# A DEVELOPMENTAL STUDY OF ASR-ENHANCED E-BOOK SOFTWARE TO IMPROVE ON-TASK INTERACTION FOR FIRST GRADE USERS

by

Matthew William Gudenius

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Educational Technology

Boise State University

May 2012

BOISE STATE UNIVERSITY GRADUATE COLLEGE

# DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Matthew William Gudenius

Thesis Title: A Developmental Study of ASR-Enhanced E-Book Software to Improve On-Task Interaction for First Grade Users

Date of Final Oral Examination: 16 March 2012

The following individuals read and discussed the thesis submitted by student Matthew William Gudenius, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

| | |
|---|---|
| Ross Perkins, Ph.D. | Chair, Supervisory Committee |
| Kerry Lynn Rice, Ed.D. | Member, Supervisory Committee |
| Lee Ann Tysseling, Ph.D. | Member, Supervisory Committee |

The final reading approval of the thesis was granted by Ross Perkins, Ph.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

# DEDICATION

This is dedicated to Sarah, who suffered through endless nights and countless weekends (and innumerable voice recording sessions), yet somehow still managed to have the energy to provide the moral support and motivation I needed to keep moving forward.

ACKNOWLEDGEMENTS

This research is just one of many stepping stones on a long journey, a journey which has enjoyed the benefit of several helpful companions and likely will seek the support of many more along the way.

I would like to thank Dr. Ross Perkins, not only for continual constructive criticism as committee chair, but also as the first professor to usher me into the online learning experience at Boise State University, and to guide me several times along the way throughout the past few years. I would also like to thank Dr. Kerry Rice and Dr. Lee Tysseling for their high standards and honest feedback.

Lastly, I would like to acknowledge the various members of the school that graciously helped with this study: the teachers who provided expert advice and critical feedback; the secretary who assisted with translations; and most of all the wonderful children who voluntarily (and even eagerly) provided the best efforts and most thoughtful feedback they could muster to assist with this developmental research.

ABSTRACT

This developmental research study explores the effects of audio-visual feedback and user input mechanisms on user behaviors and satisfaction, through development of a first-grade reading program for the computer. Specific design elements investigated include human vs. synthesized audio feedback, segmented vs. whole-word pronunciation, format of supporting graphic (image vs. animation), use of automatic speech recognition (ASR) to encourage or enforce oral reading of an e-book, and effect of tutorial with mouse-click word identification or ASR-controlled word synthesis games. The study examines a variety of quantitative and qualitative measures including use logs, recorded screen-capture videos of use sessions, one-on-one interviews, and satisfaction surveys. The results of testing for each design element are analyzed and most appropriate design choice is implemented for subsequent design phases in an iterative manner. Design guidelines are given confirm the existing literature's findings of user preference for human speech over computer-synthesized speech (TTS) and that supporting graphics can increase user satisfaction of e-books, but also have the potential for distraction and reduction of active reading tasks. ASR was found to be ineffective as an input mechanism due to user error and low success rate in this study, but was found to be better-suited as a tool for smaller discrete tasks such as word synthesis practice and games, and may be effective for practice or support when offered as an optional tool to be used voluntarily by users. *(Keywords: literacy, reading, e-books, CAI, educational technology, elementary education, animation, synthesized speech, TTS, ASR, developmental research)*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1: THE RESEARCH PROBLEM

## Introduction

Proficiency in English literacy continues to fall below desired goals and expectations in the United States. At-risk groups including ethnic minorities and students coming from poverty continue to show the lowest performance on reading assessments. Numerous interventions over the past several decades have resulted in modest performance gains for these students, but a literacy divide still exists (National Center for Education Statistics, 2010b). Meanwhile, the number of American children aged 5 to 17 who spoke a language other than English at home increased from 3.8 to 10.9 million, or from 9 to 21 percent of the U.S. population, between 1979 and 2008 (National Center for Education Statistics, 2010a). Students of low socioeconomic status and non-English-speaking households have been found to enter school with lower than average oral vocabulary, putting them at a disadvantage in literacy acquisition. These facts call to attention the importance of early and effective intervention for reading, especially for at-risk populations.

## Statement of the Problem

Unfortunately, many schools and districts frequently face limitations of resources due to budget constraints which may prevent the use of many types of effective reading interventions. Targeted interventions requiring one-on-one or small group interactions with educators may not be possible when schools are faced with insufficient budgets or a lack of sufficiently-trained personnel. How, then, can literacy intervention be administered in a consistent and cost-effective manner?

One proposal for individualized intervention without the need for special training or hiring of additional personnel has been through the use of computer-aided instruction (CAI). For decades, CAI interventions have been researched as a viable alternative to traditional teacher-based instruction in support of building reading skills. These studies have focused on a variety of methods and literacy subtasks, but one of the most promising areas has been the use of CAI software for improving phonological awareness (Barker & Torgesen, 1995; Foster, Erickson, Foster, Brinkman, & Torgesen, 1994; Reitsma & Wesseling, 1998).

Unfortunately, there are drawbacks to the use of standard CAI for phonemic instruction. Several studies have shown a need for targeted intervention for students with low oral fluency (Chuang, Porfelli, & Algozzine, 2008) as well as listening comprehension and phonemic awareness (Juel, Griffith, & Gough, 1986). One of the noted strengths of modern CAI programs is their ability to build upon foundational skills such as listening comprehension (Karemaker, Pitchford, & O'Malley, 2010); unfortunately, listening comprehension is one of the specific skills shown to be insufficient in many at-risk students, a factor which may cause difficulties in the use of CAI by these students. In addition, many at-risk students exhibit low regulatory skills including inhibitory control, working memory, and cognitive flexibility which allow students to plan, organize, and apply rules. Students who lack these regulatory skills may become distracted and unable to effectively navigate the interfaces and activities of many CAI programs, possibly eliminating a primary benefit of CAI as being a self-directed and individualized activity (Foster, Erickson, Foster, Brinkman, & Torgesen, 1994; Kegel, van der Kooy-Hofland, & Bus, 2009; Prabhu, 2010).

There are several factors which indicate that, instead of CAI limited to specific skill-building, it may be more beneficial to embed reading skills into interactive e-books. One problem is that most CAI programs focus on a narrow subset of skills inherent in systematic reading instruction. Although the design and engineering of these programs may be "research-based", focusing only on small subsets of early literacy skills and excluding pragmatic application of those skills likely limits the efficacy of these programs since the majority of reading research indicates that it is best to combine phonics instruction with authentic reading tasks including fluently reading words in context of sentences and books (Ehri, Nunes, Willows, & Schuster, 2001b; Pufpaff, 2009; Sensenbaugh, 1996). E-books may also enhance CAI due to their ability to repeatedly expose the reader to written and oral vocabulary in multiple contexts, which is an important stepping-stone to reading comprehension (Hemphill & Tivnan, 2008; Silverman & Crandell, 2010).

Aside from the limitations of content and skill-building tasks, e-books and CAI software have also suffered from ineffective design of user input, navigation, and supporting features. In particular, interactive talking storybooks often include multimedia features including games and animations which actually distract from reading of the text or clicking on other features more supportive of actual reading tasks (de Jong & Bus, 2002; Labbo & Kuhn, 2000; Lefever-Davis & Pearman, 2005; Pearman & Chang, 2010; Trushell & Maitland, 2005).

These challenges seem to be exacerbated by a lack of ability to control or guide users of CAI and e-books. Several studies have shown that, despite the promise of individualized learning, CAI and e-books are more effective when utilized in peer-pairs (Mathes, Torgeson, & Allor, 2001; Shamir, Korat, & Barbi, 2008; Shamir, 2009) and most effective when supported by an adult (Korat, Segal-Drori, & Klien, 2009). This calls attention to a potential problem of user control or interaction with such e-books and CAI programs, the majority of which rely on manual interaction as the sole means of user input. Aside from the clear benefits of non-manual input for the physically disabled, Kegel et al. (2009) also noted that students with low regulatory skills suffer from uncontrolled and misguided mouse clicks. Behaviors that are frequently observed in e-books and CAI, due to open-ended user control and voluntarily-selected support options, include minimal interaction; under-accessing (not requesting help as needed); over-accessing (clicking for support when not necessary); reliance on the computer to identify unknown words rather than developing his/her own strategies; and having no mechanism for detecting and correcting errors unless the user specifically requests assistance (Lewin, 2000; McKenna, 1998).

Lefever-Davis and Pearman (2005) make reference to the drawbacks of this "voluntary support" mechanism traditionally found in e-books and liken it to "learned helplessness." They note that human supporters such as teachers are much better equipped to deal provide feedback in such a system of reading support:

> Teachers today are encouraged to prompt children's use of cueing systems when they encounter a difficult word rather than simply pronounce a word. However, computers lack the ability to make instructional decisions regarding what type of assistance is most valuable in any given situation; rather, the computer simply provides the word pronunciation upon request. (p. 448)

However, this is not necessarily always the case for computer feedback. Over the past several years, great strides have been made in advancement of automatic speech recognition (ASR) technology (Gerosa, Giuliani, Narayanan, & Potamianos, 2009; Hagen, Pellom, & Cole 2007; Hagen, Pellom, & Hacioglu, 2009; Hagen, Pellom, Van Vuuren, & Cole, 2004; Potamianos & Narayanan, 2003) and studies have shown its application to reading and learning tasks (Adams, 2006). However, Leitch and MacMillan (2003) acknowledged that the speed and accuracy of ASR technology is not yet capable of 100% natural and fluid speech and these limitations can reduce efficiency or increase frustration in reading or learning tasks. Additionally, the speech patterns of children and especially emergent readers pose even greater challenges to ASR accuracy (Gerosa et al., 2009; Hagen, et al., 2007; Hagen, et al., 2009; Hagen, et al., 2004; Potamianos & Narayanan, 2003).

Developmental research may provide valuable insight and research-based guidelines for the design and engineering of e-books and CAI software for early literacy instruction in primary grades. Developmental research provides a means for applying and testing existing theories in a pragmatic context, allowing development of practical tools while accepting the variability and flexibility of actual application in praxis. Design-based research involves an actively evolving environment and a participatory dialogue with its inhabitants, allowing for a pragmatic extension of theory to real-world application with simultaneous testing and refinement of that theory and its relevance to real-world practice (Brown, 1992; Collins, 1992; Collins, Joseph, & Bielaczyc, 2004; Reeves, Herrington, & Oliver, 2005; Wang & Hannafin, 2005). Developmental research builds on this design-based paradigm, but specifically as a guide for developing tools including incremental phases of end-user testing and feedback (Richey, Klein, & Nelson, 2003). In this way, theoretical expectations can be both tested and pragmatically applied in iterations, incorporating key formative input from the end-users themselves and the implications of applying such research and theories to technology design. In the case of CAI and interactive e-books, important aspects that will determine effectiveness of the software include content delivery, human-computer interface and interactions, and effects on motivation and time-on-task.

## Purpose of the Study

The purpose of this study is multi-fold:

1. To explore or confirm the findings of existing theories and research as applied to user satisfaction and on-task behavior in interactive e-books and CAI for reading.
2. To develop theory-grounded and research-based guidelines for development of an all-purpose CAI program in early reading for first grade students (users aged 5 to 7 years old).
3. Begin development of a prototype CAI software demonstrating the principles representing the framework/guidelines.

The outcome of this study will be an exploratory and descriptive framework for a first-grade-appropriate CAI software / e-book, developed concurrently with design of a prototypical example of such a CAI system. The focus of this study is on increasing efficiency of time-on-task for authentic reading of interactive e-book software, while maintaining an acceptable level of user satisfaction.

## Definition of Terms

The following terms will be used throughout the study and are based upon previous research, assessments, and specifications:

*At-risk readers.* "At-risk students" refers to those students in the United States who have an increased chance of academic success due to a variety of factors, including: welfare/low socioeconomic status; a primary language other than English spoken in the home; only one parent in the household; or low levels of education attained by parents (National Center for Education Statistics, 2000). Several factors have been shown to contribute to the level of risk for academic difficulties specifically in reading. One particularly important factor is amount of initial oral vocabulary (Hemphill & Tivnan, 2008; Silverman & Crandell, 2010).

*Automatic speech recognition (ASR).* Automatic speech recognition refers to the process and mechanism of recognizing spoken speech input via microphone. The specific hardware configurations and software algorithms used for ASR vary, and continue to evolve over time; studies mentioned herein have all referred to use of pattern-matching ASR which finds the best fit between a spoken audio waveform and an existing library of model or template waveforms for words and phrases. For the purposes of this study, ASR will use a microphone-into-PC hardware setup combined with Microsoft Windows Speech Recognition using Microsoft Speech Application Programming Interface (SAPI), an ASR program interface which is speaker-independent and uses Hidden Markov Models (HMM), a popular and efficient algorithm also utilized by Dragon Naturally Speaking and IBM Via Voice (Ayres, 2006). The ASR engine being used is Microsoft Speech Engine 6.1.

*Computer-aided instruction (CAI).* Computer-aided instruction refers to the use of computers to assist specifically in instruction/tutorial or guided practice activities for attaining targeted educational skills. For the purposes of this study, all CAI interventions will be referring specifically to software programs available on personal computers.

*Computer-assisted language learning (CALL).* Computer-assisted language learning (CALL) is a subset classification of CAI designed specifically to enhance acquisition of a language. Although CALL software tends to be used for acquiring a second language (L2), it is not limited to such use and may also refer to acquisition of a primary language (L1). For the purpose of this study, all CALL software will be included under the CAI label, and thus referred to as CAI.

*Developmental research*. Developmental research combines aspects of design-based research and instructional systems design. It is a systematic, iterative process of developing a specific product or tool – formative research" or "Type 1 research" – or with a focus on the study of the developmental and evaluative processes themselves – "Type 2 research" (Richey, et al., 2003). This study is primarily a Type 1 developmental study, with a goal of creating a reading software tool. However, elements of Type 2 research will also be included, such as reflection on the efficacy of evaluative tools used.

*Interactive e-books.* This study will focus on the use of interactive, multimedia-enhanced e-books for primary reading practice. These types of programs have been called by various names in the literature, including electronic talking books, talking storybooks, and CD-ROM storybooks. These types of programs all have common features including interactivity, static or animated visuals, and support mechanisms such as audio narration or spoken feedback.

*Phonemic awareness.* Phonemic awareness refers to the ability to distinguish and recognize the small sound chunks (phonemes) that form words. Phonemic awareness includes segmentation of words to distinguish individual phonemes, as well as blending these phonemes together to create words and word parts.

*Phonological awareness.* Phonological awareness is a more encompassing term than phonemic awareness and includes additional phonological tasks such as rhymes, alliteration, and syllabification. These skills include, and build directly upon, phonemic awareness. Although phonemic and phonological awareness are slightly different, literature sometimes uses the terms “phonemic awareness”, “phonological awareness”, and “phonics” interchangeably. This study maintains a focus on early readers and, as such, will primarily be concerned with the specific subset of skills known as phonemic awareness.

## Limitations and Delimitations

Despite the numerous practical advantages afforded by developmental research, there are also many limitations. This systematic, formative research involves collecting extensive amounts of both quantitative and qualitative data from multiple sources; conclusions are drawn from this data, but much of the data may be excluded from conclusions and results, and qualitative processes including interviews and observations lend themselves to possibility of through selective attention or focus (Wang & Hannafin, 2005).

Additionally, as Kelly (2004) points out, design-based research does not lend the same level of rigorous scientific validity that true experiments would hold. The participant group is small, leaving insufficient cause for generalization of results to a larger population. In addition, the notion of experimental control is consciously violated in the direct dialogue and interaction with participants. This flexibility allows for socially-responsible and practical adjustment and refinement, but at the cost of weakening any causal or relationship claims made pertaining to any of the variables (Kelly, 2004).

As Richey et al. (2003) note, a Type 1 study, focused on creating a tool or product, is inherently context-specific. This limited scope of specificity is also emphasized by the small number of participants (15 first grade students) in this study. It is especially important to note this contextual nature in regard to the specifications of the software tool being developed: conclusions found within this study may be specific to the hardware, programming implementation, and instructional and support content designed for this tool. It is possible, and perhaps even likely, that alternative implementations of the same concepts – with different audio-visual content, hardware configurations, or ASR-implementation – could provide different results than those discovered in this study. In addition, it is important to note that in the current software prototype, not all of the commonly-found features of reading CAI or e-books will be implemented or tested. For example, effect of sound effects and music will not be explored – in fact, these audio features will be excluded entirely.

A final but significant limitation of developmental (and design) research is that these are relatively new concepts, lacking the same rigor of methodology and terminology which have grown to be widely accepted in other forms of research (Wang & Hannafin, 2005). This could cause perceived reduction of validity or generalizability in the community, but may also cause misunderstanding due to specific terminology. As such, an attempt has been made throughout this study to adhere to use of clearly-defined “working words” as suggested by Kelly: "hypothesis," "conjecture," "observation," "model," "framework," "explanation," "belief," "rationale," "logical or argumentative structure," "perspective," "analysis,” or "synthesis,” rather than empirical concepts such as development of “theories” (Kelly, 2004, p.123).

As a developmental study, an effort is made herein to describe and explain both quantitative and qualitative data, but the complex nature of the tool being developed and the goals of development (to increase both time-on-task and motivation/satisfaction levels) mean that evidence may be contradictory or inconclusive. Instead, an effort is made to present the data and observations in an explanatory context with interpretations of implications for design and for future developmental studies of educational software for first grade students.

## Significance of the Study

This study will provide a descriptive exploration of the formative results of development and testing of first grade software for practice and enhancement of elementary reading skills. Over the past two decades, many studies have examined the use of CAI and interactive e-books for elementary school users. Results and conclusions have been mixed, often due to the unique context-specific combinations of content, interface, technology, and specific implementation (Ertem, 2011; Pearman & Chang, 2010). This study seeks to further illuminate the effects of two well-documented support mechanisms (visual illustration/animation and spoken audio pronunciations) on user satisfaction and time-on-task. An additional, and less well-documented, aspect of development this study will explore is the impact of using ASR-triggered navigation and reward to encourage increased time-on-task. Implications of these findings may show encouraging directions for use of ASR to increase focus on the task at hand and reduce passive use or learned helplessness in computer-based reading programs.

The target audience for this study includes teachers, administrators, and technology coordinators looking to make evidence-driven choices for software and hardware purchases in their schools that target literacy needs of children aged 5-7. An additional target audience for this study includes instructional designers and software engineers involved in developing CAI and e-books for early literacy. This study’s findings could provide important confirmations or improvements on the design and development of content delivery and control mechanisms for CAI and e-books. Additionally, future developmental researchers may benefit by learning from the described benefits and drawbacks encountered during the evaluative processes of this study.

# CHAPTER 2: REVIEW OF THE LITERATURE

## Introduction

The use of computer-aided instruction (CAI) for reading has existed and been researched for several decades. Design and development of CAI for language and literacy acquisition has been based on findings of numerous early literacy and language acquisition studies. Such studies have explored the importance of various factors in complex models of emergent literacy that distinguish and correlate multiple components of literacy including phonemic awareness, oral fluency, word recognition, vocabulary, and overall reading comprehension.

Multiple studies have shown that phonemic awareness is an essential building block in early literacy and a causal predictor of future skills such as word recognition and comprehension. Although there is little debate that phonemic awareness is an important building block for literacy, there is some debate (e.g., Fox & Routh, 1984; Perfetti, Beck, Bell, & Hughes, 1987; Silverman & Crandell, 2010; Torgesen, Morgan, & Davis, 1992) about the importance of phonemic awareness and its validity as a predictor of reading comprehension when isolated from vocabulary acquisition and word recognition, especially as students advance beyond second grade.

This review of literature focuses primarily on experimental and quasi-experimental quantitative studies published in peer-reviewed journals from the 1980s until the present. The studies were located electronically using databases including ERIC, Web of Science, Education Research Complete, PsycINFO, Academic Search Premiere, and JSTOR. These studies represent findings and advances in emergent reading and literacy theories, effects of CAI and e-book design elements including audio-visual aids and support features, and a review of advances, benefits, and complications of automatic speech recognition (ASR) as it pertains to educational technologies.

## Theoretical Foundations and Research Findings in Early Literacy

The acquisition of language, its processes, and potential methods of improving or accelerating the process have been studied for decades. By the mid-1980s, practitioners and researchers were aware that literacy acquisition was a complex and multi-faceted process comprised of multiple components or interrelated variables which combine, sometimes synergistically, to influence emergent reading and writing.

In 1997, the Director of the National Institute of Child Health and Human Development consulted with the U.S. Secretary of Education to convene a National Reading Panel of 14 individuals to review research on alphabetic, comprehension, fluency, teacher education, and technology. The alphabetic groups consisted of two meta-analyses, one of phonemic awareness and one of systematic phonics instruction. Correlational studies showed phonemic awareness and letter knowledge as the two best school-entry predictors of how well children will learn to read during the first two years of instruction. In addition, direct instruction in phonemic awareness had a large and statistically significant impact on acquiring these skills, and such systematic instruction in both phonemic awareness and phonics resulted in moderate impacts on reading, spelling, and comprehension (Ehri, et al., 2001b; Ehri, Nunes, Stahl, & Willows, 2001a). This finding has been supported by studies showing insignificant impact of print exposure as an isolated variable because there must be some basic level of phonemic proficiency before benefits can be attained from increased exposure to print (Juel, Griffiths, & Gough, 1986).

Many of these current conclusions about early literacy support elements of social constructivist theory as espoused by Vygotsky, in which social interaction with adults or more knowledgeable peers can accelerate cognitive growth through assisted exposure to new knowledge in a child's "Zone of Proximal Development" (Vygotsky, 1978). This Vygotskian theory supports the premise of guided sequential, systematic instruction of "stepping stone" skills toward literacy and is supported by studies that have shown spoken interaction with peers to improve subsyllabic phonological awareness and phonemic segmentation, real word reading, pseudoword reading, oral reading fluency, and comprehension (Mathes, Torgesen, & Allor, 2001; Shamir, 2009).

