooms represented in sample	4 total (1 second grade, 1 third grade, 1 multiage 3-4, 1 multiage 4-5)
Number of participants	34
Age	7-11
Grade	second through fourth
Reading levels of participants	Below grade level though grade level as measured by school districts informal reading inventories
Whether participants were drawn from urban, suburban, or rural settings	Suburban
Pretests administered prior to treatment	School District's Curriculum Based Measures for fluency, sight word recognition and comprehension.
Socioeconomic status (SES)	Predominantly low SES
Ethnicity	Hispanic
Exceptional learning characteristics	Limited English Proficient (LEP) Students
First language	Spanish
Explain any selection restrictions that were applied to limit the sample of participants	No performance based restrictions applied. Some students omitted because of scheduling conflicts.
Concurrent reading instruction received in classroom	Standard district curriculum for bilingual students including reading instruction in both Spanish and English
How was sample obtained?	Volunteers were solicited from all bilingual second through fourth grade classrooms
Attrition Number of participants lost per group during the study Was attrition greater for some groups than others?	36 students were initially entered, 2 were unable to participate because of scheduling problems during pull out lab sessions
Setting of the study	Classroom for control condition, pull out computer lab for experimental condition
Design of study	Crossover design with each group receiving both treatments for one-month intervals. Students randomly assigned to group stratified by pre-test score and grade level.
Describe all treatment and control conditions; be sure to describe nature and components of reading instruction provided to control group	2004 Reading Tutor; regular classroom instruction consisting primarily of sustained silent reading along with some guided reading activities. Control condition included both English and Spanish reading activities.
Explicit or implicit instruction?	The Reading Tutor provides help on oral reading, consisting of large amounts of implicit instruction by modelling fluent reading and reading individual words. By pointing out specific instances of letter-to-sound rules (<i>a</i> here makes the sound /a/), the Reading Tutor also provides explicit instruction at the grapheme-to-phoneme level.

Difficulty level and nature of texts	Authentic text ranging in level from pre-primer through fifth grade and including a mix of fiction and non-fiction. Reading Tutor inserted short factoids to introduce some new words. Classroom instruction consisted of self selected material during independent reading time.
Duration of treatments	20 to 25 minutes per day, five days per week. Treatment duration was one month per condition. Actual usage logged by reading tutor indicated students averaged 323 minutes of total usage over 4 week treatment (standard deviation of 45.86 minutes)
Was fidelity in delivering treatment checked?	Pull out lab supervised by primary investigator, daily contact / communication with classroom teachers.
Properties of teachers/trainers	Computer based tutor for experimental condition. Control condition primarily independent study, with minimal interventions provided by regular classroom teacher.
Number of trainers who administered treatment	N/A
Computer/student ratio	1:1
Type of computers	IBM-compatible Pentium personal computers (500 MHz) running under Windows 2000 Pro
Special qualifications	The Reading Tutor listens to children read aloud
Length of training	N/A
Source of training	N/A
Assignment of trainers to groups	N/A
Cost factors	Personal computer costs ~\$2000; cost of software depends on accounting for research and development costs
List and describe other nontreatment independent variables included in the analysis of effects	Pre-test Score; Gender; Grade, strongly correlated to teacher/classroom; Treatment Month, either experimental treatment in first month or second month
List processes that were taught during training and measured during and at the end of training	N/A
List names of reading outcomes measured	School District's standard Curriculum Based Measures for fluency (total words read), fluency (words read correctly), sight word identification (words identified automatically), sight word identification (words decoded correctly), and comprehension (cloze test).
List time points when dependent measures were assessed	April 2004, May 2004 & June 2004
Any reason to believe that treatment/control groups might not have been equivalent prior to treatments?	No; assignment to groups based on pretest score.
Were steps taken in statistical analyses to adjust for any lack of equivalence?	Yes; Paired T-test used to analyse within subject data. ANCOVA analysis treated pre-test score as a covariate.
Result: 2-tailed paired T-test (n = 30)	Fluency (Total Words Read): $p = 0.00038$ Fluency (Words Read Correctly): $p = 0.00014$ Sight Words (Timed): $p = .056$
Difference: treatment mean minus control mean	Fluency (TWR): Reading Tutor > control by 13.03 words per minute Fluency (WRC): Reading Tutor > control by 13.10 words per minute Sight Words (Timed): Reading Tutor > control by 1.60 words
Effect size	Fluency (TWR): 1.16 Fluency (WRC): 1.29 Sight Words (Timed): 0.58

