Williams, S.M., Nix, D., & Fairweather, P. (2000). Using Speech Recognition Technology to Enhance Literacy Instruction for Emerging Readers. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 115-120). Mahwah, NJ: Erlbaum.

Using Speech Recognition Technology to Enhance Literacy Instruction for Emerging Readers

Susan M. Williams
Vanderbilt University, Learning Technology Center, Box 45, GPC, Nashville, TN 37203
Tel: 615-343-7948, Fax: 615-343-7556
Email: susan.williams@vanderbilt.edu

Don Nix, Peter Fairweather
IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598-028
Tel: 914-945-2138, Fax: 914-945-4395
Email: donnix@us.ibm.com, pfairwea@us.ibm.com

Abstract: In this paper, we describe advances that make it possible to realize the potential of speech recognition to support emerging readers. These advances include improvements in the technology itself and in the design of Watch Me! Read, a tutor that uses this technology to support young readers. In addition, we evaluate Watch Me! Read using a framework for thinking about issues of learning involved in any attempt to implement new technologies. In most discussions of technology implementations, the learning issues remain relatively tacit. By making them explicit, it is possible to assess their coherence as a system (see Brown & Campione, 1996) and their correspondence with what is known about human learning.

Keywords: literacy education, learning environments

Introduction

Commercially available reading software often seems somewhat alien to learning to read. Such software must resort to clever schemes to compensate for its inability to react to youngsters reading aloud. One common strategy calls upon the child to perform tasks that presume to exercise the same skills that reading requires, with directions like, “Find the word on the screen that rhymes with this picture.” These sorts of activities fail to give children much of a sense of the experience of reading—not surprising, given their orientation toward isolated word recognition and their reliance on picture interpretation.

Researchers have long sought to create computer-based reading tutors that can listen to beginning readers, evaluate their performance, and provide them assistance as needed. Although speech enabled tools have become widely available for adults, work on applications for children stalled for a wide range of reasons including a lack of processing power and speech technology that was fine-tuned to the voices and needs of adults rather than young children (1). This paper discusses recent work to realize the potential of speech recognition to support emerging readers. We begin by describing Watch Me! Read (WM!R), a reading tutor designed by researchers at IBM’s T.J. Watson Research Center. Then, we describe an implementation of WM!R in an urban school system. In this description, we focus on the aspects of the software that have the potential to enhance literacy instruction and to facilitate positive change. We describe ways teachers incorporate the software into their instruction and how students benefit from using it.

Watch Me! Read

Watch Me! Read software is designed to give young children a sense of being a reader. Specifically, the designers’ goals are to provide reading practice, comprehension awareness, and a sense of reading as communication. The software uses speech recognition to assess students’ performance and provide them individualized feedback. It works in much the same way an adult might listen to a child, provide help with the pronunciation of words when the child falters, and ask questions to probe his or her understanding of what is being read.

In the WM!R environment, books appear on the computer screen much as they do in traditional form, i.e., text and illustrations are displayed on 2 facing “pages” of a graphic book (see Figure 1). A small, animated Panda acts as a guide, walking across the surface of the book, pointing to the current reading location, and providing feedback and

encouragement. Students are asked to read the text one phrase at a time. For students who are just beginning to read, the Panda reads each phrase first and then students read only the last word when the phrase is repeated. At the most advanced level, students read the entire phrase without assistance. The phrase being read is marked in color with the text read by the Panda in blue and the text read by the student in red. If a student does not know a word, he or she can click on the word and hear the Panda pronounce it. The student's voice is recorded as he or she reads the book. This recording is used later in the performance section of the program.

At the beginning of each page, the student can choose to hear an overview of what they are about to read. At the end of each page, the Panda asks the student a comprehension or prediction question based on the contents of the current pages. These questions are customized for each book. The student's answers are recorded using a graphical "boom box" tool displayed at the bottom of the page. The student can listen to his or her answer and re-record it if they like. WM!R does not provide immediate feedback for answers to comprehension questions, but the teacher can review recorded answers at a later time.