Juel et al. (1986) proposed a literacy model based on existing research and observations and included these primary factors: phonemic awareness, cipher knowledge, lexical knowledge, and word recognition. The strength of causal relationships between these factors – as well as external influences including ethnicity, IQ, initial oral language, and exposure to print – were tested using path analysis and multiple significant relationships were found. The most statistically significant factor was found to be "cipher knowledge" (sound-letter connection), which is influenced by both phonemic awareness and exposure to print. This study reinforced the concept that phonemic awareness is a critical first-step in reaching ultimate goals of both reading comprehension and writing proficiency, but also showed that exposure to print is an essential component.

The nature of phonemic awareness as an essential building block for word recognition has also been supported by other studies (Swank & Catts, 1994). Although the literature is in agreement that phonemic awareness is an important skill in the development of emergent literacy, there is some debate over which specific sub-skills of phonemic awareness have the greatest impact. In a longitudinal study of first grade students, Perfetti et al. (1987) examined the effect of phonological analysis (identification of individual phonemes) and phonological synthesis (creation of words by connecting phonemes together) and found synthesis to be the most significant predictor of reading skills. However, in two separate training studies, Fox and Routh (1984) and Torgesen, et al. (1992) found that interventions focusing on both analysis and synthesis skills improved performance more than analysis or synthesis training alone.

Another area of some debate is the long-term value of phonemic awareness versus explicit instruction of vocabulary in which students are simply immersed into vocabulary-rich contexts. Hemphill and Tivnan (2008) examined correlations of “inside-out” (phonemic and phonological) reading skills and “outside-in” (vocabulary and oral discourse) skills to reading comprehension assessment scores over time. The findings showed that, although phonemic awareness tends to be a valid predictor of reading ability through the first grade, assessments in later grades show that this predictability of initial phonemic awareness declines and that initial oral vocabulary actually tends to be a better long-term predictor of reading comprehension by the end of second grade and into third grade. This finding is especially important for at-risk populations who have been found to enter school with much lower initial oral vocabulary.

Acknowledging the importance of direct vocabulary and semantic instruction as early as prekindergarten, Silverman and Crandell (2010) studied practical implications of different methods of teaching vocabulary in the classroom. By creating separate experimental groups for students with low and high levels of initial vocabulary (IV), interventions could be differentiated to determine whether they were effective for low IV, high IV, or both subsets of students. Results showed that the most traditional method of explicit explanation/definition of new words increases comprehension for both low and high IV students, but with only moderate results. The findings show that more significant gains in both groups were found through varied and repeated use of vocabulary in multiple contexts. However, another important implication of this study is the fact that there are interventions which may not work as well for high IV learners, but are effective when targeted toward at-risk (low IV) students, including methods such as acting out and illustrating the vocabulary.

These studies show the importance of systematic instruction of phonological awareness (including phonemic awareness and phonics) is an invaluable component of early reading instruction, but that the best approach is to combine phonics instruction with synthesis tasks (oral speaking of words), vocabulary, and authentic reading opportunities to apply and transfer phonemic skills (Sensenbaugh, 1996; Ehri, et al., 2001b; Pufpaff, 2009).

## Social Constructivist Theory and CAI Design

Important theoretical foundations guiding the design of e-books and CAI include Vygotsky's "Zone of Proximal Development" in social constructivist theory (Vygotsky, 1978) and, especially, the central tenets of cognitive load theory (Shamir & Korat, 2006).

This theory originally specified that cognitive growth would be aided through social interaction with a “more knowledgeable other” (MKO), with the implication that this “other” would be another human being. This concept has been extended to computer technology as a possible MKO, for example through the use of spoken feedback and clickable support scaffolds.

With the improvement of recorded human speech as well as ever-improving synthesized speech, there has been a growing body of literature regarding “pedagogical agents” which serve as digital teachers and learning assistants in software. The roles of these agents “reinforce the notion that social interaction is central to learning with agents” (Veletsianos, 2009, p. 346). Even without such a virtual MKO, Reeves and Nass (1996) posit that humans treat computers and other media in fundamentally social and human-like ways, ascribing the same social rules to their interactions with media that they do with other humans. This psychological phenomenon has been dubbed “The Media Equation” (Reeves & Nass, 1996). This implies that use of CAI and e-books, especially when paired with audio-visual feedback and virtual agents or assistants, may be just as effective as a human MKO.

Some studies suggest the benefits of social interaction through peer assistance or paired learning. Mathes, et al. (2001) conducted a study using peer-assisted learning strategies (PALS) as an intervention for building literacy among first grade readers. The effectiveness of the PALS intervention on reading ability was measured using multiple established reading assessments and was compared to a control group receiving no special intervention as well as a second experimental group who received 8-10 hours of additional intervention via *DaisyQuest* CAI software. The study found significant improvements in literacy for all students participating in PALS, with or without additional CAI. However, the addition of CAI intervention did not result in significant improvement above PALS alone. The finding of this study shows the value of verbal interaction, but it is unclear exactly what factors involved in the oral discourse caused improvement in performance. It is possible that feedback or interaction from the peer helped to build knowledge from a peer-based MKO. On the other hand, improved performance could be due to the reader's actual oral reading (authentic phonemic synthesis of words) due to the presence of a listener. This suggests that CAI and e-books may be best designed to serve not only as support and feedback mechanism, but also as an aural listener.

Additional studies have focused on the effect of peer collaboration and have come to similar conclusions. For example, Shamir (2009) found that, when using e-books, collaborative activities including discussions, computer sharing, and verbal initiative all positively affected sub-syllabic phonological awareness. However, the study does not examine whether the benefit of discussions and verbal initiative come from human/social aspects; is it possible that verbal interaction with a computer would provide the same benefits? Answers to these questions would be invaluable in more precisely tailoring effective interventions, and could play a role in improving the efficacy of CAI.

## Cognitive Load Theory and CAI Design

Aside from Vygotskian social constructivist theory, a great deal of focus in software engineering and interface design for CAI software has been founded in cognitive load theory (CLT). Cognitive load theory posits that cognitive processing uses limited working memory as well as vast long-term memory (Hollender, Hofmann, Deneke, & Schmitz, 2010). Conscious cognitive activity (i.e., "mental processing") uses the limited capacity resources of working memory, although this capacity can be improved by using multiple sensory channels for input, since audio and visual input is processed separately and can be processed simultaneously. Additionally, information in long-term memory is stored and accessed through use of mental schemata, or knowledge/memory models. Learning occurs through construction of schemata that can be retrieved and used efficiently in working memory (Hollender et al., 2010).

Due to the importance of limited-capacity working memory used for learning, educational software design relies heavily on CLT principles regarding cognitive load. Three basic types of cognitive load are intrinsic, extrinsic, and germane. Intrinsic load occurs due to the complexity of new information inherent in the content itself. This type of load is important to consider in sequencing, pacing, and other aspects of instructional design. Extrinsic load can be caused by delivery medium or presentation style, such as spatial presentation, colors used, amount of information shown simultaneously, and presence or lack of distractions being simultaneously present. This type of cognitive load is a central concern of usability designers and software engineers. Germane cognitive load, like the others, can slow down processing and impede efficiency. This type of load is present when information is presented with variation or redundancy, such as through multiple examples or a variety of presentations. Although this can slow processing by causing the learner to process the information in multiple ways, it has also been found to improve schemata formation and thus cause the storage to be longer lasting and more readily accessible in long-term memory (Hollender et al., 2010).

This concept is supported by some core tenets of usability design. Robertson (1994) points out that some aspects of work-oriented principles (such as efficiency, i.e., quickness of task completion) may actually be antithetical to learning, in which slowing down and processing of information may actually result in better learning through schema-building and transferability of knowledge. Due to this, it may actually be preferable to introduce expanding complexity by making a program simple enough to enter but complex enough to provide growth (Robertson, 1994).

Based on the concept of germane load, rapidly completing a task is not always beneficial and may, in fact, indicate ignoring the task at hand or a lack of genuine mental processing, such as indicated by off-task behaviors in the use of CAI and e-books that allow completely open-ended user control (Kegel, et al. 2009; Lewin, 2000; McKenna, 1998). One critical factor to be weighed and balanced in the design of CAI is the level of learner control allowed. At one end of the spectrum, computer-controlled media have sequencing and pacing completely predetermined by the software; at the other end of the spectrum, learner-controlled systems allow the user the freedom to choose from variable choices including sequencing, content, and how that content is represented (Scheiter & Gerjets, 2007). Problems in allowing learner control may arise from usability mistakes causing disorientation, distraction, or cognitive overload, but may also be moderated by learner characteristics including prior knowledge and regulatory skills. Students with low prior knowledge or poor regulatory skills have been shown to benefit from more guidance, instructional support, and pre-structured paths within the software (Scheiter & Gerjets, 2007).

The level of computer control can also be considered with regard to media output, such as spoken instructions and praise. This, too, must be finely balanced. A study by Buckleitner (2006) underscored the importance of this balance. In this study, one group of children used CAI software with audio instructions and praise for correct answers, while the other group used identical software with audio instructions and praise turned off. The children receiving no audio praise or instructions were more active, attempting three times as many problems in the same amount of time, and also expressed greater motivation and enjoyment. On the other hand, children using the more structured interface with audio feedback showed decreased interest and activity, but had a higher percentage of correct answers (84.95% vs. 67.97%), implying the benefit of some level of structure and guidance as well as the important of avoiding excessive or unnecessary feedback.

With regard to spoken and synthesized speech, research shows that children's response latencies for synthesized speech were significantly longer than for natural speech, reinforcing the notions of the processing constraint model in which cognitive processing demands slow reaction time (Drager, Reichle, & Pinkoski, 2010). Veletsianos (2009) found that enhancing verbal expressiveness through pauses and inflections caused learners to score higher on post-task exams and to perceive pedagogical agents as being more interactive. These studies imply that using recorded human speech may be preferable to the use of synthesized digital speech. However, this topic warrants further investigation because recorded human speech requires additional development time and money, and recent advancements are causing synthesized speech to approach closer resemblance of natural human speech and inflection.

Another aspect to consider is the use of pedagogical agents, which are virtual characters embedded in CAI to serve as assistants or provide feedback. The use of human-like pedagogical agents can actually impede learning and focus due to distraction caused by movements and animation (Veletsianos, 2009). Although the purpose is to improve receptiveness due to human-like expressivity, the “media effect” explained by Reeves and Nass (1996) shows that this may be an unnecessary feature due to the fact that people attribute human and social qualities to media even in absence of such features.

## Usability and Human-Computer Interface Design

Cognitive load theory plays a large role in usability and human-computer interaction (HCI) design. According to the International Organization for Standardization, usability is defined as the extent to which a user can fulfill a task using a tool effectively, efficiently, and with satisfaction (ISO 9241-11, 1998). Factors that influence usability include learnability, efficiency, memorability, handling of user errors, and user satisfaction (Nielsen, 1993).These aspects are analyzed, tested, and incorporated into the engineering cycle. However, Robertson (1994) and Nelson, Bueno, and Hufstutler (1999) emphasize that, for educational software, these general principles and phases are not enough.

“Human Factors” methodologies can be incorporated into design of educational systems. The field of Human Factors is traditionally tied to work performance, with worker ease and efficiency being a central goal and with the process being user-oriented such that workers themselves play a direct role in feedback and design decisions (Robertson, 1994). This aligns with Nelson et al. (1999), who advocate that concern for usability should permeate every aspect of design and development, with four phases including:

1. Design with a concern for usability.
2. Develop with input from users.
3. Field test with a concern for usability.
4. Revise and maintain software with a concern for usability.

The benefit of user-centered and iterative design for optimizing usability underscores the value of developmental research. By monitoring and analyzing systems using both theoretical foundations to guide expectations as well as continuous feedback from authentic users in natural environments, the developmental process can test and refine theory as well as apply it to actual educational goals in a pragmatic way.

## Automatic Speech Recognition (ASR)

One technological advance that may alleviate off-task behaviors and distractions in CAI and e-books, as well as potentially bring the realm of traditional Vygotskian sociocultural practices to CAI, is the use of automatic speech recognition (ASR). Unfortunately, the effect of ASR on phonemic awareness, oral fluency, or comprehension is underrepresented in the research literature (Adams, 2006). This may be due to the fact that ASR technology has not been widely available or effective on personal computers until recently.

Further impacting the use of ASR for early-literacy software, there are confounding difficulties in use of ASR with young emergent readers due to variances in speech patterns. Children's speech results in ASR error rates two to five times worse than adult speech (Potamianos & Narayanan, 2003). Strommen and Frome (1993) observed that children demonstrate significant variability in their speech patterns, including extra words and premature phonations more often than adults. As expected, this increased variability reduces the number of correctly-matched patterns of the template-based ASR for children. Further analyzing and detailing aspects of children's speech, Lee, Potamianos, and Narayanan (1999) studied acoustic characteristics including pitch, formants, and duration of vowels and sentences, and concluded that children's speech is much more variable than adult speech. The implication of these studies is that special design considerations must be used to control for these additional variables when designing or modifying ASR software for use by children.

ASR methods and processing have continued to evolve and improve, and several of these advancements have focused on the challenges inherent in children's speech. Various demonstrated methods to improve accuracy of ASR recognition of children's speech have included: speaker-adaptive training (Hagen, et al, 2004); subword analysis to analyze syllables and phonemes and accommodate children's variable fluency (Hagen, et al, 2007); and using algorithms to shift pitch and formant qualities of adult speech to approximate an acceptable resemblance of children's speech (Hagen et al, 2009).

Li, Deng, Ju, and Acero (2008) specifically acknowledged and addressed issues concerning use of an Automatic Reading Tutorial system, specifically as delivered via hand-held, small form-factor devices for portability and ease of use. Their developmental study describes a reading system in which an interpolated N-gram path can be used to represent positive ASR matches as well as account for miscues, mispronunciation, deletion, and insertion of words common amongst young readers. This allows for highlighting and feedback targeting the point of difficulty in a sentence. Using such a system, combined with training the ASR grammar via a large corpus of recorded children's speech, resulted in performance on par with larger, more powerful desktop computing systems.

Despite the difficulties and challenges posed by using ASR with children, results of early studies have shown some positive benefits. Higgins and Raskind (2004) tested ASR-enabled software for 17 weeks with 28 students with learning disabilities and compared its effects on reading and spelling improvement with a control group of 16 students who did not receive the intervention. Users of ASR saw significant improvements in word recognition, comprehension, and phonological awareness skills, though spelling ability was not affected. In addition, in a field test of ASR-enabled *Reading Assistant* software, Adams (2006) found that 228 students in 11 classes, grades 2-5, showed significant gains in oral reading fluency over a control group of 182 students across 10 classes, even when using the software for as little as 131 minutes on average. The message is that surprisingly little authentic reading is needed to make a difference, and ASR can provide a tool to track what, when, and how well students have read as well as which segments provided difficulty and which types of supports were provided accordingly.

Some existing studies of ASR in educational technology have focused on its use in learning a second (L2) language. For example, Neri, Mich, Gerosa, and Giuliani (2008) studied the effectiveness of pronunciation training using computer-assisted pronunciation training (CAPT) versus traditional teacher-instructed pronunciation training, and found that both CAPT and teacher instruction resulted in similar significant improvement for 11-year-old Italian learners of English. Like many other CAI studies, this study does not imply that computer training is superior to traditional teacher instruction, but that it may be an effective alternative. Although this particular study did not focus on emergent literacy in a primary language, it supports the notion that pronunciation skills are strongly tied to phonemic awareness, a skill set essential to literacy acquisition.

## Efficacy of CAI and e-books

### Effects of CAI

Based on the demonstrated importance of phonemic awareness, whole-word vocabulary, and authentic reading practice for reaching the ultimate goal of comprehensive literacy, existing software and relevant CAI research has focused on the ability of computers to build both phonological awareness and vocabulary acquisition through tools such as direct tutorial, guided practice, and interactive scaffolding including interactive e-books with digitized speech and built-in dictionary tools.

Although CAI software designed to improve literacy has existed for several decades, most of the early programs required extensive teacher training, involvement, and administration of the software, making it difficult to accurately assess effectiveness due to variables such as teacher/administrator interest or training. One of the first programs to isolate these factors and allow independent student use through built-in tutorial was *DaisyQuest*, a multimedia program designed specifically to improve phonological awareness through tasks including rhyming, identification of initial phoneme, identification of middle phoneme, and identification of ending phoneme.

Foster et al. (1994) studied CAI-based phonological awareness instruction as a viable alternative to traditional classroom instruction and found that, using the CAI intervention, only 18% of experimental group failed to improve by at least two raw score points, while 69% of the control group did not improve by at least two points. Importantly, the effect size after these interventions totaling eight hours and five hours was as large, or larger than, two previous studies that found similar results for students using the same post-assessment after seven weeks of traditional teacher-led phonological awareness training. This study demonstrated that  results similar to classroom-based intervention can be effectively achieved using CAI, with the benefit of significantly less teacher/aide preparation and instruction time, as well as less variability of results based on teacher or classroom environment. Barker and Torgesen (1995) replicated the Foster et al. (1994) experiment with a focus on below-average readers using *DaisyQuest* and its sequel, *Daisy's Castle*. The results showed significant growth for specific phonemic awareness subtasks that had been taught using CAI, including segmentation and phoneme elision. Reitsma and Wesseling (1998), however, point out that by administering no intervention to the control group, these initial studies did not control for potential confounding variables such as training effect or use of computers itself playing a role on the outcomes. In their study, they improved upon the design and methodology by requiring all groups to engage in CAI intervention, with the manipulated variable being the use (or lack) of blending-specific tasks. This study compared CAI-with-blending to CAI-without-blending for groups of computer-only and computer-plus-teacher-support, and found that the most significant results arose from the group that received both targeted CAI combined with traditional teacher assistance. This suggests that, although

CAI has demonstrated the ability to improve phonological awareness, it may be best used as a supplemental intervention rather than replacing traditional instruction.

Effects of e-books

Although there is a strong focus on phonological awareness at the emergent literacy level, several studies have also examined the effects of e-books and interactive tools such as student-triggered dictionaries on demand. Karemaker et al. (2010) compared the results of e-books versus corresponding traditional "big books" for beginning readers. Both formats were produced by the same publisher and contained similar content including difficulty, vocabulary, and length. Both big books and e-books served to improve word recognition, with the authors noting a likely causal factor of repeated high-frequency words in both formats. However, some additional advantages became apparent in e-book format, particularly the built-in audio-visual feature of highlighting individual words or sentences accompanied with audio narration, which could reinforce the connection between orthography (word appearance) and phonology (word sounds).

Other authors point out that easily-accessible "interactive hotspots" – screen locations that can be clicked by the user to cause an action – for tools such as the aforementioned audio-visual support, and especially interactive dictionaries, can improve vocabulary recognition and comprehension (Li, 2010), as well as prove motivating to many reluctant readers (Oakley & Jay, 2008).

However, many studies have pointed out the complex nature of eBooks and the important impact design choices about content, user interface, and audio-visual supports have on their efficacy (Lefever-Davis & Pearman, 2005; Shamir & Korat, 2006; Roskos, Brueck, & Widman, 2009; Pearman & Chang, 2010; Ertem, 2011). Of particular concern, several researchers have noted that hotspots can distract from a focus on actual reading tasks, and that audio-visual supports can actually be detrimental instead of supportive if they are not well-tied to the text. McKenna (1998) showed that e-book features can inhibit development of intrinsic reading strategies and thus impede literacy development in beginning readers. Lefever-Davis and Pearman (2005) found that e-books have the potential to promote reading skill and that both struggling and strong readers benefitted from use of actual reading supports, but that an overabundance of hot spots leads users to adopt a “spectator stance” by wasting valuable academic time by instead focusing on finding and triggering visual hotspots, treating the book more like a game for mere entertainment. Perhaps the most important design consideration related to this concept is whether audio-visual supports (graphics and particularly animations and sounds) are “considerate” (integral and supportive of the story) or “inconsiderate” (incidental or incongruent with the story text.) Labbo and Kuhn (2000) showed that inconsiderate multimedia features led to passive reading and did not support understanding. This was further confirmed by Trushell and Maitland (2005), who found that allowing Year 4 and Year 5 (8-9 years old) students to access cued animations and sound effect features in an e-book resulted in detrimental effects to verbal recall of the story as well as inferential comprehension questions, compared to students who could not access these multimedia features at will.

The implication of multiple CAI and eBook studies is that they do have the potential to improve motivation as well as phonemic and authentic reading skills, but that design choices in development of these tools is critical, because content choices, user interface, and level of user control can allow for off-task distractions and mere entertainment in lieu of actual learning.

## Summary

Several studies have found early reading ability in kindergarten and first grade to be valuable predictors of reading fluency and comprehension in later grades. Early intervention in foundational reading skills via systematic phonemic instruction has been found to be particularly beneficial. This may be especially true for at-risk student populations including ethnic minorities, students of low socioeconomic status, and students from households speaking a different first language. However, some of the most effective intervention methods require resources such as additional personnel or specialized training. In addition to these resources which may not be available to all schools, there are limitations to effectiveness for students who may need individualized, differentiated instruction.

Computer-assisted instruction (CAI) provides an alternative to traditional early reading interventions. Use of CAI can be more economical, easier to manage, and provide more consistent results than teacher-led intervention while providing similar benefits. However, design in CAI for early reading has suffered from narrow focus on specific sub-skills and narrowly-focused quasi-experimental research.