Summary statistics used to derive effect size	$(\text{Treatment mean} - \text{control mean}) / (0.5 * (\text{treatment SD} + \text{control SD}))$ Fluency (TWR): $(17.13 - 4.10) / (0.5 * (11.12 + 11.30)) = 1.16$ Fluency (WRC): $(17.70 - 4.60) / (0.5 * (9.83 + 10.49)) = 1.29$ Sight Word (timed): $(3.03 - 1.43) / (0.5 * (2.66 + 2.84)) = 0.58$
Number of people providing effect size information	N = 30. All participants except for 4 students who were not available for testing at least one test point (pre, mid or post).
Length of time to code study	Uncertain
Name of coder	Robert S. Poulsen Note: Table format and coded information specific to Reading Tutor functional descriptions adapted from (Mostow <i>et al.</i> 2003a)

Table 4: NRP Style Research Summary

References

- Aist, G. (2002). Helping Children Learn Vocabulary During Computer-Assisted Oral Reading. *Educational Technology and Society*, 5(2), http://ifets.ieee.org/periodical/vol_2_2002/aist.html.
- Aist, G. (2000, June). Human Tutor and Computer Tutor Story Choice in Listening to Children Read Aloud. *Proceedings of the ITS'2000 Workshop on Modeling Human Teaching Tactics and Strategies*, Montreal, Canada, 8-10.
- Aist, G. & Mostow, J. (1997, October). When Speech Input is Not an Afterthought: A Reading Tutor that Listens. *Proceedings of the Workshop on Perceptual User Interfaces*, Banff, Canada.
- Aist, G. & Mostow, J. (2000, June). Improving story choice in a reading tutor that listens. *Proceedings of the Fifth International Conference on Intelligent Tutoring Systems (ITS'2000)*, Montreal, Canada, 645.
- Aist, G., Mostow, J., Tobin, B., Burkhead, P., Corbett, A., Cuneo, A., Junker, B., & Sklar, M. B. (2001). Computer-assisted oral reading helps third graders learn vocabulary better than a classroom control – about as well as one-on-one human-assisted oral reading. In J. D. Moore, C. L. Redfield & W. L. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future* (pp. 267-277). San Antonio, Texas: Amsterdam: IOS Press.
- Arias, M. B. (1986). The Context of Education for Hispanic Students: An Overview. *American Journal of Education*. 95, 26–57.
- August, D. (2003). Supporting the Development of English Literacy in English Language Learners: Key Issues and Promising Practices. *CRESPAR Tech. Rep. No. 61*. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk. Retrieved May 30, 2004 from www.csos.jhu.edu/crespar/techReports/Report61.pdf
- August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Research Council.
- Baker, S. K. & Good, R. (1995). Curriculum-Based Measurement of English Reading with Bilingual Hispanic Students: A Validation Study with Second-Grade Students. *School Psychology Review*, 24(4), 561-578.
- Congressional Budget Officer. (1987). *Educational achievement: Explanations and implications of recent trends*. Washington, DC: The Congress of the United States.
- Chu-Chang, M. (1981). The Dependency Relation between Oral Language and Reading in Bilingual Children. *Journal of Education*, 163(1), 30-55.
- De La Rosa, D., & Maw, C. (1990). *Hispanic education: A statistical portrait*. Washington, DC: National Council of La Raza.
- Deno, S. L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education*, 37(3), 184-192.
- Denton, C. A., Anthony, J. L., Parker, R. & Hasbrouck, J. E. (2004). Effects of Two Tutoring Programs on the English Reading Development of Spanish-English Bilingual Students. *The Elementary School Journal*, 104(4), 289-305.