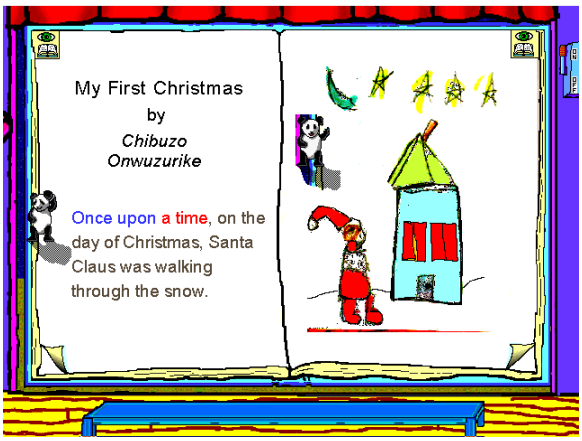

Figure 1. Reading view of Watch Me! Read showing a book written and illustrated by a student

After the student finishes reading, Watch Me! Read presents a performance of the book with the words highlighted as the book is read in the student's voice. If a camera is attached to the computer, the student can create a video introduction to the performance.

Two new tools supplement the WM!R software. The first is an authoring tool used to create new books for WM!R. Teachers, students, or other authors use this tool to type in text and import pictures. Once these elements are in place, they can create the voice of the Panda by recording the instructions, the text of the book, the comprehension questions, and the feedback messages. This tool can be used to record instructions in Spanish or other languages for children whose second language is English.

In addition to the authoring tool, a new diagnostic tool has been designed in collaboration with teachers. This tool calculates the percentage of words that each student has read correctly. Other reports give the percentage correct by student for standard word lists such as the DOLCH. Analysis can be limited to specific dates; therefore, it is possible to track a student's progress by comparing data for consecutive weeks or months on the same word list. Teachers can also prepare an individualized sight word list by having the diagnostic tool print a sheet containing words that the student did not recognize when using WM!R. A quick scan through the displays for students can easily show which students are having difficulty or those that have not used the system at all.

Issues in Adapting Speech Recognition for Young Children.

Many factors combine to make the evaluation of speech an extremely difficult task even for proficient readers (Fairweather, Nix, Oblinger, Adams, and Laffra, 1998): Oral reading contains numerous pauses, omissions, repetitions, insertions, and mispronunciations. Background noise may interfere. Microphones may be too close or too far away from the mouth. Other factors compound the difficulty when speech recognition is used in school settings with young children: Beginning readers are more likely to pause and mispronounce words. The variety of linguistic diversity is large, but training software to recognize each student's voice is impractical.

Each of these factors can increase the number of errors in judgements made by the software concerning whether or not a word is pronounced correctly. Work on WM!R has focused on improving the accuracy of these judgements using several approaches. First, more appropriate acoustic models were developed for words in the system's vocabulary. Previously, these models were based entirely on the voices of adults. In the WM!R project, 110,000 speech samples were collected from 1830 children in 20 sites all over the United States. These samples produced more accurate representations of children's relatively high-pitched voices. They also captured the regional and ethnic linguistic diversity that is present in most schools in the U.S.

In addition to the new acoustic model, additions were also made to the pool of words the tutor searched when analyzing students' speech. These new words represented common alternate pronunciations. For example, because so many children pronounce "comfortable" as if it were spelled "comfterble," that alternative was included. Miscues such as pronouncing "will" instead of "well" were also included as chaff words.

Although these modifications have reduced the recognition errors dramatically, complete accuracy can still not be assured. As one would expect, emergent readers tolerate false positive errors, generally because they are unaware of them. For example, if a child produces "this" for "that" in the text and the tutor accepts it, the child is usually unaware that anything out of the ordinary has happened. False negatives, on the other hand, frustrate children quickly if they occur too often. The WM!R interface is designed to minimize frustration by tempering the way such judgments are signaled. For example, when the panda signals that the child has not read what was asked, she never says "no" or "incorrect" but instead presents a simpler task, asking, for example, a more focused "What's this word?"

Classroom Implementation

WM!R is currently the focus of a Reinventing Education Project funded by IBM in the Houston Independent School District (HISD). Each participating classroom receives 3 computers each equipped with a CD-ROM drive, headphones, a microphone, a small camera, and the WM!R software. Although teachers are expected to use the program, they are free to decide how often they use it and how they incorporate it into their literacy instruction. Most of the teachers in this initial site used WM!R as an independent reading activity during center time. At times, students were expected to read a specific book that was part of their lessons for the week. At other times, they could choose any book. In a typical week, each student would spend approximately 30 minutes using the program.

Although WM!R is currently in over 30 classrooms in 7 schools in HISD, this paper is based on observations and interviews in the initial implementation site, an elementary school serving students in pre-kindergarten through grade five. Approximately 15% of the students in this school are in special education, 26% in Limited English Proficiency programs and 19% in gifted and talented classes. The school is 44% African American, 33% Hispanic, 20% Anglo, and 3% Asian. Sixty-two percent of the students are on free and reduced lunch.