Interactive e-books provide a way to extend these sub-skills into authentic reading practice via “exposure to print” and vocabulary exposure in multiple contexts. Well-designed e-books provide multimedia user support features such as animated graphics, audio feedback, and audio narration or pronunciation. However, such features can cause passive use and can actually detriment comprehension and literacy acquisition, especially when they are non-supportive of the text or are overused.

Human-computer interface design and usability studies, combined with recognition of cognitive load theory, can help to guide the process of effective design and development of CAI and e-books for reading. Literacy development should include phonemic analysis, sight recognition, *and* synthesis of words. These needs, combined with the user navigation problems and distractibility caused by low regulatory skills, suggest that ASR may be particularly useful as a new method of interaction and control. Although children's speech poses unique challenges and difficulties for ASR software, this relatively new technology may prove to reduce passive reading or off-task behaviors and to improve engagement and interactivity.

# CHAPTER 3: METHODOLOGY

## Research Design

This study employed developmental research as a method of pragmatically applying and testing research- and theory--grounded principles in situ. Multiple experimental and quasi-experimental studies have targeted specific features, elements, interfaces and effects of CAI software for reading. Although such studies purport to hold validity, their narrow focus and rigid experimental conditions also pose some drawbacks:

1. Unlike the physical sciences, educational sciences occur in “messy” environments with multiple variables. Attempting to create a “controlled” laboratory environment actually introduces a new variable (a controlled learning environment) that may not be present in real-world praxis, thus potentially weakening the pragmatic generalizability when attempting to apply the processes in real-world settings (Brown, 1992; Design-Based Research Collective, 2003).

2. Much like the narrow focus of the research studies, CAI software has been designed to narrowly focus on specific sub-skills rather than holistic programs that recognize the potential synergy of combined components. (Wang & Hannafin, 2005).

3. Reeves et al. (2005) have pointed out that, unlike physical sciences, social sciences (including learning and education sciences) should be socially responsible. That is, the end goal of such research is betterment of the human

> condition, and that condition may be affected by the research itself. Insofar as knowledge is being gained for knowledge sake, it is useless; that is, without a pragmatic application of findings to actually invent or improve education, educational research is moot. Even when research is valuable and relevant, it may take so long to piece together all of the narrowly-focused findings related to the myriad variables involved in a learning environment that instructional treatments cannot be modified fast enough.

Like action research and design-based research, developmental research modifies the learning environment and interacts with its inhabitants. However, action research focuses on theory-based solutions to a localized problem. Unlike action research, developmental research has the pragmatic goal of systematically creating a theoretically-sound and research-based solution, with the aid of ongoing testing, data collection, and formative assessment.

Establishing this study as developmental research means that several conditions were met:

1. Design and development of this prototype CAI/e-book tool for first grade readers was iterative and systematic, breaking down each phase to analyze specific variables and features that built upon previous design implementations.
2. Design was user-centered, incorporating observation, feedback, and interaction data of users in an authentic incorporating social cognitive theory, cognitive load theory, and dual coding theory.
3. The primary end product was a work-in-progress technology application (CAI/e-book for reading) based on data-driven design decisions.

4. A secondary product was guidelines for future design, development, or similar research and testing.

The design elements considered were:

1. User preferences for spoken audio: synthesized vs. recorded human; segmented vs. whole-word.
2. Effect of visuals (static, animated, or no graphic) on user satisfaction and time on task
3. Effect of ASR-control (optional, required, or no ASR) on user satisfaction and time on task
4. Effect of tutorial and practice (mouse-based identification or ASR-based synthesis) on oral fluency.

## Participants

The target population of this study was all end-users designed to benefit from the developed tool: first grade students (aged 5-7) learning early reading skills. Due to the in-depth nature of analysis involved in the research, this study utilized a focus group of 15 participants. Participants were selected from a population of 84 total students enrolled in 4 different first grade classes at a rural elementary school. All students were invited to participate in the study via a letter home to parents (Appendix A), and the resulting participant group consisted of 9 male students and 6 female students, ranging in age from 5.8 to 7.2 years old ($M$ = 6.4 years) at the commencement of the study (see Appendix L for timeline). Based on feedback from the Office of Human Subjects Research at Boise State

University, since this study focused on software development and used participant feedback only as a means to refine the product, I was not required to submit an application for Human Subject Research to the Institutional Review Board. The parental consent form was used as a formal invitation and to provide a detailed level of information that parents may have wanted to know. Translation of the forms into Spanish was done by the principal investigator and proofread by the bilingual school site secretary.

Three of the participants (all male) had been previously identified as advanced readers through scoring 70% or higher on a test of reading comprehension (Grade 2, Theme 6 Comprehension test developed by the Reading Lions Center in conjunction with the Sacramento County Office of Education.) Five of the students -- 2 male and 3 female -- were identified as English Language (EL) learners.

Prior knowledge and experience with the requisite computer hardware was assessed via an exposure / ability survey (Appendix B). This survey established that 100% of participants had previous experience using a laptop computer with mouth and headphone speakers. Two-thirds of participants also had access to a computer at home, but 60% indicated that they had never used a headset microphone, indicating a need for further training to build familiarity with that input device.

To help mitigate the effects of variables such as varying e-book content and sequence of test cases, these students were divided into three test groups, each of which would participate in all test cases for each phase, but with the variables applied to different content and in different sequential order. A conscious effort was made to create representative test groups by equally distributing EL learners, advanced learners, and gender amongst the groups.

## Data Collection

Like evaluation and design-based research, developmental research relies on a variety of research methods and instruments to obtain descriptive, ethnographic insights about multiple variables that comprise a system being used in its natural context (Collins, 1992; Reeves, et al., 2005; Richey, et al., 2003). Moreover, developmental research is user-focused, and usability evaluation includes testing and measuring a variety of aspects both qualitatively and quantitatively, using observations, interviews, questionnaires, expert reviews, and measures of actual use (Nelson, et al., 1999).

In this design study, participants used *DinoRead* in the natural context of a whole-class computer lab with up to 20 peers simultaneously using computers and microphone headsets. Each participant's interaction with the software was tracked and logged using a combination of data metrics programmed into the system for automatic storage in a log file (associated by each participant number) as well as screen-capture software to video-record each use session including mouse movements and clicks along with spoken audio. The software used for this task was a free screencasting program, BlueBerry Software's FlashBack Express. This software was tested prior to the study and chosen for its ease of use and its abilities to be scheduled to begin recording triggered by use of the *DinoRead* software and to run silently in the background without notifying or distracting the participant user.

In addition to the data logs (text files) and screen-capture videos for observation, participant satisfaction and feelings were solicited using a brief "child surveys"

(Robertson, 1994) at the end of each activity using a “happy face scale” (Appendix C), as well as short, scripted, audio-recorded interviews for each phase of development (Appendix D).

Development and usability testing also called for needs assessments of the subject-matter expert (SME) teacher to help guide development (Appendix E), alongside existing research studies and findings.

## Data Analysis

Data included a mixture of quantitative metrics and qualitative observational and interview data. Quantitative data included Likert-style survey questions for user satisfaction and data logging programmed into the *DinoRead* software to be saved as a log file identified by participant ID number. This data log was used to record use metrics including total time spent, time per page, time per word, words clicked, number of spoken word attempts, and the percentage of those attempts that were correct or incorrect. Quantitative metrics were evaluated to find the mean and standard deviation of each score, although these scores were designed to be supplemented and explained via use of the observation and interview tools, especially for cases of extreme outliers.

Although the quantitative data were designed to provide feedback about both time-on-task and user satisfaction, further information was needed to understand the factors affecting both of these dependent variables. Elucidation of the user process, hurdles encountered, and feelings occurred through use of automatically video-recording (screencasting) user interaction with the software, as well as post-activity interviews. For these qualitative measures, actions and interview responses were analyzed and categorized into common threads, with specific examples noted to exemplify important themes.

## Procedures

This developmental study followed an ongoing iterative process of software development over the course of four months. Development and testing focused on theory-driven aspects and empirically-guided design over several phases. Based upon the key design considerations outlined in the current literature, four key factors were investigated while developing the initial prototype of the *DinoRead* software. First, needs were determined based on indications of existing literature as well as an survey of perceived reading software needs completed by first-grade teachers at a rural elementary school. Based on these assessments, the aspects of considerate design of holistic reading software to encourage on-task reading included:

1. format of spoken audio feedback;
2. design of support graphics/visuals;
3. potential use of ASR to encourage word synthesis and authentic reading;
4. and the incorporation of tutorials and practice games to enhance phonemic skills including identification and synthesis.

Multimedia (audio-visual) feedback and design have been well-documented in the literature, but should be confirmed for the target audience. Readily-available ASR input is a newer technology and thus less well-documented in the existing literature, despite its potential utility in literacy acquisition. This calls out the importance of testing this aspect of software design (Adams, 2006), although it is important to note that there are many possible methods of implementing ASR including a variety of hardware/software permutations.

Each design aspect called for an iterative process of development in multiple phases. Testing of each phase was enabled by assigning each participant a participant ID number (1-15) which was then used as the identifier on all surveys, interview transcripts, and log and video files. This participant ID was also used to automatically assign appropriate test cases within the *DinoRead* software. Any personally-identifying information about the participants was stored in a separate file as a cross-reference to participant ID numbers.

Phases of Development

Development of each software component was guided by the literature-supported principles and developed using systematic phases:

**Phase One**: Determine needs based on review of literature, existing studies and guidelines, and subject-matter expert (SME) feedback.

**Phase Two**: Design solution based on theoretical and research-based foundations as well as SME feedback, keeping users and learning context in mind.

**Phase Three**: Test solution including user feedback interviews, satisfaction surveys, and recorded screencasting videos for observations of behavioral trends and patterns.

**Phase Four**: Use data and feedback to revise component before moving on to next component.

Development of the *DinoRead* Software Application

To ensure that testing of each design aspect could occur rapidly in succession after evaluating the previous design variable, the *DinoRead* software was developed before testing any design cases (see Figure 1). Development occurred over the course of two months and consisted of creating an Adobe Flash desktop application for Windows as the e-book platform, as well as concurrently creating the story texts and multimedia content required for each e-book. The software was programmed to allow for configuration in a variety of permutations by using configuration variables that could be switched on or off to allow for the results of test cases to be applied quickly and easily to the software to establish a given design paradigm for each aspect based on the results of testing in each phase. Thus the software developed was not static but was dynamically capable of switching on different test case conditions as a result of of/off flags set for graphics, ASR, and lessons in the configuration XML file.

The *DinoRead* software was created using Adobe Flash Builder 4.0 (Flex framework) for development of rich internet applications and desktop apps. The *DinoRead* software was designed and developed as a stand-alone desktop app using Adobe AIR (Adobe Integrated Runtime) architecture to allow it to be installed on various operating systems. The basis for the interactive book component is the Flex Large Book component by Miti Pricope (http://miti.pricope.com/2008/09/10/flex-large-book-component/), which is an extension of the Flex Book component by Ruben Swieringa (http://www.rubenswieringa.com/blog/flex-book-component-beta).

As developer of the *DinoRead* software and content, I created an additional application, *DinoRead Creator*, to facilitate automatic conversion of any text into an interactive e-book format. This application allows the user to choose which text would go onto each page, and parses the text into XML files delineating pages, sentences, and individual words with the tags <p></p>, <s></s>, and <w></w>. These files are stored in a special folder for the new book, and the book's information – including title, author, cover image, number of pages, and number of words – are added to a library file to allow the user to choose e-books within *DinoRead*. The *DinoRead Creator* also simultaneously created specific grammars for ASR, specified to listen for the whole sentence, any sub-portion of the sentence commencing from the first word, and for individual words or word pairs if speaking terminated before the end of the sentence.

The ASR grammars were written as scripts (.pie format) for GlovePIE (http://sites.google.com/site/carlkenner/glovepie), an alternative interface emulator for computers which also incorporates access to Microsoft's built-in Speech Application Programming Interface (SAPI) which is included in all current versions of Windows since Windows XP. The Microsoft SAPI allows access to the Microsoft speech recognition engine as well as the voice synthesis engine for text-to-speech functionality. For this study, the netbooks used were running Windows XP with Microsoft Speech Engine version 6.1. This engine allows for "user independent" recognition, but the speech recognition can be improved through user-specific training.

Unfortunately, the training tasks require advanced reading abilities beyond the level of most first grade students, and would be logistically difficult or impossible to incorporate in a classroom of computer resources shared amongst many students. Although the speech engine is deemed to be user-independent, it is initially trained for highest success rate with adult male speakers. Unfortunately, adult male speaker speech patterns are the least consistent with the pitch and formants of young children (Hagen, et al., 2009). Due to these constraints, the speech recognition default user profile was trained using an adult female voice, which has been shown to more closely align with the characteristics of children's speech (Hagen et al., 2009).

Upon running, the *DinoRead* software displays a library of e-book selections based on the e-books created and available. When the user clicks on an e-book, that book's demarcated XML story text is loaded and formatted into individual, clickable large-font words. These words are assigned functions to turn blue upon mouse-over, and to load and play the corresponding pronunciation file when clicked. Along with loading the words for each odd-numbered page of the book, each even-numbered page is pre-loaded with the appropriate type of graphic (static, animation, or no graphic) based on test case for the current user and selected book.

When the *DinoRead* program is run, each participant enters his or her ID number to enter the program. For ease of use, participant ID number was based on assigned seat and each user had a written version of the number for reference. The test case for each book was determined using a modulo function on participant number (numbers 1-15) + book number (numbers 0-2) to automatically set a test case of 0, 1, or 2. Depending on which design aspect was being tested (determined by variable flags in the library/configuration file), the appropriate test condition would automatically be applied for each book.

The *DinoRead* software was also programmed to record metrics including time spent on each page, total time spent, and time spent per word. These data items were automatically logged into a text file in a remote network folder, along with the results of a "happy face questionnaire" item asking, in general terms, how much the user enjoyed the activity. This survey included faces to represent a Likert-scale of satisfaction, supported by audio statements upon mouse-over (Appendix C), and upon mouse click closes the current e-book.

To accurately determine what behaviors each user was exhibiting, BlueBerry FlashBack Express screencasting software was installed and running on each Acer Aspire One netbook computer. This screencasting software was configured to automatically commence recording screen and mouse activity, along with microphone input, as a digital video file when the *DinoRead* software was run, and to terminate and save the recording upon exiting the *DinoRead* software. The purpose of this recording was to allow for observation of user behaviors and interactions at a later date and categorize those behaviors into important themes and trends. See Figure 2 for an overview of the *DinoRead* software navigation and processes.

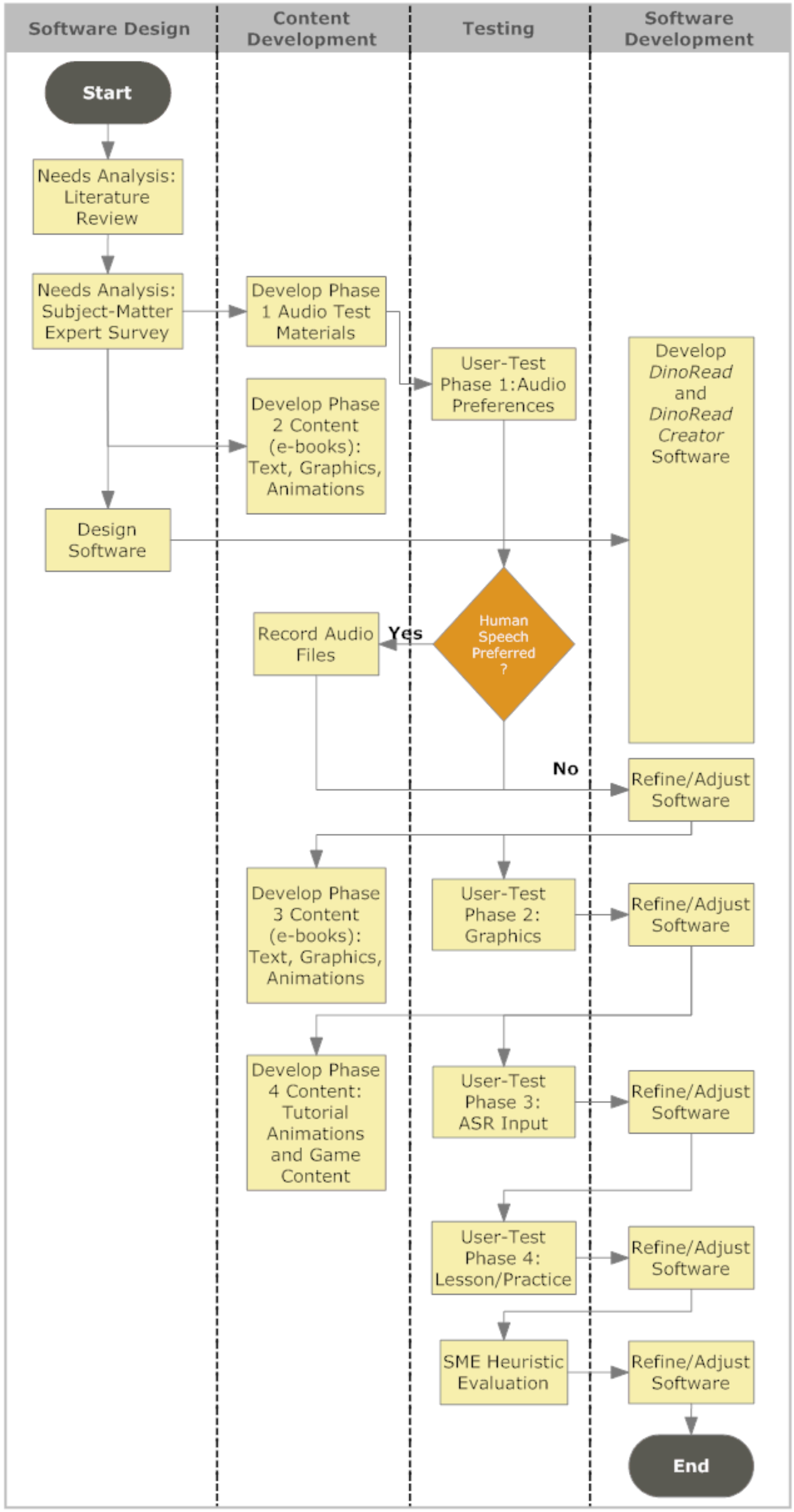



**Figure 1.** ***DinoRead* design, development, and testing sequence.**

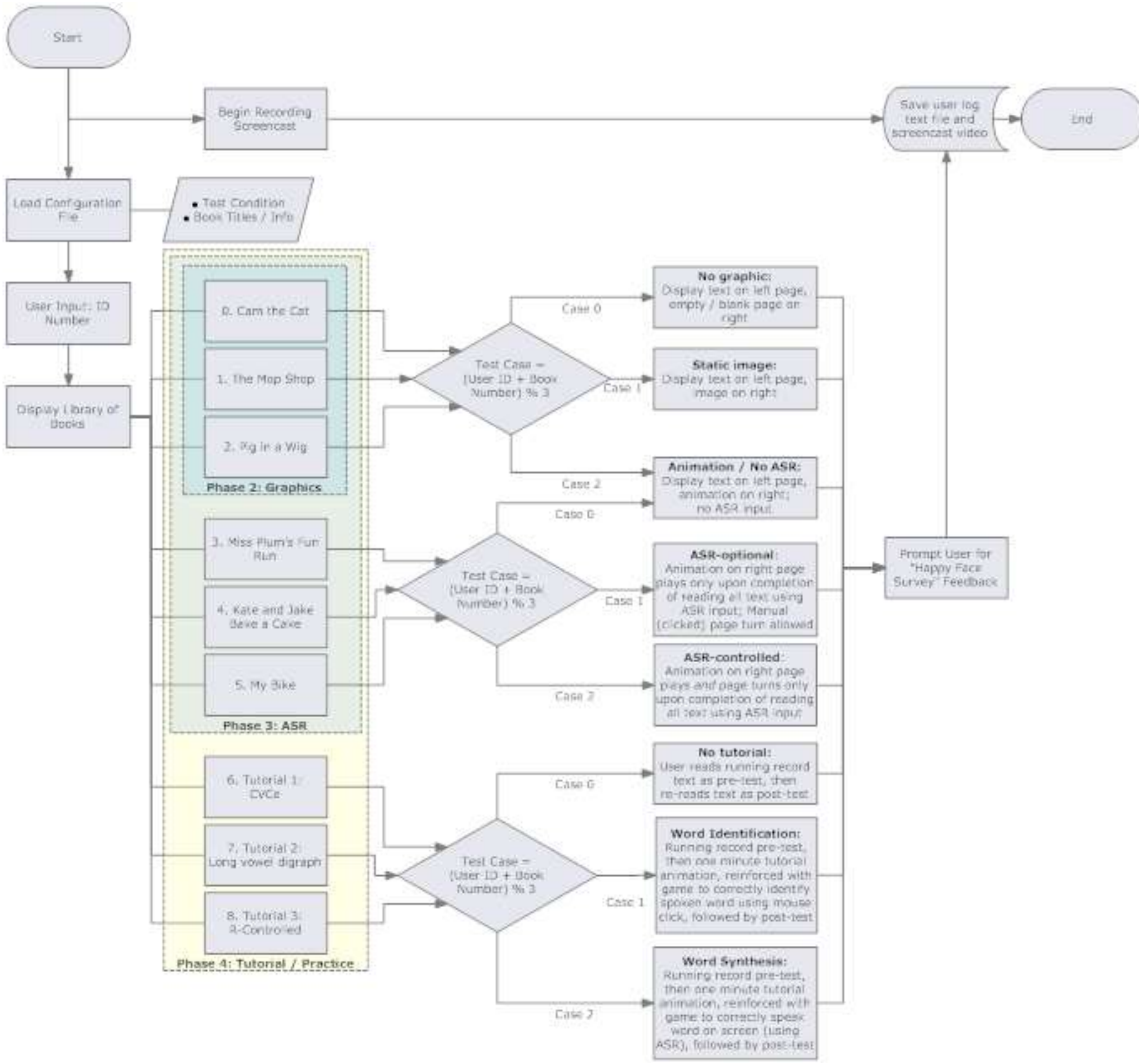



**Figure 2. Navigation and processes of *DinoRead* software.**

In addition to the quantitative metrics and observable user behaviors, participants further explained their internal feelings and feedback via post-activity one-on-one interviews, recorded and transcribed digitally into the same secure computer log folder.