- Durgunoglu, A. Y. (1998). Acquiring literacy in English and Spanish in the United States. In A. Y. Durgunoglu & L. Verhoeven (Eds.), *Literacy development in a multilingual context: Cross-cultural perspective* (pp. 135–146). Mahwah, NJ: Lawrence Erlbaum Associates.
- Foorman, B. R. & Torgesen, J. (2001). Critical Elements of Classroom and Small-Group Instruction Promote Reading Success in All Children. *Learning Disabilities Research and Practice*, 16(4), 203-212.
- Fuchs, L. S. & Deno, S. L. (1991). Effects of Curriculum Within Curriculum-Based Measurement. *Exceptional Children*, 58(3), 232-243.
- Guadalupe De la Colina, M., Parker, R.I., Hasbouch, J.E., & Lara-Alecio, R. (2001). Intensive Intervention in Reading Fluency for At-Risk Beginning Spanish Readers. *Bilingual Research Journal*, 25(4), 417-452.
- Guerrero, M. D. & Sloan, K. (2001). When Exemplary Gets Blurry: A Descriptive Analysis of Four Exemplary K-3 Spanish Reading Programs in Texas. *Bilingual Research Journal*, 25(1 & 2). Retrieved May 21, 2004 from <http://brj.asu.edu/v2512/pdf/ar9.pdf>
- Haycock, K., & Navarro, S. (1988). *Unfinished business: Fulfilling our children's promise*. Oakland, CA: The Achievement Council.
- ISBE. (2003). *Illinois School Report Card: Eugene Field Elementary School*. Illinois School Board of Education. Retrieved June 4, 2004 from http://206.230.157.60/publicsite/reports/2003/school/english/1401602102003_2003_e.pdf
- Jia, P., Beck, J. E., & Mostow, J. (2002, June 3). Can a Reading Tutor that Listens use Inter-word Latency to Assess a Student's Reading Ability? *ITS 2002 Workshop on Creating Valid Diagnostic Assessments*, San Sebastian, Spain, pp. 23-32.
- Kao, G., & Tienda, M. (1995). Optimism and achievement: The educational performance of immigrant youth. *Social Science Quarterly*, 76, 1–19.
- Li, K. (2002). The Reading Tutor: A Pilot Investigation of the Usability and Benefits For English as a Second Language Learners. *Unpublished Cust 590 Graduating Paper*, University of British Columbia, Vancouver.
- Meskill, C., & Mossop, J. (2000). Electronic texts in ESOL classrooms. *TESOL Quarterly*, 34(3), 585-592.
- Meskill, C., Mossop, J., & Bates, R. (1999). *Electronic Text and English as a Second Language Environments*. Albany, NY: National Research Center on English Learning & Achievement.
- Mirkin, P. K. & Deno, S. L. (1979). *Formative Evaluation in the Classroom: An Approach to Improving Instruction*. (Research Report No. 10). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.
- Moss, M., & Puma, M. (1995). Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity. *First Year Report on Language Minority and Limited English Proficient Students*. Prepared for Office of the Under Secretary, U.S. Department of Education by Abt Associates, Inc., Cambridge, MA
- Mostow, J. & Aist, G. (2001). Evaluating Tutors that Listen: An Overview of Project Listen. In K. D. Rofbus & P. J. Feltovich (Eds), *Smart Machines in Education* (pp. 169-234). Cambridge, MA: AAAI Press / The MIT Press.

