Voice Recognition:  A New Assessment Tool?

Darla Jones (darlajones@gci.net)
April 12, 2004

Abstract

This paper presents the results of a study that evaluated the accuracy and efficiency of using voice recognition technology to collect oral reading fluency (ORF) data for classroom-based assessments. The primary research question was as follows: Is voice recognition technology a valid and reliable alternative to traditional oral reading fluency data collection methods?

Results indicated that the VR data collection process employed in this study was clearly and significantly less accurate than traditional data collection methods. Nonetheless, several positive findings were identified. Accuracy rates, although lower than desired, were fairly stable over time, indicating that VR should be further investigated as a potential assessment tool, especially if more time were invested to improve VR accuracy rates. Findings also indicated that student accents, reading speed, colds or allergies, and background noise were not major factors influencing accuracy rates for this study. Only one factor (a speech impediment in the form of a lisp) was a major barrier to accuracy, resulting in the exclusion of one subject from the sample. The paper concludes with specific recommendations for voice recognition implementation and suggestions for future research.

Keywords

Voice Recognition, Speech Recognition, Reading, Oral Reading Fluency, Classroom-Based Assessments

Voice Recognition: A New Assessment Tool?

Over the past 20 years, numerous technological inventions have drastically altered the way educators work, sometimes radically improving efficiency. Whether locating the perfect lesson plan in cyberspace or generating 150 progress reports with the single click of a mouse, today's teacher has a myriad of technical tools available to save valuable time. Of all these tools, word processing is arguably one of the simplest, most versatile, and most powerful. However, despite the popularity of word processing, there are some drawbacks to its constant use, including repetitive motion injuries and fairly slow average typing speeds. Consequently, many technology advocates believe that voice recognition will become increasingly important in the near future (Barksdale, 2002).

Although word processing has facilitated an efficient writing and revision process, most people can type at a rate of 40 words per minute with the average range falling between 21 and 55 words per minute (Ostrach, 2002). Given the fact that most people speak at a rate of 137 words per minute with an average range between 106 and 168 words per minute (Siegler, 1995), the potential improvement in productivity is times 3. In addition, voice recognition promises to help students with motor impairments communicate their ideas more clearly and effortlessly than ever before (Goette, 2000).

Despite the many possible educational uses of VR technology, to date there have been relatively few rigorous studies assessing its effects since the technology has only recently become more practical and reliable. Most studies that have been conducted focus on specialized populations such as learning disabled (LD) students (Faris-Cole & Lewis, 2001; Higgins & Raskind, 2000) or English language learners (ELL) (Coniam, 1999; Derwing, Munro, &

Carbonaro, 2000), and most focus on older students or adults. Moreover, while a few studies focus on the area of reading (Williams, 2002; Mostow & Aist), most concentrate on the use of voice recognition for writing applications.

Hence, while there are many potential uses and benefits of VR technology for students, this paper attempts to expand the current research base by focusing primarily on the issue of teacher productivity. Since the No Child Left Behind Act (NCLB) requires teachers to collect more formative assessment data, it is essential that studies investigate technologies than can assist in this process.

## Research Questions

The primary question driving the research was as follows: Is voice recognition technology a valid and reliable alternative to traditional oral reading fluency data collection methods? Sub-questions that were explored include the following:

- Is there a significant difference between the accuracy of ORF data collected by a teacher compared to data collected via voice recognition?
- What factors affect the accuracy of the VR collection method? (Specific factors considered included student accents, speech impediments, colds or allergies, student reading speed, the number of student reading errors per session, and background noise.)
- How does the efficiency of traditional ORF data collection compare to the efficiency of VR collection methods?

Background

Of all the curricular areas targeted by statewide tests, reading is perhaps the most critical in terms of improving overall student academic performance since reading is an access skill necessary for all curriculum areas (Tindal, Heath, Hollenbeck, Almond, and Harniss, 1998). One of the most important means of measuring students' reading abilities involves frequent monitoring of oral reading fluency rates. Fuchs, Fuchs, Hosp, and Jenkins (2001) defined *Oral reading fluency* (ORF) as "the oral translation of text with speed and accuracy" (p. 239). Research has demonstrated that oral reading fluency (ORF) is highly correlated with general reading ability and comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Recent scientific evidence has reaffirmed this critical link between oral reading fluency and overall reading ability. Shaywitz and Shaywitz (2004) identified poor oral reading fluency skills as one of the most critical indicators of dyslexia and noted that observation of oral reading and timed tests requiring fluent reading are important tools for diagnosing reading problems.