Nine separate e-books were developed to cover three test cases in each of the three authentic design aspect test phases (graphic/visuals, ASR input, and phonemic tutorial/practice). The e-book story content of the first test case (graphics) was designed to be relatively easy to use as a source of practice of Theme 1 skills outlined in the 1st Grade Houghton-Mifflin Reading curriculum for California (which includes books such as *The Cat Sat, The Mat, Nan and Fan, We Can!, The Big Hit,* and *Big Pig.*) Based on the same consonant-vowel-consonant phonemic skills and a similar base of high-frequency words, three e-book stories were written: *Cam the Cat*, *The Mop Shop,* and *Pig in a Wig* (Appendix G). Additional stories for the ASR and tutorial test cases were designed to likewise correlate to Houghton-Mifflin reading curriculum pacing concurrently or imminently arising for students in their regular class instruction.

Development of considerate audio-visual features occurred in conjunction with these written texts – designed to be read sequentially with slight an increase in difficulty and word count. Each story was broken down into eight pages of text and an animation was designed to illustrate the character, setting, and actions described by the text on that page. Software and content design heeded many of the research-supported guidelines put forth by Seo and Woo (2010), including:

- Colors were high-contrast and carefully selected; the book format followed a traditional black-on-white text paradigm, color changes for text were used only for important events such as noting clicked pronunciation, or recognizing correct or incorrect user pronunciation via ASR.

- Children with low regulatory skills including inhibitory control, working memory, and cognitive flexibility are less proficient in applying rules, are easily distracted and impulsive, and have problems dealing with changing tasks (Kegel et al., 2009). For these children, it is particularly important to make instructions explicit, concise, and easy to understand with included examples, to keep task transitions minimal and unambiguous, and to avoid potential distractions such as excessive graphics and cued animation or sounds (Seo & Woo, 2010).

- Learning-disabled students have a tendency to attribute success and failure to luck or to the ease of a task rather than their own abilities. Because of this, students should receive ability attribution feedback acknowledging their efforts and successes (Seo & Woo, 2010). This feature was incorporated visual color change to green words for correct ASR pronunciations of words, and was done with both audio and visual feedback rewards for correct answers in both the mouse-clicked and ASR-triggered tutorial practice games.

Spoken Audio Format

Dual coding theory (a subset of cognitive load theory) states that cognitive load can be effectively reduced by utilizing both audio and visual channels of processing. Multiple studies have confirmed that using spoken feedback, instructions, and support mechanisms (such as clickable words for pronunciation) provides beneficial results in CAI and e-books. The studies that have compared recorded human speech and synthetic digitized speech produced by the computer have found that students show a preference for, and better response to, actual human speech moreso than synthesized speech (Lundberg, 1995; McKenna, 1998; Segers & Verhoeven, 2004; Shamir & Korat, 2006; Torgesen et al, 1992; Trushell & Maitland, 2005). It seems apparent that recorded human speech would, therefore, be preferable to using a synthetic digitized speech system. However, synthesized speech systems have continued to evolve and improve over time, with advancements in natural intonation and inflection. This is an important aspect to explore for design and development purposes, since the use of synthesized speech can save valuable time and money resources and promises to be more scalable than employing voice talent for discrete recorded passages.

Another important consideration for clickable pronunciation in e-books is whether it is preferable to use whole-word or segmented phonetic pronunciation to assist the user. If segmented speech improves phonemic awareness, it may be a worthwhile feature. However, it increases the complexity and time spent in development, and could prove detrimental by slowing down the reading process or decreasing motivation to use the support feature.

The focus of this study is on increasing efficiency of time-on-task for authentic reading of interactive e-book software, while maintaining an acceptable level of user satisfaction. As such, the effect of segmented audio on reading abilities was not addressed; instead, the study focused on user preferences and feelings about different formats of audio feedback.

To determine these preferences, participants were given a feedback sheet containing 16 boxes of words and the letters A and B underneath each word (see Appendix F). For the first 8 words on the front of the sheet, the spoken directions for the exercise stated "For each number, you will hear 2 sounds: sound A and sound B. Circle the letter of the sound *you like more*." Each number was announced, followed by "Sound A" and then an audio clip via speaker, then "Sound B" followed by a second audio clip for comparison. The final 8 words, on the back side of the spoken audio feedback form, used the same format (with different word samples) but slightly changed the connotation of purpose through the statement "Circle the letter of the sound *that is more helpful*."

In half of the cases, one sound clip was a recording of an adult female voice pronouncing the word (recorded as 44Kbps mp3 format) while the other clip was a synthesized pronunciation of the same word using the most commonly-accessible synthesized speech voice, Microsoft Anna (normal speed). Half of these cases used one-syllable words while the other half used two-syllable words to test whether length/duration of the word pronounced had an effect.

The other half of test cases prompted participants to select a preference between a word pronounced alone versus the same word segmented and pronounced as phonemes before being pronounced as a whole word. Again, these cases were divided evenly into one-syllable and two-syllable examples.

To prevent against a possible bias due to participants simply repeatedly selecting one letter or the other, the order of audio types was switched for each iteration of the test case; i.e., if the first example provided synthesized speech as Sound A, then human speech was used as Sound A in the next word example comparing these two types. Additionally, the pattern was further broken up by staggering prompts comparing human vs. computer speech and prompts comparing whole-word and segmented pronunciation (always recorded human speech in these cases).

Results of the circled choices were tabulated in a spreadsheet and calculated to determine what percentage of users preferred each of the 8 possible format choices: human (single- and double-syllable) versus computer (single- and double-syllable); and human whole-word pronunciation (single- and double-syllable) versus human segmented pronunciation followed by whole-word (single- and double-syllable).

To further illuminate why users showed a preference for synthesized or recorded human, or for segmented or non-segmented pronunciation, a short scripted one-on-one interview (see Appendix D) was conducted with each participant, in which the multi-word phrase “birds don’t have fur” was played via headphones, first as synthesized speech and then as recorded human audio clip. Next, the word “that” was played, first as whole-word pronunciation and secondly as segmented followed by whole word. The participants were asked which audio clip they preferred in these two examples, and to explain any reasons why they preferred the specified audio clips.

Format of Graphics / Visuals

As previously discussed, the use of audio-visual supports can be beneficial to motivation and comprehension when using interactive e-books, but can also prove to be detrimental by posing distractions if poorly designed (de Jong & Bus, 2002; Labbo & Kuhn, 2000; Lefever-Davis & Pearman, 2005; Pearman & Chang, 2010; Trushell & Maitland., 2005). Although it has been shown to be very important to design “considerate” support visuals and animations that support the text and avoid “inconsiderate”, or off-topic, features (Trushell et al., 2005), it is worth testing the effect of visual type (animation vs. static graphic, or no image at all) on a user’s on-task behaviors, efficiency, and satisfaction. The value of exploring this concept is multi-fold:

1. Creation of graphics requires time and resources to develop, and development of animation is even more complex and resource-intensive;
2. If visuals distract the user to take up too much time, it may be academically preferable to avoid that type of visual feature;

3. However, it is important to also consider to impact of visuals on user satisfaction. If the user is dissatisfied with the experience or unmotivated to use the software, it will likely be ineffective as a support tool (Oakley & Jay, 2008).

To control for the influences of confounding variables including story content, visual content, and sequence of reading, visual test cases were designed to be staggered such that each participant read all three stories but would receive a different visual format depending on test group. To allow for this design, each e-book needed to have both a static image – developed in Adobe Flash CS5 and exported as a JPEG image – as well as an animated version, which was illustrated and animated using Adobe Flash CS5 and exported as a .SWF (Flash) file. Both images and animations used a resolution of 400 pixels width and 400 pixels height. To ensure that duration of animation did not have an effect, all animations were 3 seconds (45 frames at 15 frames per second) in duration. The time of development of these different formats underscores the value in testing their efficacy: static illustrations required 10-20 minutes to complete, while each corresponding animation took 30-60 minutes to complete.

In addition to the graphics, based on the strong preference for whole-word pronunciations using authentic human speech, a corpus of audio clips were recorded to cover all words in the three texts included in this phase, to be triggered upon mouse-click of the corresponding word in the e-book.

For this first authentic-use test case, participants were first taught how to wear the Logitech USB microphone headset and to ensure that it is plugged into the computer and recording audio. For this segment of development, the participants were given the instructions: “You will be reading three different stories. One of them will have pictures that move, one of them will have pictures that don't move, and one of them will have no pictures at all. Do your best to read the words out loud. If you need help, you can click on a word to hear it. When you finish the story, choose the face that represents how you felt about the story, then choose the next story in line. Start with the first book on the left, called *Cam the Cat,"* accompanied by a projected demonstration of how to start the software and select the book.

For this test case, the three conditions automatically assigned to the test case numbers were as follows: Test case Graphics-0 resulted in only text being displayed, with no accompanying visual supports. Test case Graphics-1 resulted in static image use, while test case Graphics-2 would embed the animation file for that page. Animations would automatically be triggered to play as soon as the page was turned.

As previously noted, the *DinoRead* software was configured to capture user-specific metrics in a log file for each book. For this design phase, the log file included total time spent, time per page, time per word and user satisfaction as indicated by clicking on the "happy face survey" item at the end of the book. These items were analyzed, along with a post-activity interview to glean information about user feelings and processes, and a meticulous review of recorded screencast videos to determine the effects of graphics and animations on user behaviors including word clicks, spoken word attempts, and skipped pages.

<u>ASR Input</u>

Traditionally, CAI software has utilized only the mouse and keyboard for student input. Yet authentic practice and assessment of early readers requires children to read words aloud verbally. Is it enough to prompt users to do this and expect them to do so, or could the process be enhanced by automatic speech recognition (ASR)? According to the theory of germane cognitive load, ASR technology may enforce reading e-books aloud, which could cause the task to take longer but would result in improved long-term skill acquisition and post-task performance. What effect might ASR have on user satisfaction? I hypothesize that slowing down the process and enforcing authentic reading activity rather than allowing the student to access cued animations and sounds on demand may reduce the user satisfaction level of the e-book.

Testing of this design scenario followed the same format as the previous design testing for graphics: three new e-books were introduced to students, corresponding to phonemic concepts of Houghton-Mifflin Themes 2-5, including silent-e long-vowel words and short-u with consonant blends (Appendix H). Content for these books included animations as suggested by the findings from the previous test phase. However, the precise implementation of animation varied as a function of potential use of ASR.

The three test cases for ASR were as follows:

1. Test case ASR-0 had no ASR functionality. User could turn the page as desired, and animation would trigger on page turn. This test case was identical to case Graphics-2, although used a new (previously unread) e-book for testing.

2. Test case ASR-1 allowed user to speak and ASR would highlight the words in green if spoken correctly, in orange if spoken correctly after one mispronunciation, or in red if ASR did not recognize the pronunciation. User could turn the page at will, and the when the page was turned the new page's animation would begin to play.

3. In test case ASR-3, user navigation was strictly controlled using ASR, requiring the participant to attempt to read all words on the page. Correct pronunciation recognized by ASR would be marked green if correct on first try, orange if a second try was needed, and red if mispronounced three times in a row. Animation would play and the page would automatically turn *only* once all words were attempted and changed color.

Like the previous design phase, the ASR test phase automatically logged metrics of use including time per book, time per page, and time per word, as well as words clicked, number of successfully-spoken words, and number of ASR-identified errors. In addition, user opinions and perspectives were obtained via post-activity one-on-one interviews, and recorded screencasts were reviewed for the same behavioral patterns as already noted (under-utilization, over-utilization, spoken word attempts) in addition to analysis of the success rate of the ASR in correctly identifying spoken input.

Tutorial and Practice

Most CAI software currently available falls into one of two camps: skill-building and rote practice, or authentic application. In early literacy, this takes the shape of software designed to build phonemic awareness versus software that functions as e-books. According to literacy research and cognitive theory, stronger schemata are built when more complex tasks are presented, such as authentic task performance; low-complexity tasks would be rote knowledge / drill and practice (Ehri, et al., 2001a; Ehri, et al., 2001b; Hollender, et al., 2010). However, according to Vygotsky, these cognitive connections are only likely to be made if the student is working in his or her ZPD along with guidance from a “more knowledgeable other.” Perhaps a more ideal solution would be phonemic tutorials immediately followed by authentic practice using e-books scaffolded with multi-modal tools as described within this study.

The design consideration in this test phase was conducted to determine: (a) what effect a phonemic tutorial with practice drills may have on fluency; (b) whether word identification or word synthesis may be preferable for building said fluency. Three animated audio-visual tutorials, ranging in length from 35 to 40 seconds runtime, were developed to test these relationships. The three tutorials each focused on a specific grade-appropriate phonemic awareness skill: silent-e long-vowel (cvce) words; long vowel digraphs involving a (ai, oa, ea); and r-controlled ("bossy r" such as ur, ir, er) words.

Each tutorial featured an animal character representing the important alphabetic letter/sound involved in the phonemic task, and a human female narrator explained the skill along with anecdotal mnemonics for the skill accompanied by visual examples (Appendix I). As soon as each animated lesson is finished, the user is prompted to practice with 10 interactive examples in a game format (Appendix J), in which 10 points are given for each correct answer. Which of the two formats was presented was determined by test case number.

The word identification game requires the user to distinguish between two similar words printed on the screen, via spoken audio prompt stating "Click on the word that says…" and followed by the spoken word. The second format of practice was a word synthesis game in which one word is randomly selected from the practice word bank and shown on the screen, with the prompt "Say the word you see on the screen." In both formats, the user is rewarded with 10 points and a visual picture representing the word if they get the word on the first try, but the user is prompted to try again if first attempt is incorrect. To prevent an impasse in the ASR case, only 3 attempts are allowed per word before the game moves on to the next word.

The effect of the tutorial and practice on reading fluency was measured via running record, in which the participant read the text on the screen to the best of their ability, recorded by the video recording software. During review and analysis of the video, the investigator noted skipped words, miscues, and incorrect readings, and finally calculated how many words were read correctly in that time span. Running record texts were specifically designed for this activity to include a high frequency (~20%) of the specified phonemic task (Appendix K). The same running record was conducted as a pre-test prior to the lesson and practice, and again immediately following the intervention . To control for the expected beneficial effect of reading the text a second time, the control did not receive an intervention but did perform the same running record two times in a row to establish a baseline level of expected improvement from pre-test to post-test.

In test case Tutorial-0, only static (non-clickable) text was presented to the participant and the participant's reading of this text was recorded and analyzed. Following a reading of the passage, the participant was prompted to immediately read the passage a second time, with no phonemic lesson or practice being provided.

In test case Tutorial-1, the participant was prompted to read the running record text and then click a green arrow button to trigger the animated tutorial related to the phonemic task in that text was shown. This was immediately followed by the mouse-click word identification practice game. Upon completion of 10 practice exercises, the participant was prompted to read the original running record text a second time.

Test case Tutorial-2 operated identical to test case Tutorial-1, except that the practice exercise game consisted of completing each exercise by speaking a word shown on the screen and moving forward to the next word if ASR correctly identified the spoken input.

# CHAPTER 4: FINDINGS

## Introduction

The findings of this study are organized by design feature. The testing and results of each design and development phase will be discussed in chronological order, with an analysis of all quantitative and qualitative data gathered in each testing phase and their implications on design implementation to be incorporated into the subsequent design phase. In addition, findings regarding successes, challenges, and lessons learned about the development study process itself will be shared to provide guidance in future developmental research.

## Spoken Audio Format

Design of spoken audio feedback and clickable pronunciation of words was centered on: (a) whether the feedback should be computer-synthesized or recorded human speech; (b) whether pronunciations would be preferable as whole-word pronunciations or would benefit from phonemic segmentation. This test phase included a preference survey activity (Appendix D) as well as one-on-one interviews (Appendix F). Participants were prompted to listen to a series of sequential audio clips, played in pairs, representing the same word in two ways – computer vs. human, or human segmented vs. human-whole word. Participants were instructed to choose whether they preferred the first or second audio clip. There were 16 total audio prompts. Half of the prompts featured synthesized versus human speech, while half featured segmented versus whole-word pronunciation. Additionally, half asked the participants to select which they *liked* more and half asked the participants to indicate which they found more *helpful*. Half of the words used were single syllable while the other half contained two syllables.

Due to an absence during the testing session, the sample included 14 participant responses. The results of the survey were tallied with a mark being given to the preferred choice in each audio prompt, then calculated as a percentage of the overall prompts of similar type and purpose. Participants showed an overall preference for whole-word human speech, with both human speech and whole-word pronunciation receiving more than two-thirds of preferences when compared to the alternatives. Results of the human versus synthesized audio can be seen in Table 1, while results of whole word versus the addition of segmented pronunciation are found in Table 2.

**Table 1**
*Preferences for Recorded Human or Computer Synthesized Speech*

| **Audio Condition** | **One-Syllable (like)** | **Two-Syllable (like)** | **One-Syllable (helpful)** | **Two-Syllable (helpful)** | **Overall** |
|---|---|---|---|---|---|
| Recorded human speech (adult female, 44 kbps mp3) | 64% | 64% | 75% | 79% | 71% |
| Computer synthesized speech (Microsoft Anna) | 36% | 36% | 25 % | 21% | 29% |

Note: *n* = 15 for all conditions

**Table 2**
*Preferences for Whole-Word or Segmented Pronunciation.*

| Pronunciation Condition | One-Syllable (like) | Two-Syllable (like) | One-Syllable (helpful) | Two-Syllable (helpful) | Overall |
|---|---|---|---|---|---|
| Whole-word (human) | 54% | 64% | 75% | 79% | 68% |
| Segmented + whole-word pronunciation (human) | 46% | 36% | 25% | 21% | 32% |

Note: *n* = 15 for all conditions

This information was confirmed and elaborated upon during one-on-one interviews. One challenge of interviews with first grade students is that the interviews must be kept short and, although the children were able to form and express their own opinions, they often found it difficult to elaborate upon the reasons for those opinions. This is reflected by many circular and non-descriptive explanations such as “I liked it because it sounds good” and “I liked it because… I just liked it.” Aside from these nondescript elaborations, several students remarked on the clarity of audio (they found the human speech to be more “clear”) and on the rate of speech: that the synthesized speech was spoken too quickly and that they preferred the human pronunciation which was pronounced more slowly. Table 3 shows the rate of each incident and gives examples of statements in each category, as well as underscoring the fact that human speech was strongly preferred over the computer-synthesized counterparts.

**Table 3**
*Themes of Audio Speech Format Interview Responses*

| Response | Count | Percent |
|---|---|---|
| Preferences | | |
| Preferred human speech (single multi-syllabic word) | 13 | 87% |
| Preferred computer speech (single multi-syllabic word) | 2 | 13% |
| Preferred human speech (four word sentence) | 14 | 93% |
| Preferred computer speech (four word sentence) | 1 | 7% |
| Reasons | | |
| Non-descript explanation | 15 | 50% |
| “It says it good.” | | |
| “It sounded better.” | | |
| “I just liked it.” | | |
| “It sounds pretty” | | |
| Reason: Slower rate of speech | 5 | 17% |
| “It said it slower, but the other one said it fast.” | | |
| “You could hear all the sound, not too fast.” | | |
| “It’s like you’re making a word fast” | | |
| Reason: Clarity | 5 | 17% |
| “It’s more clear.” | | |
| “You could hear all the sound” | | |

## Format of Graphics/Visuals

It was important to test multiple formats of visuals due to their complex nature as being potentially beneficial or potentially detrimental features of e-books, depending on design and implementation. This study focused on examining the effect of presenting static graphic, animated graphic, or no graphic accompanying clickable text in an e-book format. The visuals were specifically designed to be "considerate" by being directly illustrative of the text on each page and by only presenting a single image/animation per page.

User interactions with the system were recorded by log files automatically generated by the *DinoRead* software, accompanied by "screencast video" which captured the screen image, mouse movements and clicks, and audio input via headset microphone. Using this combination, each usage session was reviewed with a critical eye for both *active* reading behaviors and *passive* reading behaviors actions. Active reading included on-task authentic reading tasks and practice including: (a) attempts to read each word aloud; and (b) clicks on the words combined with reading aloud (either as a pre-speaking scaffold, or as a post-speaking confirmation.) Passive interaction behaviors were ones that did not involve an active oral reading role, including: (a) passive reading, indicated by clicking on words to hear them without an attempt to actually try to synthesize the word (though it must be noted that there may still be some academic value to this task, and there may also be internal processing which we were unable to observe); and (b) aborting reading early on a page, or skipping pages entirely.

During this phase, some data items were unable to be included. One participant terminated the screen recording software prior to saving the screencast videos, thus making it impossible to observe or review user interaction, although satisfaction survey data was able to be recorded. Another participant erroneously entered the wrong participant ID number for two of the three activities, creating incorrect test cases and thus skewing and invalidating the data. These missing and invalid data were not used, as noted in the sample numbers.

The automatically-recorded metric used to determine “time on-task” was the amount of time spent per word, since the stories had variable words per page and total words. A slightly higher time-per-word was expected for the static image test case, and yet longer duration per word for the animation case, to account for the processing time students would spend viewing and processing the picture or animation, though this would not be likely to add more than 5 seconds per page or, in this study, 40 seconds per story. As an extension of this, it stands to reason that for books that have more words per picture, the amount of impact of time spent on such visuals as a will diminish as a ratio of the overall reading time. This was found to be true by examining the time-per-word correlated to each book (Table 4), since the books were read in the same sequence by all test groups. However, the two word difference between Book 2: The Mop Shop, and Book 3: Pig in a Wig does not appear to account for the difference in time-per-word. In this case, it is likely a result of either diminishing patience/motivation, or a direct result of the story content itself.