Evaluation Framework

Our goal in this discussion is to describe WM!R's potential to support new views of learning and instruction and to report ways in which teachers understanding and use of the software mirrored these views. We organize this discussion using a framework introduced by Bransford, Brown, and Cocking in their recent book *How People Learn* (1999). This framework (see Figure 2) embodies a set of design principles for learning environments based on what the field of the learning sciences has learned about expertise, learning and transfer. The framework is not prescriptive. Rather, it asks four important questions to help designers and teachers reflect on the extent to which an environment is consistent with evidence on learning. These questions examine the extent to which an environment is student centered, knowledge centered, assessment centered, and community centered. In the next section, we discuss how the WM!R environment is consistent with each of these perspectives. We supplement the discussion with evidence from classroom

observations and teacher interviews. Although the perspectives are discussed separately, it is important to note the ways in which they overlap and support one another.

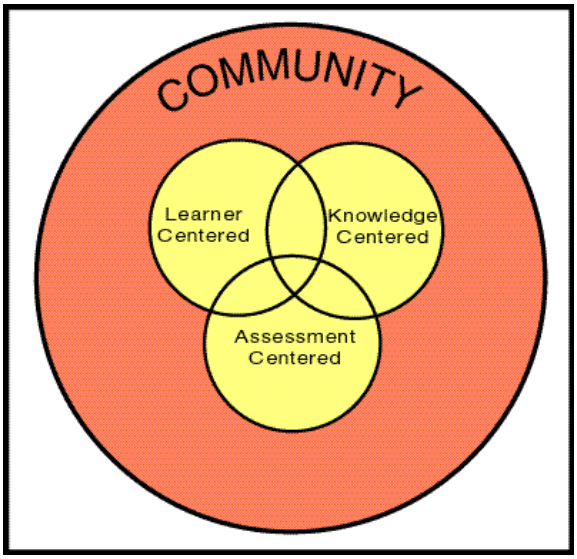

Figure 2. HPL framework for evaluating design environments

Learner Centered

Learner-centered environments are created by teachers who understand that students construct their own meanings beginning with the beliefs and knowledge that they bring to the classroom. These environments include a respect for students' current knowledge as representing a starting point for further learning. Many efforts to teach reading in large classrooms fall short of the goal of being learner centered. Students march through the curriculum in lock step, reading the same books, and working on the same aspects of decoding skills at the same time. There are few opportunities for individualizing instruction due to the number of students per teacher and the time available. The WM!R environment lets students choose their own books and read them at their own pace. The feedback provided lets even the beginning reader work independently and get as much (or as little) support as they need.

In the implementation site, where 26% of the children were classified as limited in English proficiency, teachers immediately recognized the importance of an environment that responded individually to their students needs. All teachers spontaneously mentioned how much their Limited English Proficiency and bilingual students benefited from using WM!R. For example, bilingual student's classes are conducted primarily in Spanish and Spanish is the primary language in their homes and communities. Only two students out of twenty in the bilingual class at the test site had English spoken at home. Thus, they have few opportunities to try out their English skills and are very insecure about trying to say the words out loud. Using WM!R allows these students to hear their own voices and compares their pronunciation with that of the Panda. It helps them acquire confidence to use their developing English skills.

Special needs students also benefit from using WM!R. Students who exhibited symptoms of attention deficit disorder were more engaged in reading practice with WM!R than when reading on paper. One such student reported that he liked reading on the computer because it didn't take as long. Data indicated that he actually spent almost four times as long reading on the computer as he did in reading a printed version of a similar story. One teacher described a hearing-impaired student who was not fitted for a hearing aid until spring of the school year. WM!R was his best opportunity for getting feedback on his pronunciation because the volume could be adjusted so that he could hear well.

Learner-centered environments exhibit sensitivity to the individual cultural practices of students. The WM!R environment supports a wide range of dialects and languages. Alternate pronunciation of words that occur commonly in

some dialects are included in the acoustic models and accepted as correct. Instructions and feedback are available in Spanish for students in the bilingual class. Using the authoring system, aides were able to create instructions in Turkish for a student with limited English proficiency.