Despite the volume of studies documenting the benefits of ORF measures, advocates such as Fuchs and Fuchs (1992) have acknowledged that educators often avoid using this technique due to time constraints. Consequently, Yell, Deno, and Marston (1992) recommended that computer programs be developed to automate the ORF measurement process. Although this idea was a mere fantasy a few short years ago, recent major advances in voice recognition technology have made this dream worthy of pursuit.

Admittedly, the complexity of voice recognition has ensured numerous challenges in its evolution, several of which have prevented its widespread use in the classroom. Earlier generations of VR software required long voice training sessions and discrete word input with relatively low accuracy rates (Coniam, 1999). Environmental concerns such as excess noise and changing vocal patterns have also hindered the success of VR sessions (Goette, 2000; Berliss-Vincent & Whitford, 2002).

However, many of these problems have been solved with better hardware and much more sophisticated software. Most current software now utilizes continuous speech input, allowing users to speak normally rather than halting after each word. Training sessions have also shortened dramatically, and many users report accuracy rates above 98% (*Voice Recognition News*, 2003). Moreover, better microphones in the form of noise-canceling headsets have improved transcription accuracy.

## Voice Recognition Research

Although there are currently relatively few studies examining the use of voice recognition in education, a few seminal researchers have begun to define the field and pose important questions. Studies by Coniam (1999) and Derwing, Munro, and Carbonaro (2000) demonstrated that voice recognition software is often less successful with English Language Learners (ELL) than English-only speakers. The use of voice recognition as an assistive technology for students with special needs has also been the focus of recent research. Faris-Cole and Lewis (2001) found that earlier versions of Dragon NaturallySpeaking and PowerSecretary were highly inaccurate and ultimately slower than writing by hand.

However, other studies have produced more positive findings. Raskind and Higgins (1999) found that students who used either discrete or continuous voice recognition technology for writing dictation showed significant gains in word recognition, spelling, and reading comprehension as compared to a control group utilizing technology without voice recognition.

In addition to research focusing on ELL students and those with special needs, a number of software programs are beginning to utilize speech recognition to assist students with their reading. For example, Project LISTEN's software entitled The Reading Tutor analyzes a student's reading patterns and provides assistance when it detects problems with word identification or other bottlenecks preventing fluent reading (Mostow & Aist, 1999). A similar program called The Reading Assistant is currently under development by Soliloquy Learning. Improving oral reading fluency is a primary goal of the new software, and it is one of the only under development for the Macintosh platform (Adams, 2003).

## Method

### *Setting and Participants*

Fourteen students from a single fourth grade class in Anchorage, Alaska, participated in the study. (Although the initial sample size was 15, one student was ultimately excluded from the final data analysis because his pronounced lisp prevented the software from functioning properly.) The final sample consisted of 9 girls and 5 boys with a mean age of 9.5 at the start of the study. The ethnic make-up of the sample was as follows: 9 Caucasian, 2 Hispanic, 1 Pacific Islander, and 2 Alaska Native. Of this group one student was classified as an English-language learner, and 3 students had noticeable accents. Reading levels were also mixed, with 3 students reading at below-average rates, 7 reading at average rates, and 2 reading at proficient rates based on an initial reading assessment I conducted at the beginning of the study.

### *Procedures*

This study utilized the software program Dragon NaturallySpeaking Preferred version 7.2 (ScanSoft) running on a Dell Inspiron 5150 laptop with 1 Gigabyte of RAM and a 60 Gigabyte hard drive. During the initial voice training session, students established a voice profile by reading the passage entitled “Stories Written by Children” (one of the choices provided in the program) since it had been designed specifically for younger children. Students then read a second passage (not provided in the program) and were asked to pause after each paragraph so the program’s correction window could be used to improve recognition accuracy rates. Use of the correction window entails highlighting the incorrect phrase, using the designated hot key combination to bring up the correction window, and selecting (from a list provided by the program) or typing the correct phrase.