- Mostow, J., & Beck, J. (2003, November 3-4). When the Rubber Meets the Road: Lessons from the In-School Adventures of an Automated Reading Tutor that Listens. *Conceptualizing Scale-Up: Multidisciplinary Perspectives*, Park Hyatt Hotel, Washington, D.C.
- Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., Huang, C., Junker, B., Sklar, M. B., & Tobin, B. (2003a). Evaluation of an Automated Reading Tutor that Listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research*, 29(1), 61-117.
- Mostow, J., Aist, G., Huang, C., Junker, B., Kennedy, R., Lan, H., IV, D. L., O'Connor, R., Tassone, R., Tobin, B., & Wierman, A. (2003b). 4-Month Evaluation of a Learner-controlled Reading Tutor that Listens. In P. DelCloque & M. Holland (Eds.), *Speech Technology for Language Learning*. Lisse, The Netherlands: Swets & Zeitlinger Publishers.
- Mostow, J., Aist, G., Bey, J., Burkhead, P., Cuneo, A., Junker, B., Rossbach, S., Tobin, B., Valeri, J., & Wilson, S. (2002, June 27-30). Independent practice versus computer-guided oral reading: Equal-time comparison of sustained silent reading to an automated reading tutor that listens. *Ninth Annual Meeting of the Society for the Scientific Study of Reading*, Chicago, Illinois.
- Mostow, J., Roth, S. F., Hauptmann, A. G., & Kane, M. (1994, August). A prototype reading coach that listens. *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Seattle, WA, 785-792.
- NCES. (2003). National Assessment of Educational Progress: The Nations Report Card Trial Urban District Assessment, Reading Highlights 2003. from <http://nces.ed.gov/nationsreportcard/pdf/dst2003/2004459.pdf>
- NCLB. (2001). The No Child Left Behind Act of 2001, from <http://www.nclb.gwu.edu/pubs/legislation/nclb/nclb.pdf>
- NRP. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction, Reports of the Subgroups*. (NIH Publication No. 00-4754). Washington, DC: National Institute of Child Health & Human Development.
- Orfield, G. (1986). Hispanic education: Challenges, research, and policies. *American Journal of Education*, 95, 1-25.
- Padolsky, D. (2002). *How has the Limited English Proficient student population changed in recent years?* The National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs (NCELA). Retrieved April 27, 2004 from <http://www.nclb.gwu.edu/expert/faq/08leps.htm>.
- Quintana, C. C. (2001). The Development and Description of an Inventory to Measure the Reading Preferences of Mexican Immigrant Students. *Bilingual Research Journal*, 25(4), 417-435.
- Ruiz de Velasco, J., & Fix, M. (2000). *Overlooked and underserved: Immigrant students in U.S. secondary schools*. Washington, DC: The Urban Institute.
- Seder, R. C. (1998). *Bilingual Education: Reading, Writing, and Rhetoric*. Los Angeles, CA: Reason Public Policy Institute (RPPI).
- Slavin, R. E. & Cheung, A. (2003). Effective Reading Programs for English Language Learners: A Best-Evidence Synthesis. *CRESPAR Tech. Rep. No. 66*. Baltimore,

- MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk. Retrieved May 30, 2004, from www.csos.jhu.edu/crespar/techReports/Report66.pdf
- Slavin, R. E. & Cheung, A. (2004). How Do English Language Learners Learn To Read? *Educational Leadership*, 61(6), 52-57.
- Snow, C. E., Burns, M. S. & Griffin, P. (Eds.). (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Thomas, W. P. & Collier, V. P. (2001). *A National Study of School Effectiveness for Language Minority Students Long-Term Academic Achievement*. Center for Research on Education, Diversity & Excellence (CREDE).
- Van Hook, J., & Fix, M. (2000). A profile of the immigrant student population. In J. Ruiz de Velasco, M. Fix, & B. Chu Clewell (Eds.), *Overlooked and underserved: Immigrant students in U.S. secondary schools*. Washington, DC: Urban Institute.
- Verhoeven, L., & Aarts, R. (1998). Attaining functional biliteracy in the Netherlands. In A. Y. Durgunoglu & L. Verhoeven (Eds.), *Literacy Development in a Multilingual Context: Cross-cultural Perspectives* (pp. 111–133). Mahwah, NJ: Lawrence Erlbaum Associates.
- Zehler, A. M., Fleischman, H. L., Hopstock, P. J., Stephenson, T. G., Pendzick, M. L., and Sapru, S. (2003). *Descriptive Study of Services to LEP Students and LEP Students with Disabilities. Volume I: Research Report*. Submitted to U.S. Department of Education, OELA. Arlington VA: Development Associates, Inc.