**Table 4**
*Seconds Per Word, by Book Sequence and Graphic Format*

| **Condition** | ***n*** | ***Mean*** | ***SD*** |
|---|---|---|---|
| Book / Sequence | | | |
| Book 1: Cam the Cat (51 words) | 13 | 3.22 | 0.80 |
| Book 2: The Mop Shop (83 words) | 13 | 2.45 | 0.92 |
| Book 3: Pig in a Wig (85 words) | 13 | 1.90 | 0.69 |
| Graphic Format | | | |
| No graphic | 13 | 2.24 | 1.00 |
| Static image | 13 | 2.65 | 0.70 |
| Animation | 13 | 2.69 | 1.10 |

Viewing the data, it also appears that we can see that the time spent per word, according to the log file alone, confirmed a slight increase with images and animations. It appears that presenting an image did, in fact, slow down the activity process slightly, but not to an excessive degree.

However, using time spent per word as a sole indicator is insufficient for determining time on task. It was essential to examine the recorded interaction sessions to determine exactly what that time was spent doing and to determine if there were trends tied to the graphic/visual variable. Upon reviewing each participant's usage of each book, important active and passive reading behaviors as already noted were tallied. Additional events were also noted, as would be expected in an authentic classroom context as this developmental study purposefully tested within. These distractions included: excited exclamations about content (particularly, illustrations and animations); peer conversation and assistance; and other miscellaneous pauses and breaks, caused by classroom distractions, restroom needs, or other sources. These incidences underscore the importance of viewing the automatically logged data metrics cautiously, and of supplementing them with observations such as via screen recording as utilized in this study. Most of these additional time expenditures were coincidental and did not appear to have any relationship to any known variables, such as sudden restroom needs, or a neighbor asking for assistance.

Particularly important for indicating attention to authentic reading tasks are the active (Table 5) and passive (Table 6) behaviors already discussed; specifically, the positive interactions of percentage of words read aloud correctly, the percentage of words clicked for support (either pre-read-aloud or post-read for confirmation), and the potentially negative interactions of passive reading (clicking without attempting to read the word) and skipped pages. The sequence of use seems to have a clear impact on engagement, particularly on the number of words read aloud, the number of words clicked for support, and the number of pages skipped, though the percentage of clicks that were passive seems to be fairly constant and is more likely caused by the use tendencies of individual users.

**Table 5**

*Active Behaviors, by Book Sequence and Graphic Format*

| Condition | *n* | Read Aloud % *M* | Read Aloud % *SD* | Word Click % *M* | Word Click % *SD* |
|---|---|---|---|---|---|
| Book / Sequence | | | | | |
| Book 1: Cam the Cat (51 words) | 9 | 65 | 43 | 55 | 43 |
| Book 2: The Mop Shop (83 words) | 9 | 46 | 31 | 47 | 38 |
| Book 3: Pig in a Wig (85 words) | 8 | 23 | 36 | 38 | 43 |
| Graphic Format | | | | | |
| No graphic | 9 | 47 | 38 | 39 | 42 |
| Static image | 8 | 53 | 43 | 48 | 38 |
| Animation | 9 | 36 | 36 | 55 | 42 |

**Table 6**

*Passive Behaviors, by Book Sequence and Graphic Format*

| Condition | *n* | Passive Click % *M* | Passive Click % *SD* | Pages Skipped % *M* | Pages Skipped % *SD* |
|---|---|---|---|---|---|
| Book / Sequence | | | | | |
| Book 1: Cam the Cat (51 words) | 9 | 58 | 43 | 1 | 4 |
| Book 2: The Mop Shop (83 words) | 9 | 54 | 34 | 22 | 24 |
| Book 3: Pig in a Wig (85 words) | 8 | 66 | 18 | 39 | 39 |
| Graphic Format | | | | | |
| No graphic | 9 | 50 | 36 | 26 | 37 |
| Static image | 8 | 60 | 25 | 22 | 22 |
| Animation | 9 | 68 | 36 | 11 | 24 |

The format of graphic does not appear to have much of an impact on most active or passive interactions, although increasingly-engaging graphics did result in declining rates of page skips. It should be noted that these numbers wildly varied in all scenarios, as can be seen in the standard deviations. It is likely that individual user preferences and characteristics such as patience and regulatory skills may play the biggest factor of all. For example, four participants increasingly skipped pages as time went on, possibly indicating a shrinking level of patience or attention span, or cognitive overload from continual use of the software. On the other hand, three participants never skipped any pages, and two of those users continued to demonstrate high read-aloud rates even as time progressed. Likewise, one participant read 100% of the words aloud in every scenario, while three participants had consistently low read-aloud rates, and the remainder had read-aloud attempts progressively diminish as a function of how long they had been using the software.

It is also important to keep user satisfaction in mind, as an unmotivated user is more likely to lose patience or terminate the task. Satisfaction in each test case was determined via a “happy face survey” (Appendix C) immediately following each e-book activity, combined with one-on-one interviews to garner more information about the reasons for satisfaction levels. As noted in Table 7, satisfaction ratings were overall positive (mean greater than 3 on a 4-point Likert scale) for all three books, indicating that sequence did not seem to adversely affect enjoyment levels to the same extent that it affected many participants’ interaction behaviors. However, the format of graphics did seem to play a role in enjoyment of the task; animations scored an average of 28% higher satisfaction rating than the control case with no visuals.

**Table 7**
*User Satisfaction, by Book Sequence and Graphic Format*

| **Condition** | ***n*** | ***Mean*** | ***SD*** |
|---|---|---|---|
| Book / Sequence | | | |
| Book 1: Cam the Cat (51 words) | 14 | 3.36 | 1.08 |
| Book 2: The Mop Shop (83 words) | 14 | 3.21 | 1.12 |
| Book 3: Pig in a Wig (85 words) | 14 | 3.07 | 1.07 |
| Graphic Format | | | |
| No graphic | 14 | 2.79 | 1.25 |
| Static image | 14 | 3.29 | 0.91 |
| Animation | 14 | 3.57 | 0.94 |

To gain further insight about underlying reasons for user satisfaction or frustrations, one-on-one interviews were conducted and common threads (Table 8) were that the most enjoyable books were the ones that "were funny" and "had moving pictures." All books were represented as favorites, but not equally; the first book in the sequence had the highest number of votes, while the last book read in the activity received the least; this once again indicates that task duration, attention span, patience, or increasing difficulty or word count may play a role.

Surprisingly, two participants responded that their favorite book was one that had no graphics at all, but were unable to elaborate on why they preferred those stories. No respondents favored a story featuring static images; however, two participants did state that overall they preferred static images over the animations, with one stated reason being that "it's hard to see what happens with the pictures moving."

**Table 8**
*Themes of Graphic Format Interview Responses*

| Theme | Count | Percent |
|---|---|---|
| Book preference | | |
| Book 1: Cam the Cat | 7 | 50% |
| Book 2: The Mop Shop | 4 | 29% |
| Book 3: Pig in a Wig | 3 | 21% |
| Reasons for book preferences | | |
| Animation | 3 | 21% |
| "Because when I turn the page they move." | | |
| "…it's sort of like a video. I like videos." | | |
| Action | 3 | 21% |
| "Because he was trying to catch the rat" | | |
| "...when the wig falls down and it gets a new wig." | | |
| Humorous content | 3 | 21% |
| "Cam the Cat because it was so funny." | | |
| "Because they were dressing up kind of funny." | | |
| Graphic format preference | | |
| Animation | 11 | 79% |
| Static pictures | 2 | 14% |
| No preference – "I like all of them" | 1 | 7% |
| Reasons for format preferences | | |
| Clarity / ease of understanding | 3 | 21% |
| "…it's hard to see what happens with the pictures moving." | | |
| "I had never seen a wig but in the story I saw a wig." | | |
| "…because it helped you read." | | |

## ASR Input

Development and testing of ASR input for program control followed a nearly-identical sequence as the graphic format design and testing process. Based on the results of the graphic format tests, animations were chosen to be incorporated into the software due to their beneficial effect on user satisfaction and page skips. However, animations did appear to cause more time spent than alternative visual support possibilities, while correlating with a shift from authentic oral reading to passive reading. Since increasing active oral reading is a central goal of this developmental study, this poses a concern.

The aim of testing ASR input was to determine whether the use of ASR input via headset microphone could encourage or enforce more authentic attempts at reading (and, possibly, more clicks on word pronunciations to support this effort when necessary). The two cases tested were using ASR as an optional tool intended to motivate readers by rewarding them with animation only after all words on the page are read. In this way, the animation is actually a motivational reward while the ASR mechanism is used to check input to trigger the reward (until triggering the animation, the first frame would appear on the page as a static image). However, the user maintained control over whether or how long to attempt to read the words on the page, and could turn the page on command by clicking the page corner.

The second format of ASR input for control was designed to prevent users from skipping authentic reading or page skips altogether. As noted by the previous test phase, when users are given complete control, they may opt to skip reading words, may opt for a passive audience role, or may even bypass as many as 40% of the pages in a book, taking just enough time to view the visual in these cases. The intention of ASR control was to ensure that the user attempted to orally read every word in the story; even if words are read incorrectly, there is a mechanism to move on to a new word if the user is struggling even after receiving support. Both of these ASR formats were also compared to a format identical to the animation test condition from the previous design phase, which triggers animation upon page turn but does not utilize ASR input at all.

Data to be analyzed for ASR format testing included the same data types and sources as the graphic format testing: (1) logged metrics and recorded screencasts to determine time spent as well as on-task and off-task behaviors; (2) happy face surveys and one-on-one interviews to determine satisfaction levels and personal preferences, along with further elaboration when possible. However, there were some additional complexities involved due to the need to analyze the efficacy of ASR, plus a higher rate of aborted use due to difficulties or frustrations caused by ASR.

In reviewing the amount of seconds spent per word on the three new books for this development phase (Table 9), a pattern emerges similar to the one previously seen: time spent per word diminishes as overall time on the activity – along with book complexity and word count – increase. There appears to be a clear effect of either overall time elapsed, book complexity, or word count on time spent per word, though investigating these particular concerns was outside the scope of this study.

The introduction of ASR input as an optional or required mechanism had a noticeable effect on time spent per word, especially when users had to attempt to speak all words on the page to progress to the next page. This input mechanism nearly doubled the mean time spent on the task. However, was this due to increased reading attempts and support clicks, or due to difficulties or time wasted? The fact that there were so many early terminations of the program – 9 out of 14 users aborted the activity before reaching the end of the book – implies that the ASR mechanism, as implemented, is either ineffective or bogs down the process too much such that user satisfaction drops to unacceptable levels.

**Table 9**
*Seconds Per Word, by Book Sequence and ASR Format*

| **Condition** | ***n*** | ***Mean*** | ***SD*** |
|---|---|---|---|
| Book / Sequence | | | |
| Book 4: Miss Plum's Fun Run (99 words) | 13 | 3.09 | 1.19 |
| Book 5: Kate and Jake (134 words) | 8 | 2.68 | 1.46 |
| Book 6: My Bike (171 words) | 10 | 1.93 | 1.24 |
| ASR Format | | | |
| No ASR | 14 | 2.07 | 0.91 |
| ASR optional | 12 | 2.63 | 1.53 |
| ASR required | 5 | 3.96 | 0.79 |

To further understand the actual interaction behaviors that occurred and how they may have impacted time-on-task, the recorded screencast videos were analyzed and interactions recorded similar to the previous development phase. These behaviors included active reading behaviors (Table 10) including correct oral reading of words and clicking on words for pronunciation support if necessary, as well as passive reading behaviors (Table 11) including the percentage of clicks that were not accompanied by a reading attempt and the percent of pages completely skipped with no reading attempted. Unfortunately, the introduction of ASR input did not have the desired effect; in fact, it resulted in the opposite behaviors occurring. The more ASR was required to control and manipulate the e-books, the lower the percentage of words that were spoken aloud or clicked on, and the more pages ended up being skipped. As with previous testing, there was a lot of variability by individual user preferences and interaction habits, as seen in the standard deviation for each mean.

**Table 10**
*Active Behaviors, by Book Sequence and ASR Format*

| | | Read Aloud % | | Word Click % | |
|---|---|---|---|---|---|
| **Condition** | ***n*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Book / Sequence | | | | | |
| Book 4: Miss Plum's Fun Run (99 words) | 9 | 53 | 38 | 9 | 21 |
| Book 5: Kate and Jake (134 words) | 9 | 51 | 42 | 3 | 5 |
| Book 6: My Bike (171 words) | 8 | 52 | 43 | 13 | 31 |
| ASR Format | | | | | |
| No ASR | 8 | 72 | 37 | 14 | 33 |
| ASR optional | 10 | 45 | 36 | 10 | 21 |
| ASR required | 8 | 40 | 43 | 2 | 2 |

**Table 11**
*Passive Behaviors, by Book Sequence and ASR Format*

| | | Passive Click % | | Page Skips % | |
|---|---|---|---|---|---|
| **Condition** | ***n*** | ***M*** | ***SD*** | ***M*** | ***SD*** |
| Book / Sequence | | | | | |
| Book 4: Miss Plum's Fun Run (99 words) | 9 | 34 | 52 | 36 | 37 |
| Book 5: Kate and Jake (134 words) | 9 | 21 | 22 | 29 | 33 |
| Book 6: My Bike (171 words) | 8 | 7 | 12 | 51 | 40 |
| ASR Format | | | | | |
| No ASR | 8 | 2 | 4 | 16 | 28 |
| ASR optional | 10 | 32 | 42 | 33 | 34 |
| ASR required | 8 | 18 | 22 | 69 | 29 |

In this ASR testing phase, it was important to note many additional behaviors and interactions not unique to the test conditions. As noted in the review of literature, ASR has come a long way but is not yet capable of 100% accuracy with normal, fluid speech. In addition, children's pitch, formants, and speech patterns pose special difficulties.

Although this study attempted to help accommodate the physical characteristics of children's speech through training with a high-pitched adult female voice (a closer approximation of young children than the standard adult male profile), there were still many difficulties that prevented sufficient success of the ASR system, as noted in Table 12.

Although there were some navigation difficulties with manually turning pages of the books, as well as two instances of activity termination due to computer (not user) error, the vast majority of problems appeared to stem from the user's interaction with the ASR system. Despite having a training session in proper use of the headset microphone setup prior to the activity, seven of the participants were not using the microphone correctly, with four of them yelling into the microphone or holding it directly onto their mouths, while three others failed to put the microphone down in front of their mouths, or were mumbling/whispering at such a low volume that they could not be heard. Two participants had natural speech impediments – lisp and rhotacism -- and ASR for these participants had a 0% success rate, though it may have been a combination of pronunciation as well as other voice characteristics.

When users interacted correctly with the system – reading at normal volume and at a steady, medium pace without miscues, mispronunciation, halting, speeding through, repeating or skipping words, and pausing at the end of each sentence – the ASR system demonstrated much higher success rates, averaging a 70.26% effective recognition rate when utilized correctly by the user. However, this type of fluent reading is not common in first grade students, and thus this type of interaction occurred only rarely. Overall, the

ASR recognition actually only reached a mean ASR success rate of 55.64%, with a standard deviation of 28.72.

Some interaction issues appeared to be a result of not understanding the verbal directions or getting confused by the color-coding used to indicate which words were recognized, which were logged as an error, and what the next word to pronounce should be. Likewise, some users would not wait to at all to see a response when trying to pronounce a word that had not been detected; instead, they chose to repeatedly say the word without pausing. This lack of a pause impeded the system's ability to distinguish the word as a finite unit and thus register it as a successful pronunciation.

It is also possible that participant found the system navigation mechanisms confusing because they changed from test condition to test condition, and if the participant didn't listen carefully to the verbal prompt at commencing each e-book, they may not realize the differences from a previous e-book.

**Table 12**
*Observed Interaction Problems and Difficulties*

| **Interaction Problem** | **Count** |
|---|---|
| ASR problems | 27 |
| Speaking too loudly / too close to microphone | 4 |
| Speaking too quietly or microphone not in front of mouth | 3 |
| Words not correctly identified, but user continues speaking | 6 |
| Speech impediment causing mispronunciation | 2 |
| Repeating word after reading it | 4 |
| Confusion about color-coding indicators | 8 |
| Not attempting ASR at all | 2 |
| Halting progress to go back to highlighted word | 5 |
| Difficulty manually turning page | 5 |
| Activity aborted due to technical error (page won't turn) | 2 |

Given these challenges, it is not surprising that participant satisfaction levels (Table 13) were lower for this activity than for the previous testing phase. However, it should be noted that aborted attempts did not allow user selection of a satisfaction level, and were automatically assigned a "1" on the scale; if a user is ending the task prematurely, that seemed like a strong indicator of dissatisfaction. On the 4-point Likert scale being used, a 2.5 would indicate neutral feelings, with anything below that tending toward negative satisfaction levels and anything above 2.5 indicating positive satisfaction. In this case, the only condition that resulted in slightly-positive satisfaction was the non-ASR condition, using standard e-book with animations that play on page-turn. However, even this level was lower than recorded for the same conditions in the previous test phase. This may be due to cognitive fatigue or lingering frustrations from already participating in the ASR test conditions, or may simply be that the novelty of the software has started to wear off in this second use session.

**Table 13**
*User Satisfaction, by Book Sequence and ASR Format*

| **Condition** | ***n*** **(*n* aborted)** | ***Mean*** | ***SD*** |
|---|---|---|---|
| Book / Sequence | | | |
| Book 4: Miss Plum's Fun Run (99 words) | 14 (3) | 2.43 | 1.40 |
| Book 5: Kate and Jake (134 words) | 14 (5) | 2.29 | 1.20 |
| Book 6: My Bike (171 words) | 14 (3) | 2.21 | 0.97 |
| ASR Format | | | |
| No ASR | 14 (0) | 2.93 | 1.07 |
| ASR optional | 14 (2) | 2.21 | 1.12 |
| ASR required | 14 (9) | 1.57 | 0.94 |

One-on-one interviews were conducted to better understand whether participants did or did not like being able to use ASR input, and specific reasons to explain why. Participants' feelings about ASR were mixed, with half of the 14 respondents indicating that they had no desire to use the ASR, while the other half appreciated the option of using ASR, for a variety of reasons (see Table 14). Only one respondent noted, however, that they enjoyed the condition that required ASR input in order to turn the page.

Users who did not like ASR indicated personal preference for silent reading or the desire to avoid being bogged down by the tedium of the ASR reading, but also illustrated important logistical points about technical difficulties including ineffectiveness of the speech recognition as well as distractions and background noise caused by multiple users in a classroom setting. Participants who liked having access to ASR did so because they felt it was easy to use, fun to manipulate the words using speech, and that speaking the words aloud helped them with practice and memory of the words encountered.

**Table 14**
*Themes of ASR Format Interview Responses*

| Theme | Count | Percent |
|---|---|---|
| Disliked ASR | 7 | 50% |
| Technical difficulties | 1 | 7% |
| "I was saying it out loud and I was trying my best but I couldn't." | | |
| Prefer silent reading | 2 | 14% |
| "I like just reading" | | |
| Speed / facility | 3 | 21% |
| "I'm kind of lazy so I don't really like... I just like to go on and do something else, you know what I mean?" | | |
| "The pictures just moved" | | |
| "I just look at the pictures" | | |
| Reading aloud distracts peers | 1 | 7% |
| "I liked not reading the words out loud so you don't bother other people." | | |
| Liked ASR | 7 | 50% |
| No reason given | 2 | 14% |
| Ease of use | 2 | 14% |
| "I liked the page turning by itself" | | |
| "you could read the words if you want to because I think it gives easy words." | | |
| Enjoyable / fun | 2 | 14% |
| "I like reading the words" | | |
| "The one where the words turn red and green because it was fun." | | |
| Helpful for learning/memory | 1 | 7% |
| "Because we could learn... we could hear them out loud and we could memorize. When we don't do that I kind of forget, that's why I like talking in the computer." | | |

## Tutorial and Practice

The tutorial and practice phase was designed to explore whether these common elements of CAI – including brief animated lesson and subsequent practice of basic phonemic skills – may prove useful to incorporate into an e-book, either as a pre-reading lesson or as a reactive support within the software. The three conditions tested were: lesson plus mouse-clicked word identification; the same lesson plus ASR-triggered synthesis of a word; or a control containing no lesson/practice at all. Screen video recording was used to capture all user interaction, as well as a tool to perform a running record analysis of oral reading fluency before and after the intervention.

Unfortunately, students did not hear or did not follow the directions, or were confused by the presence of an identical story as both the pre- and post-test; 12 participants engaged in this lesson, and 100% of them read only the pre-test running record, while ignoring the 2nd viewing of the same story, which would appear accompanied by the verbal directions “Now read the story out loud again. If you get stuck on a word, try your best to say the word, then move on to the next word. Click on the green arrow at the bottom when you are finished reading all the words.” Instead, all participants immediately clicked the arrow, which quit the activity.

This eliminated ability to use the testing tool to determine the original central focus: immediate effect on oral fluency. However, there was still valuable information to be found pertaining to the secondary purpose of this test phase: interaction behaviors and user satisfaction gained from the exercise. By keeping a running tally of common behaviors and interactions in this CAI component (Table 15), it was determined that, as had already been observed in the previous phase, many of the users were speaking too loudly by yelling or shouting into the microphone, or by holding the microphone directly up to their mouths as they spoke. This was confirmed by viewing the audio waveforms in the FlashBack Express video viewer; when input volume is too loud, the wave gets "clipped", giving it a rectangular appearance (see Figure 3). Because the Microsoft SAPI ASR system is designed to recognize speech by comparing its waveform pattern to a closest matching pattern, having such clipped audio will guarantee reduced efficacy of ASR.