Following initial voice training, oral reading fluency data collection began. Each research session resulted in the collection of 2 oral reading fluency samples: one via the traditional method, and one via the voice recognition method for a total of 10 traditional and 10 VR samples per student at the end of the study. For traditional collection samples, students read a passage while data was recorded manually onto a copy of the passage. I calculated the number of words read correctly (WCPM) by counting the total words attempted (pronounced correctly and self-corrected) and subtracting words read incorrectly (words omitted, mispronounced, substituted, added or transposed).

For VR data collection samples, the student wore a microphone and read the selected passage into a Microsoft Word Document for a period of one minute. While the student read, I simultaneously collected data by hand to allow calculation of the WCPM for later comparison to the computer transcription. Following the VR session, the Word Count feature in Microsoft Word was used to count the total number of words transcribed and the playback feature in Dragon NaturallySpeaking was used to count the number of computer transcription errors. (Transcription errors included words omitted, words added, or words incorrectly recorded by the software. Punctuation was not scored since it was not a factor in determining the number of words read correctly in a minute.) This process enabled the subsequent calculation of the software's accuracy percentage rate by dividing the number of errors by the total number of words recorded and subtracting the quotient from 100.

For the purposes of this study, I designated an accuracy level of 95% as the minimum acceptable standard for content validity. I selected this benchmark because the 95% level has been the technology industry's standard definition of an acceptable accuracy rate for continuous speech input (Barker, 1998), although ScanSoft now advertises a 98% accuracy rate for their software (*Voice Recognition News*, 2003).

In order to ensure the reliability and consistency of the passages, all reading probes were at the fourth grade level and drawn from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) system (Good & Kaminski, 2002). Time records were kept on a log for each student along with ORF scores. Each session was audiotaped to enable data verification by an independent evaluator. At the end of each session, field notes were taken to record any relevant observations.

In addition to the standard procedures followed in all sessions, special procedures were followed during four sessions in an attempt to boost VR accuracy. Two primary methods are used to improve Dragon NaturallySpeaking's accuracy rates. First, use of the correction window is an essential tool for improving recognition. Accordingly, I initially chose to perform phrase corrections following Session 3 and Session 5. For both sessions, I counted the number of transcription errors and then made phrase corrections using the correction window. I then recorded the total amount of time invested for later analysis.

The second method of boosting performance involves reading additional training passages in order to expand users' voice profiles. Hence, I decided to try this approach with students immediately prior to Session 7, again maintaining time logs to determine total student time invested. For this procedure students read "Sales Letters," which was one of the passages labeled "easier reading" in the list of options. Problematically, there was a near-universal drop in VR accuracy rates following this additional training. I later determined that the vocabulary had been too difficult for students, causing mispronunciations that resulted in lower accuracy rates for subsequent sessions. Eventually, I was forced to restore voice profiles from a backup copy to improve recognition rates.

## Results

### *Voice Recognition Accuracy Rates*

Although 11 of the 14 students participating in the study had at least one session in which a 95% accuracy rate was achieved, results from this study were not consistently high enough to be considered valid or reliable. The first six data collection sessions proceeded smoothly with very few problems. Since the additional voice training session that occurred immediately prior to Session 7 caused a drop in accuracy rates (most likely due to the difficulty of the vocabulary in the training passage), I chose to analyze two sets of data: results from Sessions 1-6 only and results from all sessions combined, including those following the accuracy drop (Sessions 1-10).

Table 1 presents accuracy rate data for all students in the study. Every student had a higher average accuracy rate for the first six sessions than for all sessions combined, indicating that overall results might have been higher had the technical problem not occurred in Session 7.