One important outcome of a learned-centered environment is a high level of motivation. Teachers reported that students were always motivated to use WM!R. Here are some of their comments: “I do feel that (my students) are more motivated to read while they’re on the computer, because … it’s very interactive, whereas when they’re having to read independently …they are more likely to be off task compared to how they act when they’re at the computer.” “They are very on task when they are there … And when I need them then I actually have to go over there. I mean I can’t call their names and have them turn to me. They’re very focused." “Basically they love the computers and they will go to that first if they are given a choice.”

Knowledge Centered

A knowledge-centered perspective focuses attention on what is to be taught and why it is important. This is based on an analysis of what students need to know and do at the end of instruction. Most teachers would agree that to read well students should have a knowledge of letters and sounds, be able recognize and know the meaning of a large number of words automatically, and to have strategic knowledge that facilitates comprehension. However, many might disagree about the best kind of instruction to accomplish these goals. In fact in recent years, there has been a great debate over whether a phonics or a whole language approach is *the* best way to teach reading. In reality, most teachers, including the teachers at our implementation site, incorporate aspects of each approach in their instruction.

With its stated goal of giving young children the experience of reading, WM!R adopts a somewhat holistic approach to reading instruction with learning occurring in the context of meaningful literacy activities. Word recognition is supported through the computer's pronunciation of individual words in the broader context of reading a book. Students use the recording tools to record answers to comprehension questions, summarize what they have read, and make predictions about what is about to happen.

WM!R does not provide decoding instruction. It does support children's discovery of patterns and rules through the use of books that focus on specific sounds.

Assessment Centered

Assessment-centered environments provide opportunities for students to make their thinking visible, receive feedback, and revise their performance. Teachers have limited time to assess students and technologies such as WM!R can help solve this problem.

WM!R's assessment goals are formative: The environment enables students and teachers to monitor progress toward the knowledge goals discussed earlier. Teachers reported that students benefit from having their voices recorded and played back. Most had no idea what they sounded like when they read and this feedback helped them improve their oral presentation skills. Teachers also indicated that listening and reflecting on their own performance encouraged students to read with expression and to practice to build fluency.

Diagnostic records kept by WM!R are available only to the teachers who reported they used them primarily to confirm their judgments about specific students. They also indicated that the diagnostic tool proved useful to print out individualized sight word lists for students to study. Although it would be possible to use these records for report card grades or other high-stakes assessments, the accuracy of the speech recognition and the informal settings in which the records are gathered makes this unwise (2).

Community Centered

New developments in the science of learning suggest that community norms and community connections play an important role in learning. One important aspect of a community-centered classroom is the extent to which there are opportunities for students to create products that are discussed and valued by others. Although reading to a computer, may seem like a solitary activity, the WM!R environment provides several opportunities for building community within and across classrooms. First, students can use WM!R to present a reading of a book with the words highlighted as the book is read in the student's voice. Teachers mentioned that students created these performances and showed them to

their classmates during a time set aside for sharing center activities. This motivated students to polish their reading skills.

Students and teachers also used the authoring component of WM!R to create stories for their classmates. In fact, it was seeing their teachers write and publish books in the WM!R environment that led to students' realization that they too could be authors.

Summary

Our purpose in this paper is twofold. First, we describe advances that make it possible to realize the potential of speech recognition to support emerging readers. These advances include improvements in the technology itself and in the design of WM!R, a tutor that uses these models to support young readers.

Second, we evaluate WM!R using a framework for thinking about issues of learning that are involved in any attempt to implement new technologies. In most discussions of technology implementations, the learning issues remain relatively tacit. By making them explicit, it is possible to assess their coherence as a system (see Brown & Campione, 1996) and their correspondence with what is known about human learning.

Endnotes

(1) Important pioneering work in this area has been done by Mostow and his colleagues at CMU (Mostow, et al., 1994). Their reading tutor is based upon adult voices and has been tested with older children.

(2) Interestingly, not a single teacher mentioned problems with speech recognition as a barrier to their students using the software. When teachers were specifically asked about failures in the speech recognition they indicated that the positive aspects of WM!R far outweighed occasional problems with recognition.

References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289-325). Mahwah, NJ: Erlbaum.

Mostow, J., Roth, S., Hauptmann, A.G., & Kane, M. (1994) A Prototype Reading Coach that Listens. *Proceedings of the Twelfth National Conference on Artificial Intelligence.* Menlo Park, CA. American Association of Artificial Intelligence.

Nix, D., Fairweather, P. and Adams, W. Speech Recognition, Children and Reading. *Human Factors in Computing Systems, 1998*. New York, NY: Association for Computing Machinery.