**Table 15**
*Observed Interaction Problems for ASR Input in Tutorial Testing*

| ASR Input Problem | Count | Percent |
|---|---|---|
| Speaking too loudly / microphone too close to mouth | 7 | 64% |
| Rapid repetition of word | 2 | 18% |
| Too quiet or microphone not in front of mouth | 1 | 9% |

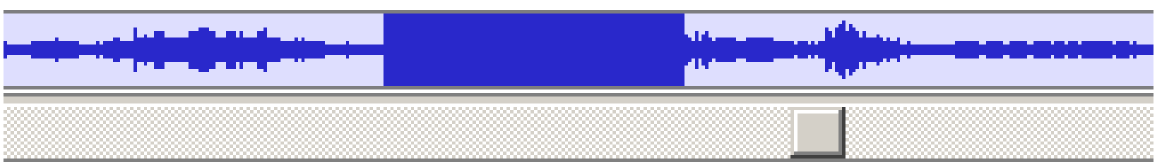


**Figure 3. Audio clipping due to excessive volume.**

The rectangular block in the center shows clipping which distorts audio and diminishes waveform pattern recognition by ASR. Normal, unclipped waveform patterns are seen on the left and right.

Another pattern that emerged is that the ASR seemed to recognize speech input from male participants much more readily than it did for females. The one female participant who pronounced the words correctly and at an appropriate volume still received 0% word recognition from the ASR system, while both males who spoke at an appropriate volume experienced at least 80% rate of speech recognized. This is likely due to the fact that the voice profile was trained with the voice input of a female adult, whose speech pattern has been shown to most closely align with that of a male child (Lee, et al. 1999; Hagen, et al. 2009). Table 16 illustrates that the ASR design as used in this study may be sufficient for male first graders using correct microphone-speaking procedures, but will fail in the presence of excessive volume and may not work sufficiently for female first grade users.

**Table 16**
*Mean ASR Success Rates in Tutorial Practice*

| | Male | | | | Female | | | |
|---|---|---|---|---|---|---|---|---|
| | Correct volume | | Too loud / too close | | Correct volume | | Too loud / too close | |
| **Condition** | ***n*** | ***M*** | ***n*** | ***M*** | ***n*** | ***M*** | ***n*** | ***M*** |
| ASR Success | 2 | 82% | 4 | 6% | 1 | 0% | 4 | 18% |

One on one interviews were conducted with the 12 participants involved in the tutorial/practice study to determine their feelings about both the lesson content as well as preferences for the word identification game using mouse or the spoken synthesis game using ASR via microphone (Table 17). Preference of lesson animation seemed to be determined by characters, their appearances and their actions; no participants made an y mention of the difficulty or academic content of any of the tutorials. Most participants preferred the mouse-click identification game, stating that it was more fun or helpful. This may have been due to familiarity with style of traditional mouse-clicked identification task, but was also likely influenced by the low overall success rates of ASR input, as shown in Table 16.

**Table 17**
*Themes of Tutorial/Practice Interview Responses*

| Theme | Count | Percent |
|---|---|---|
| Favorite lesson / animation | | |
| Long vowel / silent-e | 4 | 33% |
| “because she was a girl” | | |
| “she was pretty and nice” | | |
| Long vowel digraphs | 2 | 17% |
| “because I liked him, he was cool” | | |
| R-controlled / bossy r | 6 | 50% |
| “you had to listen and I like the pictures” | | |
| “raccoon because he said rrrr” | | |
| “it was funny” | | |
| Practice game format preferences | | |
| Liked them both - “because it was so fun” | 3 | 25% |
| Word identification (click) | 7 | 58% |
| “because it teaches me stuff I could know” | | |
| “one you clicked because you finished the game” | | |
| “because it was fun to click” | | |
| Word synthesis (ASR) | 2 | 17% |
| “it’s fun talking into the computer” | | |

## Interpretation and Discussion

### *DinoRead* Software Development

As an iterative development study, testing and tweaking were an ongoing, formative process and results of each testing phase would guide the features to be incorporated into the next phase. There were often clear implications of the findings and how they impact reading software design for this particular population of users. However, determining a “best” implementation based on those implications was not always so clear. The purpose of this study was to develop a software tool that maximizes on-task academic behaviors in reading software, while maintaining a high level of user satisfaction.

These two goals are not always congruent, and in fact are often at odds with each other; after all, the problem is that many e-books and CAI programs have distracting features that have been determined to be “mere entertainment” but to reduce entertainment would also logically be expected to reduce user satisfaction levels (Lefever-Davis & Pearman, 2005; Trushell & Maitland, 2005). Thus a balancing act must occur, choosing considerate features (Labbo & Kuhn, 2000) that will encourage authentic learning tasks, while maintaining a sufficient level of user satisfaction to ensure motivation and engagement (Seo & Woo, 2010).

For the purposes of this study, the assumption is that user satisfaction levels were unacceptable if:

- Common interview themes included comments about boredom or technical difficulties.
- Mean satisfaction survey responses were below neutral (ie. less than 2.5 on 4 point scale).
- Users prematurely aborted the activity.

The first phase of this developmental study examined user preferences and feelings about audio clips for spoken pronunciations and feedback, and found a strong preference for recorded human speech over computer-synthesized speech. This aligns with findings of previous studies (Veletsianos, 2009; Drager, et al., 2010), and is an important aspect to consider in software development, because recording speech can be resource-intensive, both in the sense of time, money, and human resources needed for production, as well as digital file resources that can take up valuable storage space or online streaming bandwidth. Synthesized speech offers benefits of ease and unlimited scalability.

Although users noticed unnatural and undesirable characteristics in the audio, it must be noted that these results are limited to the specific conditions used: adult female speech recorded at 44 kbps, compared to a common synthesized voice, Microsoft Anna. There are other cutting-edge text-to-speech engines that exist, such as Ivona (http://www.ivona.com/en/), Loquendo (http://www.loquendo.com/en/), and Cepstral (http://cepstral.com/), and it is possible that users may find these to be natural enough in tone, formant, and pacing to be a suitable replacement for recorded human speech. This is a worthwhile area to continue researching, especially as technologies advance.

Additionally, this study focused only on user perception and preferences; it did not closely examine the actual effects of such audio files on learning tasks or efficacy of user interaction. For example, users preferred the brevity and natural sound of whole-word pronunciation, but it is still possible that phonemic segmentation of pronunciation may provide more academic benefit.

The second phase of development reviewed a topic that has been extensively researched in the existing literature: how design and format of a visual aid affects user behaviors and satisfaction levels in e-books. The findings of this study supported much of the existing literature by showing that users find graphics and, especially, animations to be engaging and supporting, but that these visuals can also prove to be distractions from the authentic reading task.

This poses a dilemma in software design and development, especially because the most satisfactory visual supports (animations) detracted the most from authentic reading oral practice, and encouraged a more passive stance in the user. It would seem from this, then, that a happy medium would be to use static images instead animated ones. This would certainly be beneficial from a developmental resource standpoint, as animations are much more complex to develop and incorporate than static images are (in this study, they required approximately triple the time span to produce).

However, in this study, the developmental choice based on the feedback was to use a hybrid approach: to display a static image automatically, and to use the more satisfactory animations as a form of reward to potentially be triggered and played based on successful reading, detected via ASR.

In the ASR phase, this study examined whether it is feasible to use a common speech recognition engine (Microsoft Speech API) to encourage or enforce oral reading of an e-book. The findings were complex, and indicated that users encountered too many challenges, difficulties, and frustrations with the particular ASR implementation used in this study to allow it to replace the mouse as a primary control mechanism. Enforcing ASR-recognized reading of all of the words on a page as a requirement for playing the animation and progressing to the next page was found to have a significantly detrimental impact on user satisfaction. This was influenced by some participants' preferences for silent reading, for short activity durations, and for more control, though the most significant factor may have been the unacceptably-low success rate of ASR input.

Successful recognition of speech was impacted by a variety of factors, including: user error with microphones and spoken input procedure; speech impediments and acoustic characteristics of children's speech; and speech patterns common to early readers including halts, miscues, segmentation, and repetition of words, in addition to classroom logistical factors such as background noise.

For this particular context, the ASR design as implemented could not be relied on as a tool to coerce children to read the stories aloud; instead, it encouraged them to give up on the task. However, in the one on one interviews, half of the participants indicated that they enjoyed and were intrigued by the process of speaking into the computer. Accordingly, it seems beneficial to allow this feature as an optional tool for e-book users to voluntarily elect to use as they see fit.

There are a variety of ways the ASR experience in this study could potentially be improved, based on the findings:

- Change color-coding and instructions to more clearly indicate where successful, unsuccessful, and current ASR recognition is occurring.
- Embed demonstration of proper headphone use and speech input procedure for users.
- Provide more training and practice with ASR input.
- Train the speech profile with a more accurate acoustic model of children's speech

In this study, Microsoft Speech API was chosen for its widespread and readily-available accessibility, as well as ease of development. However, it would be worthwhile to further investigate ASR input and control using alternative ASR implementations which have been developed specifically for children, trained with a corpus of children's speech and incorporating features such as subsyllabic recognition. Some possible choices for ongoing research include SONIC, developed at the University of Colorado – Boulder (http://www.bltek.com/virtual-teacher-side-menu/sonic.html) and Carnegie Mellon University's Sphinx toolkit (http://cmusphinx.sourceforge.net/).

The final phase of development of a *DinoRead* prototype involved the potential for incorporating lesson tutorials and game-style practice, including mouse-clicked word identification and ASR-triggered word synthesis tasks. The recorded user interaction videos show that users were confused by the pre-test, post-test setup being employed. Since a large degree of user control was allowed, participants often opted to skip the pre-test or end it before completion and try to move on to the lesson/game. In addition, 100% of participants skipped the post-test rereading of the text, reducing the efficacy of this tool.

However, from a motivational and engagement standpoint, there were some encouraging findings from this phase. In interviews, participants indicated that they found the word identification game to be fun and engaging, and some also found the ASR-input practice to be fun and engaging, as well. The ASR system was found to work more proficiently for the discrete task of speaking single words, with up to 80% success rate of recognizing spoken input, which was higher than any use case in the previous test phase for authentic reading tasks. However, the ASR system was much less effective at recognizing female speech, indicating a need to better-train the voice profile, or possibly to use a different ASR engine. The majority of participants were still demonstrating a lack of proficiency with appropriate spoken input via microphone, further underscoring the importance of proper training and practice for ASR input skills. Further research will be necessary to determine the efficacy of various phonemic lessons and practice tasks, and whether they are beneficial to use as an intervention before or during e-book reading sessions.

Based on all of the findings, a final prototype was developed which incorporated: recorded human spoken audio feedback and clickable-word pronunciations; static images that play animation as a reward for reading all of the text on a page; and lessons and practice games prior to an authentic reading task via e-book that focuses on the specified phonemic skill.

# CHAPTER 5: SUMMARY, CONCLUSION, AND RECOMMENDATIONS

## Summary

This developmental study examined the effects of four different design aspects of an e-book based CAI software tool. The central problem being addressed is that many CAI programs and e-books for early readers encourage passive or off-task behaviors. The aspects researched included: user preferences for feedback and pronunciation audio; effects of visuals and animations on user interaction and satisfaction; effects of ASR input on user interaction and satisfaction; and effect on fluency of phonemic tutorial paired with mouse-clicked word identification or ASR-enhanced word synthesis.

Although the research was exploratory in nature, based on existing literature I hypothesized that users would prefer recorded human speech; that engaging animated graphics would result in higher satisfaction levels; and that increasing ASR-control would increase on-task behaviors, but may reduce user satisfaction.

The study was conducted in systematic, iterative phases, beginning with development of a software tool that could have the various design elements turned on or off for testing of any combination of them. For each design aspect, four phases of design and testing occurred, including reviewing needs, designing and developing the solution, user-testing the solution in an authentic context, and revising or cementing the design choices accordingly. Testing of each design element occurred with 15 first grade student study participants as users, eliciting data and feedback via logged metrics, observation/review of recorded interaction, one-on-one interviews, and satisfaction surveys.

## Conclusion and Recommendations

Reading CAI / eBook Development

Based on the results of this developmental study, there are several guidelines for developing multimedia supports, ASR input controls, and practice games in a CAI system featuring e-books. Elements and aspects interpreted at a glance as being engaging may not always be the most educational or beneficial in a reading program. Multimedia and support features should be carefully assessed to determine whether they are being used in ways that support learning, or instead serve as mere entertainment. A fine balance between learning and entertainment must be maintained, or there may be little educational value in using the tool at all. Software designers and developers of reading programs for early readers should carefully balance on-task authentic learning activities with motivational and satisfactory design elements such as audio-visual supports.

When it comes to spoken audio feedback, first grade users seem to prefer recorded human speech over digitally synthesized speech, which aligns with findings of previous studies (Drager, et al., 2010; Veletsianos, 2009). However, digitally synthesized speech may save development time and resources, so this aspect is worth continuing to research with various cutting-edge speech synthesis engines and voices to determine whether some may work sufficiently as an alternative to recorded human speech.

For pronunciation of words in e-books, children preferred simple whole-word pronunciations instead of segmented speech to break down the phonemic pronunciation.

Although this was a user preference, there is some evidence in the literature that segmented pronunciation in CAI and e-books can prove beneficial for building phonemic awareness skills. Further research is warranted in this area to compare levels of user satisfaction and levels of fluency improvement when using segmented audio pronunciation.

Inclusion of considerate visuals congruent with the text increase user satisfaction when using e-books. However, visuals can also create distractions that lead to passive reading, changing the role of the user from *reader* to *audience*. Such supports may reduce time spent on authentic reading tasks. This is especially true for animated graphics.

Use of ASR input is complex and extremely variable in its implementation and effects. Using ASR input allows for recording of reading difficulties and immediate user feedback or interventions, but significantly increases the time required and can lead to lower user satisfaction. When ASR is a required mechanism of interaction and does not sufficiently recognize correct speech, users become frustrated and often give up on the task entirely. Microsoft SAPI ASR, when used with minimal voice profile training, insufficiently recognized children's speech, especially for young girls. It is possible additional voice profile training in Microsoft SAPI may also improve results, though the voice training system requires the user to be a proficient and fluent reader, rendering it inaccessible to most early readers. Use of other, more child-oriented ASR systems may alleviate this problem. In addition, some users have indicated enjoyment, engagement, and memory benefits of speaking into the computer, while other users prefer silent reading. However, for games/practice involving input of a discrete word or phrase, the ASR input mechanism has a much higher rate of success and resulted in higher user satisfaction. One useful approach may be to use rely on ASR as a required mechanism for whole-word synthesis practice games, while relegating ASR as an optional input for encouragement (rather than enforcement) of authentic reading of passages.

Further research is warranted, especially in the areas of ASR-input and embedded tutorials and practice game interventions for e-books. Some particular topics worthy of exploration include: efficacy of alternative ASR systems; use of game-style points or rewards to encourage ASR use (when optional); and possible use of ASR for recording reader difficulties and dynamically adapting texts or triggering interventions. It would also be beneficial to test the design aspects explored in this study with other contexts and populations to confirm or expand upon the findings, or to find specific findings pertaining to different target populations, such as struggling or EL readers, or gifted/advanced students.

Developmental Research Best Practices – Lessons Learned

In addition to the user-testing which drove findings that often confirmed findings in the existing research literature, it is also useful to examine benefits and challenges of the developmental research process itself. Not only are there unique challenges inherent in designing CAI and e-books to build early literacy, but there are difficulties specific to working with participants of such a young age.

Although the purpose of the developmental study is to guide development of a tool, the difficult irony is that the development process may sometimes be too complex or time-consuming to occur concurrently with the developmental research and testing phases. In the case of this research study, all software tools, tutorials, story texts, and audiovisual content were developed by the principal investigator. In such a paradigm, there simply wouldn’t have been time to develop the required tools in a reasonable time to complete the study with the same group of participants during the first grade school year. Because of this, the software and most content were developed in advance of the study.

The implication of this is that, rather than streamlining the development process, the process is complicated and lengthened because a dynamic tool must be developed that can be used as some permutation of the different design aspects being tested. In other words, one must develop a malleable tool that can quickly be converted into a prototype featuring any permutation of the disputed features. This may not be necessary in all scenarios, but would likely require (a) additional contributors to work on the software and/or content development, and/or (b) an extended duration of time for the developmental study.

The most critical element of a developmental study is the choice of data to be collected and the methods of collection chosen. Development of a software tool affords a unique opportunity to program many metrics to be automatically logged by the software. However, as outlined by the data above, it is likely that such use metrics will not be sufficient to tell the entire story of how the software is being used and what effect it is having on users. Confounding variables and incidents such as external distractions, bathroom breaks, and peer interactions can significantly skew numerical metrics, especially when used in an authentic classroom context.

The use of screencasting software proved to be an invaluable tool for being able to accurately observe all interactions of all users, while maintaining a realistic test environment and limited time span for test periods. The BlueBerry FlashBack Express screencasting software was chosen for its ease of cost, ease of use, and ability to be programmed to begin and end at a certain time or coinciding with running of a specific software application. The participant population is one that has had very limited proficiency with computers, and thus the software use process needed to be as streamlined and “invisible” as possible. The screencasting software would start automatically and run as a hidden/minimized process so that it did not pose confusion or distraction.

However, there were still some challenges and problems caused by this setup. To avoid extraneous, unnecessary video recording, the software was configured to begin when the application started and to automatically terminate and save a video file upon closing the application. This posed a problem because user error could result in closing the application early (either by accident, or by intentionally aborting the task), which would also terminate screen capturing and the FlashBack software would not automatically reset itself; user interaction was necessary to ensure the video capture was recording or waiting to record at specified runtime. This resulted in unobtainable user interaction data for 5 participants who prematurely aborted the recording software. A preferable system would be one that is entirely invisible to the end-user and does not allow or require any user interaction beyond their normal use of the tool being developed. Another potential drawback of the system, as used in this study, is that only computer interactions and audio input can be obtained; however, some tools (including FlashBack) also allow you to capture webcam input, which may give an even fuller picture of what actions and behaviors – or even facial expression responses – are occurring with the user.

User control and input of test mechanisms such as testing measures and participant IDs should be minimized or eliminated as much as possible to prevent user error from reducing or invalidating data. One source of user error in this study was caused by requiring each user to enter his/her participant number into the program for automatically setting test cases and logging use data. Although a simple typing mistake could cause this to occur with any user, it is of particular concern with young or inexperienced computer users. For the most part, the participants were proficient enough to complete this task without incident; however, one user entered ID number "34" instead of "3" into the software, causing placement into a different test group and test conditions for that session. What is worse is that this was done for only two out of the three activities for the session; the third use resulted in a correct ID number, but this duplicated a test condition that had already occurred, thus data could not be validly compared between control and test conditions for this user. The benefit of this user-input ID method is that it is flexible and easy to implement. However, it is likely worthwhile to use a static, hard-coded method rather than user-input which may be prone to error; for example, different computer numbers could be set up to automatically have different ID numbers assigned in the software configuration, and thus any user sitting at that assigned computer will automatically log in with the correct ID.

Collecting feedback about feelings and thought processes from young children also poses unique challenges. These participants expressed that they had never participated in a study like this, and had never been involved in an interview. Although they were excited and eager to participate, there were developmental challenges inherent in the communication process. For example, children of this age are not yet used to thinking or elaborating upon the causal reasons behind their thoughts or feelings. They were forthcoming with unbiased subjective views and opinions, but many participants gave circular or non-descript elaborations such as "I liked it because I just liked it." This may reflect meta-cognitive skills that have not yet developed or been taught, and may also be indicative of a child's smaller vocabulary. This smaller vocabulary also calls out the importance of word choice in interviews, surveys, and questionnaires, and the importance of including illustrative graphics (see Appendices B and C) and reading aloud all prompts and text (due to lack of reading ability at this age.) These best-practices may prove useful with other target populations, as well.

# REFERENCES

Adams, M. J. (2006). The promise of automatic speech recognition for fostering literacy growth in children and adults. In M.C. McKenna, L.D. Labbo, R. D. Kieffer, & D. Reinking (Eds.), *International Handbook of Literacy and Technology, Volume 2* (pp. 109-128). Mahwah, NJ: Lawrence Erlbaum Associates.

Ayres, Tony (2006). Voice activated command and control with speech recognition over wifi. *Science of Computer Programming, 59,* 109-126. doi: 10.1016/j.scico.2005.07.007

Barker, T. A. & Torgesen, J. K. (1995). An evaluation of computer assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research, 13*, 89-103. doi:10.2190/TH3M-BTP7-JEJ5-JFNJ

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141-178. doi:10.1207/s15327809jls0202_2

Buckleitner, W. (2006). The relationship between software design and children's engagement. *Early Education and Development, 17*(3), 489–505. doi: 10.1207/s15566935eed1703_8

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). Berlin: Springer Verlag.