Table 1: *Accuracy Rate Results*

| | All Sessions | | Sessions 1-6 | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Student 1 | 90.1 | 5.5 | 92.0 | 4.6 |
| Student 2 | 90.9 | 2.9 | 91.7 | 3.4 |
| Student 3 | 88.2 | 3.8 | 90.5 | 2.3 |
| Student 4 | 90.3 | 4.7 | 91.0 | 4.7 |
| Student 5 | 88.3 | 4.9 | 90.3 | 2.3 |
| Student 6 | 88.3 | 4.8 | 91.0 | 3.3 |
| Student 7 | 89.1 | 4.2 | 91.8 | 2.6 |
| Student 8 | 87.8 | 5.3 | 90.7 | 3.5 |
| Student 9 | 91.4 | 3.6 | 94.0 | 1.1 |
| Student 10 | 91.2 | 3.4 | 92.8 | 3.3 |
| Student 11 | 89.8 | 4.6 | 92.3 | 3.2 |
| Student 13 | 91.1 | 4.6 | 94.0 | 2.1 |
| Student 14 | 89.4 | 3.3 | 91.5 | 2.2 |
| Student 15 | 92.4 | 3.4 | 94.7 | 2.0 |
| All Students/All Sessions | **89.9** | **4.3** | **92.0** | **3.1** |

*Note.* Student Number 12 was excluded from the study, so his data is not included in this table.

Table 1 shows that the first six sessions were more successful than those following the second voice training session. All students had a mean accuracy rate of at least 90% for Sessions 1-6, whereas the majority of students had mean accuracy rates below 90% when all sessions are considered.

*Reading Speed and Passage Error Rate Data*

In order to analyze whether or not reading speed or the number of reading errors in a session were connected with accuracy rates, I conducted a correlation analysis for each of these factors. First, I computed the correlation between students' initial informal reading inventory score and their mean accuracy rate for all sessions. The resulting correlation coefficient was .0034, indicating that there was no relationship between the students' informal reading inventory scores and their mean accuracy rates. Next, I calculated correlation coefficients for each student by comparing student reading errors (words skipped, mispronounced, or inserted) for each session with the VR accuracy rates for each session. Again, the correlation analysis did not indicate a significant relationship between the number of reading errors in a session and the accuracy rate for that session. Only 2 students showed correlations above .5 for this analysis, and most correlations were near zero.

*Time Data*

In addition to the question surrounding the validity and reliability of VR accuracy rates and factors influencing those accuracy rates, the issue of efficiency was a central question in this study. Since the ultimate goal is to improve teacher productivity and promote more frequent oral reading fluency assessments, it is critical to examine the time investment required for the voice recognition process. Table 2 presents a summary of time invested in order to achieve the accuracy rates discussed in the previous section.

Table 2: *Time: Minutes per Student*

| | *M* | *SD* |
|---|---|---|
| Voice Training Session 1 | 21.2 | 3.3 |
| Voice Training Session 2 | 8.7 | 2.7 |
| Total Passage Correction Time (3 sessions) | 20.2 | 3.8 |
| Total Time Investment | 50.1 | 6.8 |

*Note.* n=14.

Voice Training Session 1 represents the time for the entire initial training procedure. This total also includes the time it took to correct transcription errors for the additional passage using the software's correction window. Voice Training Session 2 was the session immediately prior to data collection Session 7 in which students read the selection entitled "Sales Letters." This was the training session that ultimately proved to be problematic, resulting in a drop in accuracy rates. Finally, the Total Passage Correction time represents the cumulative total of correction window procedures for Probes 5, 9, and 17.

## Discussion

As a small, exploratory case study, it would be inappropriate to generalize results broadly. While the sample was mixed in terms of ethnicity, gender, and reading speed, subjects in this study were limited to volunteers from a single fourth grade classroom in Anchorage, Alaska. It is possible that student enthusiasm to participate in the study contributed to higher voice recognition accuracy rates and that less willing students might have been less successful. Moreover, although the sample initially included one special education student with a pronounced speech impediment, this student's results were ultimately excluded from the study when it became evident that the voice recognition process was totally inadequate for him. Since this exclusion left no special education students in the final sample, this further limits the scope of the study.

The fact that all subjects were from the fourth grade is another important consideration. Oral reading fluency assessments are also frequently conducted with younger students in grades 1-3. Since this study did not include any students from these earlier grades, it would be important to conduct further research to determine if voice recognition could be successful with students at all grade levels.

### *The Question of Accuracy*

Clearly, the VR data collection process employed in this study was significantly less accurate than traditional data collection methods. Even when ignoring the odd punctuation and capitalization patterns generated by the automatic transcription process, an overall error rate of just over 10 words for every 100 spoken would make it difficult for a teacher to interpret what their students had actually read to the computer. As shown in Table 1, even when only the most successful six sessions were considered in isolation, the average error rate was still an unacceptably high 8 words per 100 spoken.

Of course, it is possible and perhaps even likely that students might have eventually achieved the software's advertised 98% accuracy rates had more time been devoted to voice training and passage correction procedures. Barksdale (2002) reported achieving consistently high results with secondary students after two to three weeks of regular training, even with poor readers. However, because this study was designed to utilize voice recognition as an assessment tool rather than a more general-purpose writing and productivity tool, the goal was to minimize training time, which resulted in lower accuracy rates.

*Factors Affecting Accuracy*

Before the study began, I hypothesized that certain factors might influence the level of accuracy achieved in VR sessions. Given the fact that this was a small-scale case study with only 14 participants, it was impossible to conduct formal statistical analyses for most of these factors. Two correlation analyses indicated that there was little or no relationship between reading speed or reading errors and accuracy rates. For the remaining factors, I analyzed my observational notes to identify trends.

Since previous studies have demonstrated lower voice recognition accuracy rates for English language learners, I expected that students with accents might have lower accuracy rates. However, this did not prove to be the case for this study. Three students in the final sample had noticeable accents (one of whom was classified as ELL), but their average accuracy rates were not lower than average. In fact, one student in this group had initial accuracy rates (during Sessions 1-6) higher than the average participant. Although I also anticipated that colds or allergies might negatively affect performance, a review of the session logs did not indicate any obvious effects.

Speech impediments, on the other hand, are possibly problematic. As previously noted, one student who participated in the study had a noticeable lisp. I ultimately chose to exclude his data from the final analysis because his results were consistently poor and significantly below the other participants. Hence, it may still be true that voice recognition software has problems interpreting speech from students with certain types of speech problems, even when a student consistently pronounces a word in the same way.

Another expectation I had formulated was that students who read more fluently might have higher VR accuracy rates. However, based on results from this study, this does not appear to be true. For example, Student Number 2, although a lower-performing reader (with a mean rate of 67.5 WCPM), spoke very clearly and deliberately into the computer and always remembered to skip a word that he did not understand rather than stumbling with it. As a result, his accuracy rates remained relatively stable even after the ill-fated voice training prior to Session 7. Student Number 3, by contrast, was one of the most fluent readers in the sample with a mean rate of 140.6 WCPM. However, she spoke very quickly and softly with resulting accuracy rates consistently below average. Her particular speech pattern caused the computer often to miss or confuse articles such as “a” or “the.” Another fluent reader (Student Number 1) tended to change her vocal rhythm fairly frequently, speeding up as she reached the end of a sentence and failing to pause before starting the next. This resulted in accuracy rates often lower than other students.

In addition to student speech patterns, I expected that background noise might play a significant role in accuracy rates. Although the majority of sessions were conducted in fairly quiet surroundings since the setting was in a small room connected to the library, there were some notable exceptions. On several occasions, physical therapy sessions in the adjacent room became noisy. However, none of the sessions where this was noted in the log appeared to be affected by this situation, indicating that the noise-canceling headset did a good job filtering out the background noise.

Interestingly, the lack of findings for this research question is one of the more positive outcomes of the study. The observation that accents, reading speed, colds or allergies, and background noise did not seem to have a noticeable impact on accuracy rates indicates that the software did a good job compensating for these factors for this group of students.

*The Question of Efficiency*

If accuracy is the most critical issue underlying this study, then efficiency is a close second. As previously discussed, one of the primary reasons that teachers seldom or never collect oral reading fluency data is a lack of time. Therefore, the primary objective in automating such assessments would be a substantial increase in productivity. Because efficiency is such an important issue if a new assessment technique is implemented, I chose to minimize voice training and correction procedures to simulate more accurately what might be realistic for a typical elementary teacher. However, this choice had a clear and direct impact on the software's ability to improve recognition, and it is possible that a higher time investment would have resulted in more acceptable accuracy rates.