Collins, A.M., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15–42. doi: 10.1207/s15327809jls1301_2

Chuang, C., Porfeli, E., & Algozzine, B. (2008). Development of oral reading fluency in young children at risk for failure. *Journal of Education for Students Placed at Risk, 13*, 402-425.

de Jong, M. T., & Bus, A. G. (2002). Quality of book-reading matters for emergent readers: An experiment with the same book in a regular or electronic format. *Journal of Educational Psychology, 94*(1), 145-155. doi:10.1037/0022-0663.94.1.145

de Jong, Maria T., & Bus, A.G. (2003). How well suited are electronic books to supporting literacy? *Journal of Early Childhood Literacy, 3*(2), 147-164. doi: 10.1177/14687984030032002

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Drager, K. D. R., Reichle, J., & Pinkoski, C. (2010). Synthesized speech output and children: A scoping review. *American Journal of Speech-Language Pathology, 19*. 259–273 doi:10.1044/1058-0360(2010/09-0024)

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001a). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393-447. doi:10.3102/00346543071003393

Ehri, L. C., Nunes, S. R., Willows, D. M., & Schuster, B. V. (2001b). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250-287. doi:10.1598/RRQ.36.3.2

Ertem, I. S. (2011). Understanding interactive CD-ROM storybooks and their functions in reading comprehension: a critical review. *International Journal of Progressive Education, 27*(1), 28-44.

Foster, K. C., Erickson, G. C., Foster, D. F., Brinkman, D., & Torgesen, J. K. (1994). Computer administered instruction in phonological awareness: Evaluation of the DaisyQuest program. *Journal of Research and Development in Education, 27*, 126-137.

Fox, B., & Routh, D. K. (1984). Phonemic analysis and synthesis as word attack skills: revised. *Journal of Educational Psychology, 16,* 1059-1064.

Gerosa, M., Giuliani, D., Narayanan, S., & Potamianos, A. (2009). A review of ASR technologies for children's speech. Proceedings of the 2nd Workshop on Child Computer and Interaction WOCCI 09. ACM Press, 1-8. doi:10.1145/1640377.1640384

Hagen, A., Pellom, B., Van Vuuren, S., & Cole, R. (2004). Advances in children's speech recognition within an interactive literacy tutor. *Proceedings of HLT-NAACL 2004: Short Papers*, Center for Spoken Language Research. Retrieved from http://acl.ldc.upenn.edu/N/N04/N04-4007.pdf

Hagen, A., Pellom, B., & Cole, R. (2007). Highly accurate children's speech recognition for interactive reading tutors using subword units. *Speech Communication, 49*, 861-873. doi:10.1016/j.specom.2007.05.004

Hagen, A. Pellom, B., & Hacioglu, K. (2009). Generating synthetic children's acoustic models from adult models. Proceedings of NAACL HLT 2009: Short Papers, pp 77–80. Boulder, Colorado: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology-new/N/N09/N09-2020.pdf

Hemphill, L., & Tivnan, T. (2008). The importance of early vocabulary for literacy achievement in high-poverty schools. *Journal of Education for Students Placed at Risk, 13*, 426-451. doi:10.1080/10824660802427710

Higgins, E. L. & Raskind, M. H. (2004). Speech recognition-based and automaticity programs to help students with severe reading and spelling problems. *Annals of Dyslexia, 54*(2), 365-388. doi:10.1007/s11881-004-0017-9

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human-computer interaction. *Computers in Human Behavior, 26*, 1278–1288. doi:10.1016/j.chb.2010.05.031

ISO (1998). ISO 9241-11 *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.* Retrieved from http://www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf

Juel, C., Griffith, P. L., & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. doi:10.1037/0022-0663.78.4.243

Karemaker, A., Pitchford, N., & O'Malley, C. (2010). Enhanced recognition of written words and enjoyment of reading in struggling beginner readers through whole-word multimedia software. *Computers & Education, 54*, 199-208. doi:10.1016/j.compedu.2009.07.018

Kegel, C. A., van der Kooy-Hofland, V. A., & Bus, A. G. (2009). Improving early phoneme skills with a computer program: Differential effects of regulatory skills. *Learning and Individual Differences, 19*, 549-554.doi:10.1016/j.lindif.2009.07.002

Kelly, A. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences*, *13*(1), 115-128.doi:10.1207/s15327809jls1301_6

Korat, O., Segal-Drori, O., & Klien, P. (2009). Electronic and printed books with and without adult support as sustaining emergent Literacy. *Journal of Educational Computing Research, 41*(4), 453-475. doi:10.2190/EC.41.4.d

Labbo, L., & Kuhn, M. (2000). Weaving chains of affect and cognition: A young child's understanding of CD-ROM talking books. *Journal of Literacy Research, 32*(2), 187–210. doi:10.1080/10862960009548073

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameter. *Journal of the Acoustical Society of America*, *105*(3), 1455-1468.

Lefever-Davis, S., & Pearman, C. (2005). Early readers and electronic texts: CD-ROM storybook features that influence reading behaviors. *Reading Teacher 58*(5), 446-454. doi: 10.1598/RT.58.5.4

Leitch, D., & MacMillan, T. (2003). Liberated learning initiative innovative technology and inclusion: current issues and future directions for liberated learning research, year III report. Saint Mary's University, Nova Scotia.

Lewin, C. (2000). Exploring the effects of talking book software in UK primary classrooms. *Journal of Research in Reading, 23*(2), 149-157. doi:10.1111/1467-9817.00111

Li, J. (2010). Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning*, *23*(3), 253-275. doi:10.1080/09588221.2010.483678

Li, X., Deng, L., Ju, Y., & Acero, A. (2008). Automatic children's reading tutor on hand-held devices. *Proceedings of International Speech Communication Association*, 1733-1736.

Lundberg, I. (1995). The computer as a tool of remediation in the education of students with reading disabilities: A theory-based approach. *Learning Disability Quarterly*, *18*(2), 89-99. doi:10.1109/ICASSP.2007.367196

Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal, 38*(2), 371-410.

McKenna, M. C. (1998). Electronic texts and the transformation of beginning reading. In D. Reinking, M. C. McKenna, L. D. Labbo, & R. D. Kieffer (Eds.), *Handbook of Literacy and Technology: Transformations in a Post-Typographic World* (pp. 45-59). Mahwah, NJ: Erlbaum.

National Center for Education Statistics, U.S. Department of Education. (2000). Closer look 2000, Entering kindergarten: a portrait of American children when they begin school. Retrieved from http://nces.ed.gov/programs/coe/analysis/2000-e07.asp

National Center for Education Statistics, U.S. Department of Education. (2010a). *The condition of education 2010* (NCES 2010-028). Retrieved from http://nces.ed.gov/pubs2010/2010028.pdf

National Center for Education Statistics, U.S. Department of Education. (2010b). *Digest of education statistics 2009* (NCES 2010-013). Retrieved from http://nces.ed.gov/pubs2010/2010013_0.pdf

Nelson, W.A., Bueno, K. A., & Hufstutler, S. (1999). If you build it, they will come. But how will they use it? *Journal of Research on Computing in Education, 32*, 270-286.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning, 21*(5), 393-408. doi:10.1080/09588220802447651

Nielsen, J. (1993). *Usability engineering*. San Francisco, CA: Morgan Kaufman.

Oakley, G. & Jay, J. (2008). "Making time" for reading: factors that influence the success of multimedia reading in the home. *The Reading Teacher, 62*(3), 246–255. doi: 10.1598/RT.62.3.6

Pearman, C. & Chang, C. (2010). Scaffolding or distracting: CD-ROM storybooks and young readers. *Tech Trends, 54*(4), 52-56.

Perfetti, C. A., Beck, I., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.

Piaget, J. (1950). *The psychology of intelligence.* (M. Piercy and D. Berlyne, Trans.). New York: Routledge.

Potamianos, A. & Narayanan, S. (2003). Robust recognition of children's speech. *IEEE Transactions on Speech and Audio Processing, 11*(6). Retrieved from http://sail.usc.edu/publications/pona_childasr_review.pdf

Prabhu, M. T. (2010). Speech recognition, mobile apps help build reading skills. *eSchool News, 13*(7), 22.

Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychology in the Schools, 46*(7), 679-691. doi:10.1002/pits.20407

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, MA: University Press.

Reeves, T. C., Herrington, J., & Oliver, R. (2005). Design research: A socially responsible approach to instructional technology research in higher education. *Journal of Computing in Higher Education*, *16*(2), 97-116. doi:10.1007/BF02961476

Reitsma, P., & Wesseling, R. (1998). Effects of computer-assisted training of blending skills in kindergartners. *Scientific Studies of Reading, 2*(4), 301-320. doi:10.1207/s1532799xssr0204_1

Richey, R., Klein, J., & Nelson, W. (2003). Developmental research: Studies of instructional design and development. In D. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology,* (pp. 1099−1130). Hillsdale, NJ: Lawrence Erlbaum.

Robertson, J. W. (1994). Usability and children’s software: A user-centered design methodology. *Journal of Computing in Childhood Education, 5*(3/4), 257-271.

Roskos, K., Brueck, J., Widman, S. (2009). Investigating analytic tools for e-book design in early literacy learning. *Journal of Interactive Online Learning, 18*(3), 218-341. Retrieved from http://www.ncolr.org/jiol/issues/pdf/8.3.3.pdf

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19*, 285–307. doi:10.1007/s10648-007-9046-3

Segers, E., & Verhoeven, L. (2004). Computer-supported phonological awareness intervention for kindergarten children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 35*, 229-239. doi:10.1044/0161-1461(2004/022)

Sensenbaugh, R. (1996). *Phonemic awareness: An important early step in learning to read. ERIC Digest.* Blooming, IN: ERIC Clearinghouse on Reading English and Communication. Retrieved from http://www.ericdigests.org/1997-2/read.htm

Seo, Y., & Woo, H. (2010). The identification, implementation, and evaluation of critical user interface design features of computer-assisted instruction programs in mathematics for students with learning disabilities. *Computers & Education, 55*, 363–377. doi:10.1016/j.compedu.2010.02.002

Shamir, A. & Korat, O. (2006). How to select CD-ROM storybooks for young children: the teacher's role. *The Reading Teacher, 59*(6), 532-543.doi:10.1598/RT.59.6.3

Shamir, A., Korat, O., & Barbi, N. (2008). The effects of CD-ROM storybook reading on low SES kindergarteners' emergent literacy as a function of learning context. *Computers & Education, 51*, 354–367. doi: 10.1016/j.compedu.2007.05.010

Shamir, A. (2009). Process and outcomes of joint activity with e-books for promoting kindergarteners' emergent literacy. *Educational Media International, 46*(1), 81-96. doi:10.1080/09523980902781295

Silverman, R., & Crandell, J. D. (2010). Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly, 45*(3), 318-340. doi:10.1598/RRQ.45.3.3

Strommen, E. F., & Frome, F. S. (1993). Talking back to Big Bird: Preschool users and a simple speech recognition system. *Educational Technology Research and Development, 42*(1), 5-16. doi:10.1007/BF02297088

Swank, L. K., & Catts, H. W. (1994). Phonological awareness and written word decoding. *Language, Speech, and Hearing Services in Schools, 25,* 9-14. doi:10.1044/lle6.1.26

Torgesen, J. K., Morgan, S., & Davis, C. (1992). The effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology, 84*, 364-370. doi:10.1037/0022-0663.84.3.364

Trushell, J., & Maitland, A. (2005). Primary pupils' recall of interactive storybooks on CD-ROM: inconsiderate interactive features and forgetting. *British Journal of Educational Technology, 36*(1), 57-66. doi:10.1111/j.1467-8535.2005.00438.x

Veletsianos, G. (2009). The impact and implications of virtual character expressiveness on learning and agent–learner interactions. *Journal of Computer Assisted Learning, 25*, 345–357. doi:10.1111/j.1365-2729.2009.00317.x

Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development, 53*(4), 5-23. doi:10.1007/BF02504682

APPENDIX A

# Parental Consent Letter

**Parental Consent Letter**

Dear Parent,

My name is Matthew Gudenius and I am the computer teacher at Calistoga Elementary School. I am also a graduate student pursuing a Master of Science in Educational Technology at Boise State University. I am currently working on a Master's Thesis in which I am developing a new software tool to assist reading skills for first grade students.

I would like to invite your child to participate in this developmental project by testing the program and providing feedback on various aspects. As a participant in the developmental study, your child would use the reading software to read e-books and provide feedback about the enjoyment and effectiveness of audio feedback, visual graphics and animations, and speech recognition. All feedback and results will be anonymous and confidential.

If you give permission for your child to participate, please read and sign the consent form on the back and return it to school. If you have any questions, feel free to contact me by telephone or email.

Sincerely,

Matthew Gudenius
707-942-4398 x728
mgudenius@calistoga.k12.ca.us

**Principal Investigator:** Matthew Gudenius

**Brief description of the project:**

The purpose of this research is to develop a learning tool to assist students in reading first-grade level stories in English. Students will use the program to read 6 different e-books, supported in varying degrees by audio, pictures, animations, speech recognition, and lessons. All activities would take place at the child's school during regular school hours. No extra time commitments would be required.

**Potential risks or benefits:**

No known risks are present. Benefits include contribution to research and best practices for developing educational software tools for literacy. Participants may also benefit from receiving additional computer-assisted reading practice with speech recognition.

**My signature below indicates that:**

- Participation in the project is voluntary.
- If I choose not to participate, there will be no penalties and my child's access to educational services will not be affected.
- I understand that the project involves use of a computer and the child's activities in the program will be recorded.
- I understand who to ask if I have questions about the project: Matthew Gudenius at mgudenius@calistoga.k12.ca.us or (707) 942-4398 x728.
- I understand that I can stop participating at any time, and that deciding not to participate would not affect my relationship with the school.
- I understand that the results of this study will not be associated with my child's name, and the information about my child will be kept confidential.

---

I give permission for information about the study to be used, including possible future presentations and publications about the design study.

Name of child: ______________________________________________________

Signature: ___________________________ Date: _________________________

Estimado Padre,

Me llamo Matthew Gudenius y yo soy el maestro de las computadoras en la escuela primaria de Calistoga. Soy también un estudiante de posgrado por un Maestría de la Ciencia en la Tecnología Educativa en Universidad Pública de Boise. Trabajo en la Tesis Maestría en

La que desarrollo una nueva software para ayudar la lectura de estudiantes en primer grado. Querría invitar su niño a tomar parte en este proyecto por probando el programa y proporcionar reacción en varios aspectos. Cada participante en el estudio va a utilizar el software de lectura para leer libros electrónicos y proporcionar reacción acerca del placer y la eficacia de audio, de la gráfica y las animaciones visuales, y del reconocimiento del habla. Toda la reacción y los resultados serán anónimas y confidenciales.

Si da permiso para su niño a participar, leer por favor y para firmar la forma de consentimiento en la espalda y regresar a la escuela. Si tiene cualquier pregunta, siéntase libre contactarme por teléfono o correo electrónico.

Sinceramente,

Matthew Gudenius
707-942-4398 x728
mgudenius@calistoga.k12.ca.us

**Principal Investigador:** Matthew Gudenius

**La descripción breve del proyecto:**

El propósito de esta investigación es de desarrollar una herramienta de aprendizaje para ayudar a estudiantes en leer primer-grado historias planas en inglés. Los estudiantes utilizarán el programa para leer 6 libro electrónico diferente, apoyado en variar grados por audio, por las imágenes, por las animaciones, por el reconocimiento del habla, y por las lecciones. Todas las actividades sucederían en la escuela del niño durante horas de clase regulares. Ningún compromisos extra del tiempo serían requeridos.

**El potencial se arriesga o los beneficios:**

Ningunos riesgos son presentes. Los beneficios incluyen mejorar diseño de software educativo para la capacidad de leer. Los participantes también pueden beneficiar de recibir la práctica adicional de lectura con reconocimiento del habla por la computadora.

Mi firma debajo de indica que:

- Participación en el proyecto es voluntaria.
- Si escojo no participar, no habrá penas y el acceso de mi niño a servicios educativos no serán afectados.
- Comprendo que el proyecto implica el uso de una computadora y las actividades del niño en el programa será registrado.
- Comprendo quién preguntar si tengo preguntas acerca del proyecto: Matthew Gudenius en mgudenius@calistoga.k12.ca.us o (707) 942-4398 x728.
- Comprendo que puedo parar participar en tiempo, y eso decidiendo no participar no afectaría mi relación con la escuela.
- Comprendo que los resultados de este estudio no serán asociados con el nombre de mi niño, y con la información sobre mi niño será mantenida confidencial.

---

Doy permiso para la información para ser utilizado para éstos investiga propósitos, inclusive presentaciones y publicaciones posibles sobre el diseño.

El nombre de niño: ______________________________________________

Firma: ______________________________ Fecha: _________________________

APPENDIX B

# Technology Exposure / Knowledge Survey

## Technology Exposure / Knowledge Survey

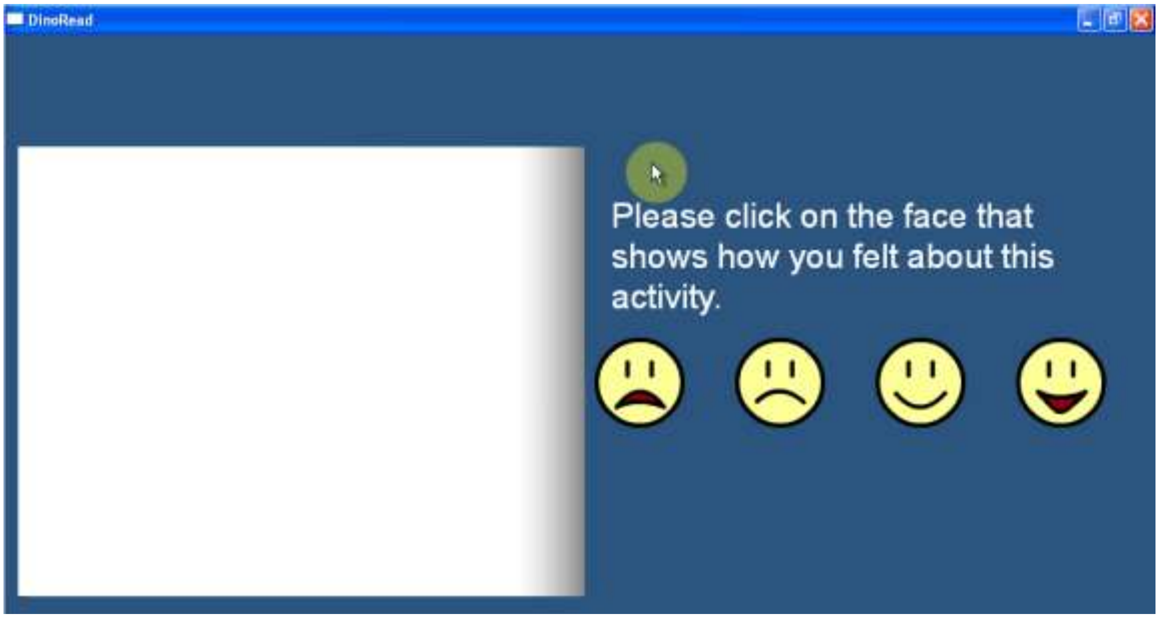
Each of these questions will be read aloud to you. For each answer, please **circle** YES or NO.

| | |
|---|---|
| 1) Have you ever used a **laptop computer**?  | |
| **YES** | **NO** |
| 2) Do you have a **laptop**  or **computer**  at home? | |
| **YES** | **NO** |
| 3) Have you ever used a **computer mouse**?  | |
| **YES** | **NO** |
| 4) Have you ever used **headphones**?  | |
| **YES** | **NO** |
| 5) Have you ever used a headset **microphone**?  | |
| **YES** | **NO** |

APPENDIX C

**User Satisfaction Survey "Happy Face" Form**

**User Satisfaction Survey "Happy Face" Form**




Adapted from "Usability and children's software: A user-centered design methodology," by Robertson, J. W., 1994, *Journal of Computing in Childhood Education, 5*(3/4)

Spoken audio:

"Please click on the face that shows how you felt about this activity."

Response buttons (audio spoken on mouse-over):

"I really didn't like it at all."

"I didn't like it that much."

"I liked it a little bit."

"I really liked it a lot."

APPENDIX D

# Interview Scripts

## Interview Scripts

Project: Development of CAI for First-Grade Reading
Time of Interview:
Date:
Place:
Interviewer:
Interviewee:
Position of Interviewee: Study Participant (first grade student)

**Design Phase 1 (audio) Questions/Script:**

I'm going to play some sounds for you through these headphones and ask you some questions about what you think of them. Is that okay?

1. Here is the first sound (play "weather" as Microsoft Anna synthesized speech) and here is the second sound (play "weather" as recorded adult female human voice). Which one do you like more?
Why?

2. Okay, we're going to listen to a couple more sounds. Here is the first one (play "birds don't have fur" as recorded human voice)... and here is the second one (play "birds don't have fur" as synthesized speech). Which one do you like more?
Why?

**Design Phase 2 (graphics) Questions/Script:**

I'm going to talk to you for a few minutes about the activity you did and I'm going to ask you some questions about it, okay?

1. Do you remember the three stories you read in the activity? What were they?

2. Which story was your favorite one to read? Why?

3. Did that story have any pictures? If so, were they moving pictures?

4. Which do you like the most: the story with no pictures, the story with the moving pictures, or the story with the regular pictures that didn't move?  Why?

**Design Phase 3 (ASR) Questions/Script:**

I'm going to talk to you for a few minutes about the activity you did and I'm going to ask you some questions about it, okay?

1. Which story was your favorite one to read? Why?

2. There were 3 different types of stories: one that didn't do anything when you read the words; one where you could read the words to make them change color but you didn't have to; and one where you *had* to read the words to go to the next page. Which one of those did you like doing the most?

Why?

**Design Phase 4 (lesson/practice) Questions/Script:**

I'm going to talk to you for a few minutes about the activity you did and I'm going to ask you some questions about it, okay?

1. Which story did you like the most? Why?
2. Did you like the game where you click on the words? Why?
3. Did you like the game where you had to say the words out loud? Why?