On the other hand, the average total time invested in each student was just over 50 minutes, which is still considerable given the fact that traditional oral reading fluency assessments can easily be given in 1 1/2 to 2 minutes. For the time being, if VR data collection were to be considered as a reasonable alternative to traditional ORF assessment techniques, then it would be critical for students to complete voice training procedures individually (without teacher assistance) or simultaneously (in a computer lab, for instance). Moreover, students would need to be trained to make corrections on their own so that they could improve the computer's accuracy rates on a consistent basis. Under this scenario, a handful of class trips to the computer lab could make the initial time investment sensible, especially if students utilized the technology for additional purposes such as writing.

*The Primary Question*

The primary research question explored by this study was as follows: Is voice recognition technology a valid and reliable alternative to traditional oral reading fluency data collection methods?

At this point, the realistic answer to this question must be a qualified "not yet." Accuracy rates did not reach a consistently high enough level to be considered valid or reliable, and the time investment would likely be viewed as too high unless students could be trained to operate the program independently and/or for a variety of purposes. However, a primary objective of exploratory case studies is to generate questions worthy of further investigation. The fact that average accuracy rates were fairly consistent and reasonably high (although not high enough for assessment purposes) indicates that voice recognition may have advanced to the point that it should be investigated more thoroughly with younger students.

In this study students consistently achieved accuracy rates above 86% even after the technical problem occurred and without completing correction procedures after each session. Since the software is designed to improve based on more frequent corrections, further investigation into the use of VR technology as an assessment tool is warranted.

The technical problem that occurred after Session 6 also highlights some key points that should be considered when implementing voice recognition technology:

1. Proper training for teachers is critical before utilizing the software with students, even though the program is fairly simple to learn. A thorough understanding of how voice profiles are created, enhanced, backed up and restored is necessary for smooth operation of the software.
2. Voice training passages should be selected very carefully and with great attention to the vocabulary in the passage. The training passage entitled “Stories Written by Children” worked very well, but the unfamiliar vocabulary in “Sales Letters” caused a significant degradation in accuracy for the students in this study.
3. After initial voice training with a suitable passage, frequent use of the correction window is probably the best way to improve accuracy rates, especially with younger students.
4. Voice profiles and customized vocabulary lists should be backed up frequently in order to minimize frustrations should problems arise.

## Suggestions for Future Research

To date, very little research has examined the use of voice recognition with elementary students. Although voice recognition software has historically had more problems interpreting young children’s speech (Williams, 2002), more research is needed in this area, particularly since this study showed consistent accuracy rates above 90% with a minimal amount of training.

Questions worth exploring in future, larger-scale studies include the following:

1. How much time, including voice training and correction procedures, must be invested to achieve consistent accuracy rates of at least 95%?
2. What is the youngest age that voice recognition can be utilized effectively?
3. Does voice recognition now work as accurately for English language learners as it does for other students?
4. Does voice recognition work well in a computer lab environment with students sitting near one another as they speak?
5. Can young students complete the voice training process independently or in a lab setting?
6. Can young students learn to make corrections independently in order to improve accuracy rates?
7. If young students use voice recognition software to practice their reading for assessment purposes, would this process have a secondary benefit of improving their reading skills?

## Conclusion

Since voice recognition is becoming a more viable technology, programs that are currently under development and specifically targeted to reading (such as The Reading Tutor and Reading Assistant) may ultimately prove to be more useful than Dragon NaturallySpeaking for reading applications. However, if voice recognition becomes as widespread as many predict, then schools might be well-advised to train students to develop strong dictating skills in multipurpose applications such as Dragon NaturallySpeaking. This would enable students to perform a variety of tasks with the added benefit of improving their general literacy skills, as Higgins and Raskind found (2000). Since emerging brain research has demonstrated the critical importance of oral reading fluency skills (Shaywitz & Shaywitz, 2004), the use of voice recognition technology for individualized oral reading practice deserves further exploration.

It is also much too early to rule out voice recognition as a potential assessment tool. In the area of reading, it would enable students to practice reading aloud much more frequently, collect oral reading fluency data independently, and establish a permanent assessment record reflecting their reading progress.

# REFERENCES

Adams, M. J. (2003). A research-based approach ensures reading success. Retrieved March 15, 2004, from http://www.soliloquylearning.com/research_paper.html

Barker, D. (1998). Microsoft research spawns a new era in speech technology: Simpler, faster, and easier speech application development. *PCAI, 16*(6), 18-27. Retrieved March 3, 2004, from http://www.pcai.com

Barksdale, K. (2002). Speech recognition: How do we teach it? *Business Education Forum, 56*(3), 52-55.

Berliss-Vincent, J., & Whitford, G. (2002). Talking speech input. *Communication Disorders Quarterly*, *23*(3), 155-157.

Coniam, D. (1998). The use of speech recognition software as an English language oral assessment instrument: An exploratory study. *CALICO Journal*, *(15)*4, 7-23.

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27*(1), 49-64.

Derwing, T. M., Munroe, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, *34*(3), 592-603.

Dragon NaturallySpeaking Preferred (Version 7.2) [Computer software]. (2003). Peabody, MA: Scansoft.

Dragon Naturally Speaking [Software Manual]. (2003). Peabody, MA: Scansoft.

Faris-Cole, D., & Lewis, R. (2001). Exploring speech recognition technology: Children with learning and emotional/behavioral disorders. *Learning Disabilities: A Multidisciplinary Journal, 11*(1), 3-12.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*(1), 45-59.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256.

Goette, T. (2000). Keys to the adoption and use of voice recognition technology in organizations. *Library Computing*, *19*(3-4), 235-244.

Good, R.H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Higgins, E. L., & Raskind, M. H. (2000). Speaking to read: The effects of continuous vs. discrete speech recognition systems on the reading and spelling of children with learning disabilities. *Journal of Special Education Technology*, *15*(1), 19-30.

Mostow, J., & Aist, G. (1999). Giving help and praise in a reading tutor with imperfect listening--because automated speech recognition means never being able to say you're certain. *CALICO Journal, 16*(3), 407-424.

Mostow, J., Aist, G., Burkhead, P., Corbett, A., Cuneo, A., Rossbach, S., et al. (2002). Independent versus computer-assisted reading: Equal-time comparison of sustained silent reading to an automated reading tutor that listens. Retrieved December 12, 2003, from http://www-2.cs.cmu.edu/~listen/pubs.html

Ostrach, T. R. (2002). Typing speed: How fast is average. *Five Star Saffing Inc.* Retrieved January 5, 2004, from http://www.fivestarstaff.com/publication_typing.htm

Raskind, M. H., & Higgins, E. L. (1999). Speaking to read: The effect of speech recognition technology on the reading and spelling performance of children with learning disabilities. *Annals of Dyslexia, 49*, 251-281.

Reading Assistant [Computer software]. (2003). Needham Heights, MA: Soliloquy Learning.

Scansoft (2003). NaturallySpeaking 7. Retrieved August 17, 2003, from http://www.scansoft.com/NaturallySpeaking/

Shaywitz, S. E., & Shaywitz, B. A. (2004). The new science of reading and its implications for the classroom. *Education Canada, 44*(1), 20-23.

Siegler, M. A. (1995). Measuring and compensating for the effects of speech rate in large vocbulary continuous speech recognition. Retrieved January 5, 2004, from http://www.cs.cmu.edu/~msiegler/publish/Masters/MS.pdf

Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). Accommodating students with disabilities on large-scale tests: An experimental study. *Exceptional Children, 64*(4), 439-450.

*Voice Recognition News*. (2003). Retrieved September 20, 2003, from http://www.voicerecognition.com/vrnews

Williams, S. (2002). Speech recognition technology and the assessment of beginning readers. In *Technology and assessment: Thinking ahead--proceedings from a workshop* (pp. 40-49). Washington: National Academy Press.

Yell, M. L., Deno, S. L., & Marston, D. B. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique*, *18*(1), 99-112.