APPENDIX E

# Needs Assessment Survey

# Needs Assessment Survey

**1. Of the reading skills listed below, how would you rate their importance for inclusion in a computer program to help first graders learn how to read?**

| | A little bit useful | | | | Very useful / important |
|---|---|---|---|---|---|
| Vocabulary | ○ | ○ | ○ | ○ | ○ |
| Rhyming Words | ○ | ○ | ○ | ○ | ○ |
| Phonetic decoding skills (blending, segmentation, onset/rime, digraphs, etc.) | ○ | ○ | ○ | ○ | ○ |
| Oral reading fluency | ○ | ○ | ○ | ○ | ○ |
| Reading comprehension questions | ○ | ○ | ○ | ○ | ○ |

**2. Please rate each of these features based on how important or useful you think they would be in a reading program for the computer:**

| | Not very important / useful | | Indifferent / Not Sure | | Very important / useful |
|---|---|---|---|---|---|
| Tutorials to teach concepts/skills | ○ | ○ | ○ | ○ | ○ |
| Games to practice individual word or phonetic skills | ○ | ○ | ○ | ○ | ○ |
| E-books to practice reading level-appropriate stories | ○ | ○ | ○ | ○ | ○ |
| Speech recognition to assess pronunciation and fluency | ○ | ○ | ○ | ○ | ○ |
| Targeted interventions to scaffold problem areas | ○ | ○ | ○ | ○ | ○ |
| Audio support for word pronunciation (word spoken when clicked) | ○ | ○ | ○ | ○ | ○ |
| Visual vocabulary support (picture of word when clicked) | ○ | ○ | ○ | ○ | ○ |

Other Recommendations?

[ ]

**3. Which method of pacing and control would you prefer in a reading program for first graders (if you think a combination of these would work, select more than one to represent which choices you would like to use)**

- [ ] The program allows students full control of selecting activities, lessons, or ebooks
- [ ] The program uses a strict sequence of skills, activities, lessons and ebooks for all students to progress through
- [ ] The program provides practice and lessons based by determining areas of difficulty for the student
- [ ] The program allows a teacher or administrator to limit or assign the lessons, activities, and/or ebooks

**4. Which method(s) of audio support & feedback do you think would be best?**

- [ ] Pronounce word when clicked/touched
- [ ] Segmented sound-out of each phoneme/sound when word is clicked/touched
- [ ] Read aloud entire sentence when clicked/touched
- [ ] Automatically read-aloud the entire page
- [ ] Allow user to click button for read-aloud of page

**5. Please rate how helpful these types of data and feedback about student performance would be:**

| | Not so important or useful | | Somewhat helpful | | Very helpful / important |
|---|---|---|---|---|---|
| Fluency (words per minute) | ○ | ○ | ○ | ○ | ○ |
| Pronunciation / oral reading errors | ○ | ○ | ○ | ○ | ○ |
| Word assists (words clicked by student) | ○ | ○ | ○ | ○ | ○ |
| Words read (list of words encountered by student) | ○ | ○ | ○ | ○ | ○ |
| Reports of phonetic decoding difficulties | ○ | ○ | ○ | ○ | ○ |
| Vocabulary matching / recognition | ○ | ○ | ○ | ○ | ○ |
| Comprehension questions answered correctly | ○ | ○ | ○ | ○ | ○ |
| Time spent per book / page / word | ○ | ○ | ○ | ○ | ○ |
| Audio recording of student reading | ○ | ○ | ○ | ○ | ○ |
| Interventions / lessons / tutorials provided to student | ○ | ○ | ○ | ○ | ○ |

**6. Are there any other features you would like to see included in a reading program for first graders?**

Done

APPENDIX F

# Spoken Audio Feedback Form

**Spoken Audio Feedback Form**

For each number, you will hear 2 sounds: sound A and sound B. Circle the letter of the *sound you like more*.

| | |
|---|---|
| **1. rake**<br>Which sound do you like more? A or B<br>**A** **B** | **2. with**<br>Which sound do you like more? A or B<br>**A** **B** |
| **3. since**<br>Which sound do you like more? A or B<br>**A** **B** | **4. that**<br>Which sound do you like more? A or B<br>**A** **B** |
| **5. sprinkler**<br>Which sound do you like more? A or B<br>**A** **B** | **6. flowers**<br>Which sound do you like more? A or B<br>**A** **B** |
| **7. painted**<br>Which sound do you like more? A or B<br>**A** **B** | **8. window**<br>Which sound do you like more? A or B<br>**A** **B** |

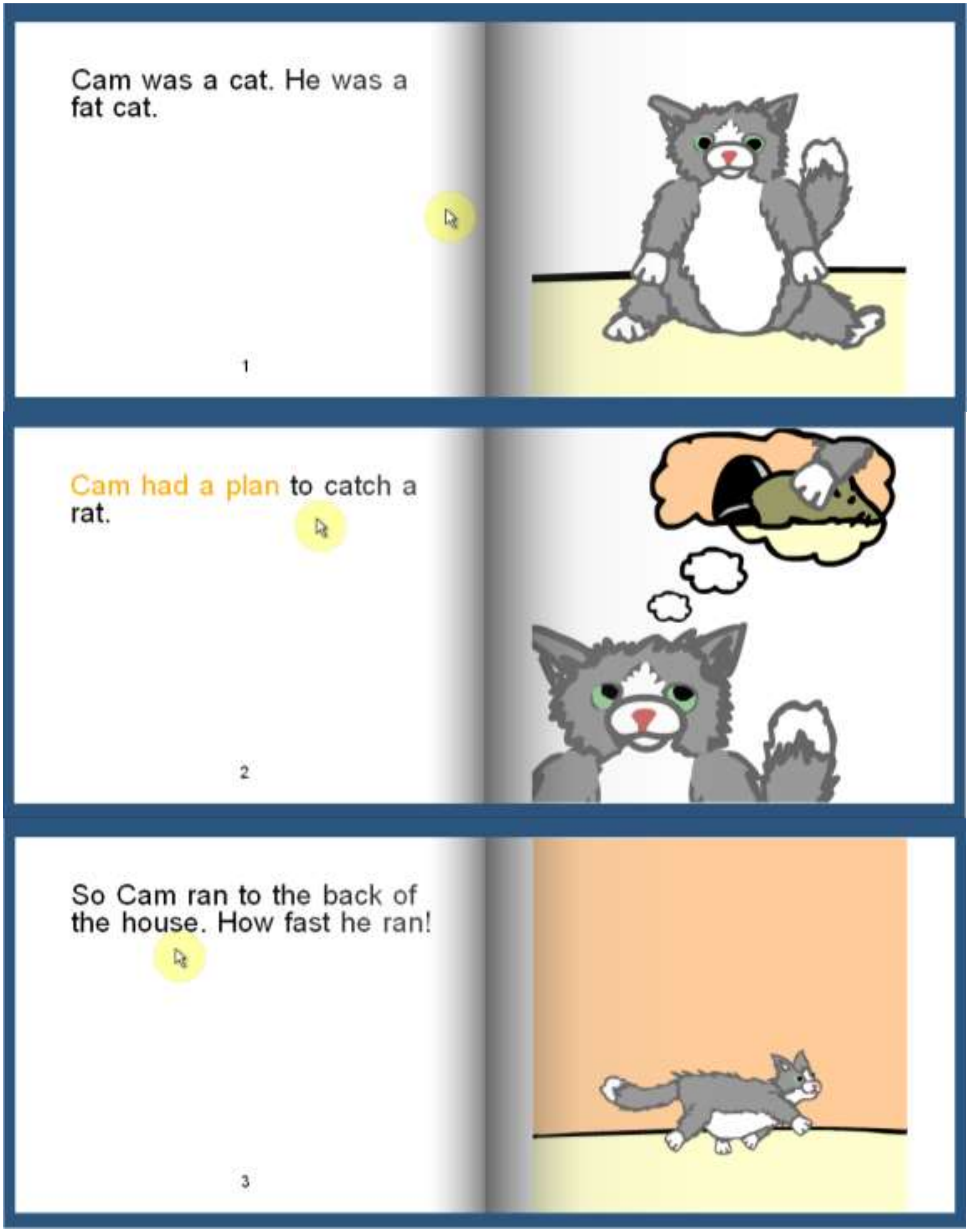
For each number, you will hear 2 sounds: sound A and sound B. Circle the letter of the *sound that is more helpful*.

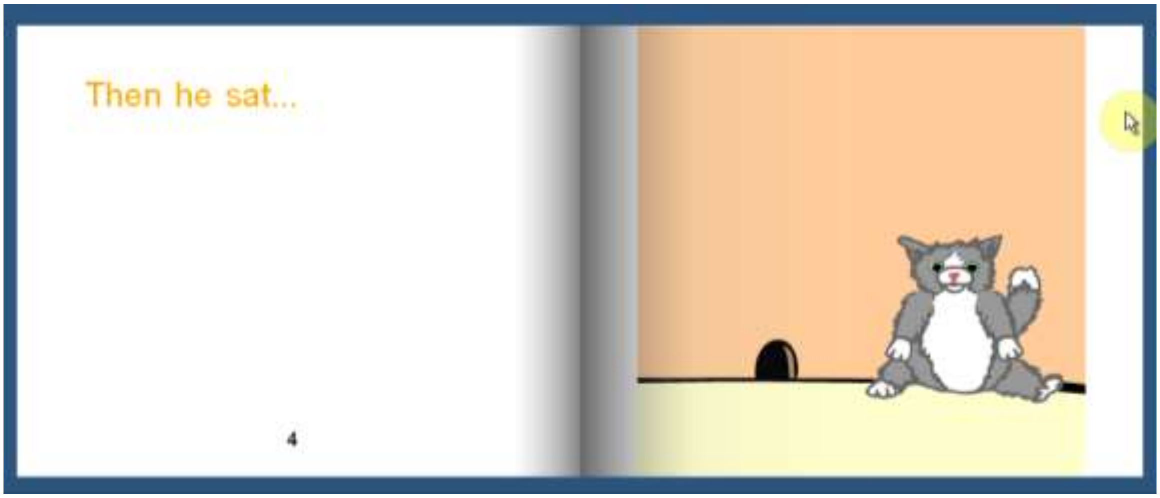
| | |
|---|---|
| **9. blue**<br>Which sound is more helpful? A or B<br>**A** **B** | **10. that's**<br>Which sound is more helpful? A or B<br>**A** **B** |
| **11. green**<br>Which sound do you like more? A or B<br>**A** **B** | **12. don't**<br>Which sound is more helpful? A or B<br>**A** **B** |
| **13. weather**<br>Which sound is more helpful? A or B<br>**A** **B** | **14. waiting**<br>Which sound is more helpful? A or B<br>**A** **B** |
| **15. maybe**<br>Which sound is more helpful? A or B<br>**A** **B** | **16. outside**<br>Which sound is more helpful? A or B<br>**A** **B** |

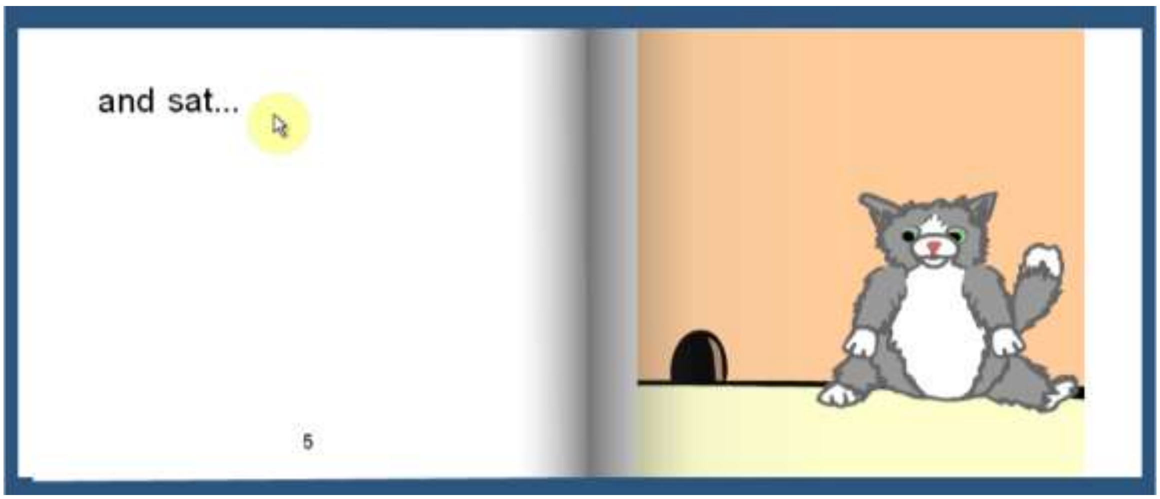
APPENDIX G

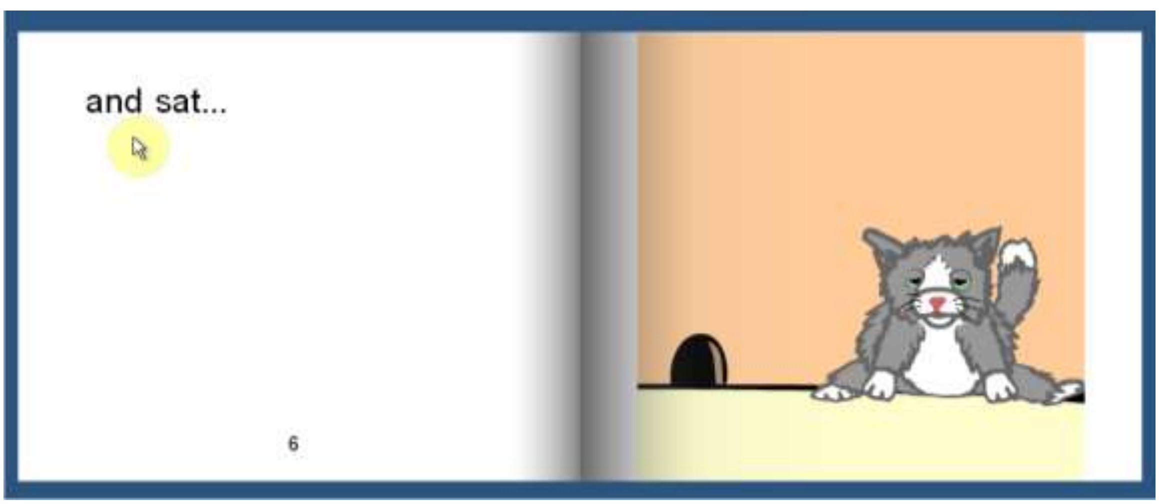
# Visuals Test Case E-Book Screens

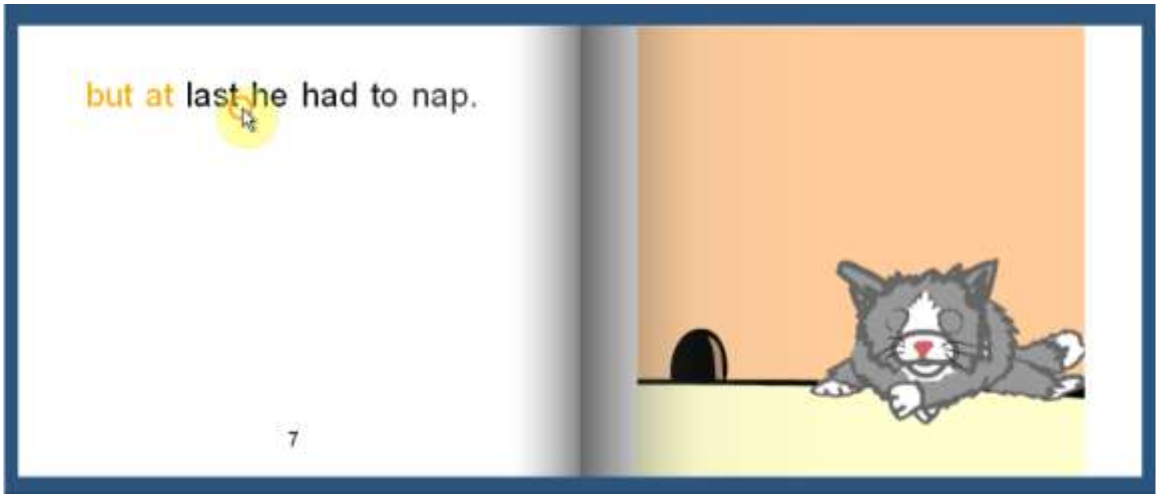
**Visuals Test Case E-Book Screens**

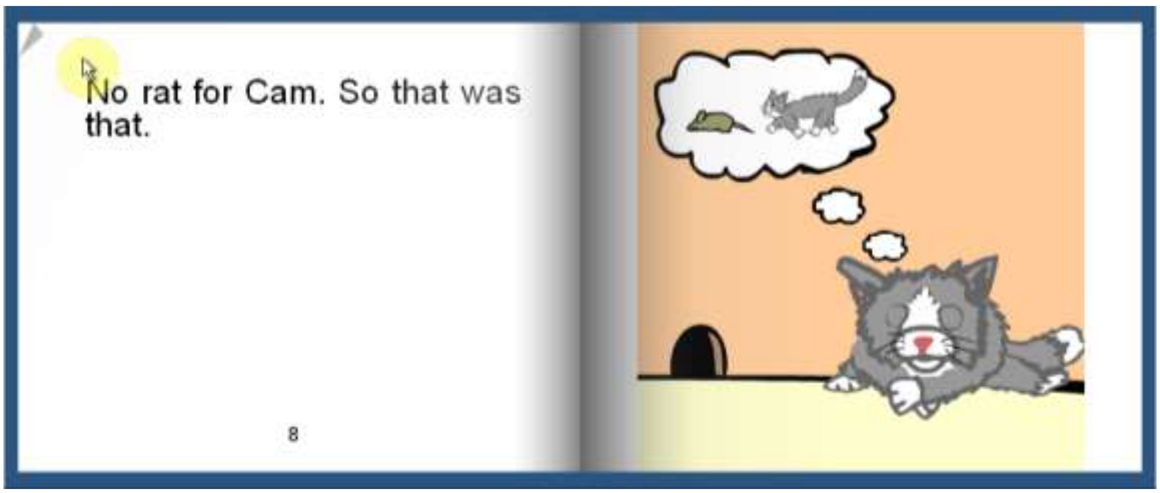


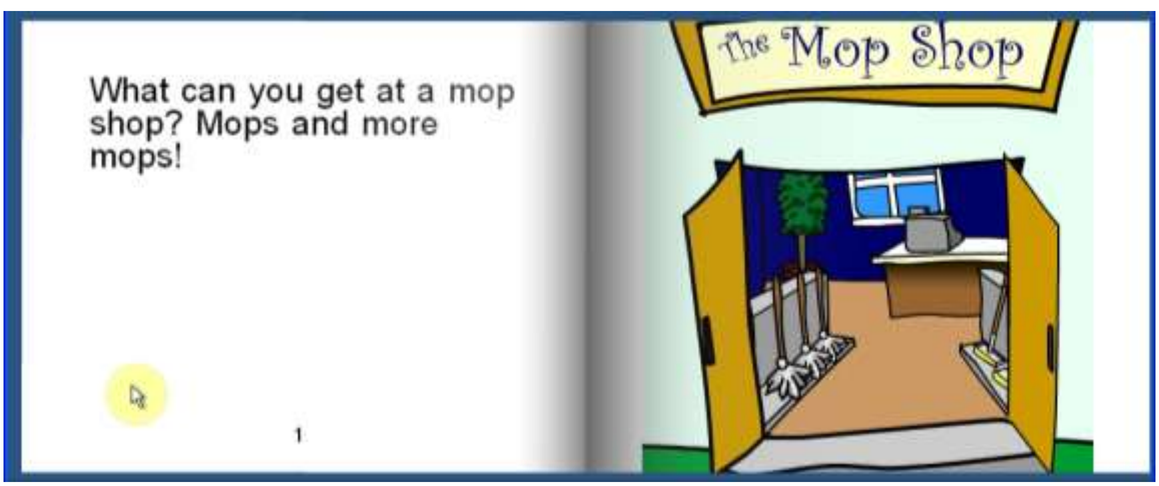

Then he sat...
4

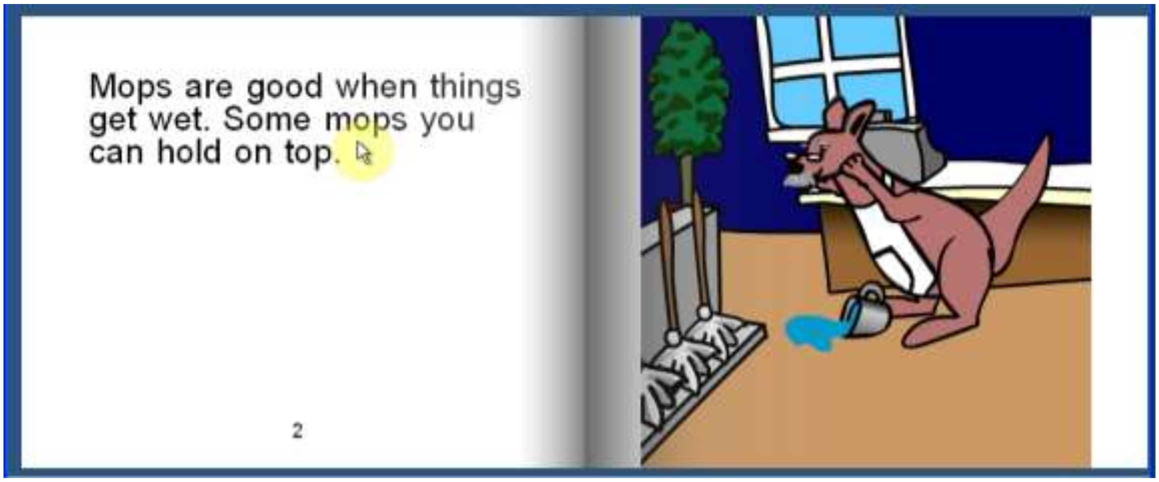

and sat...
5

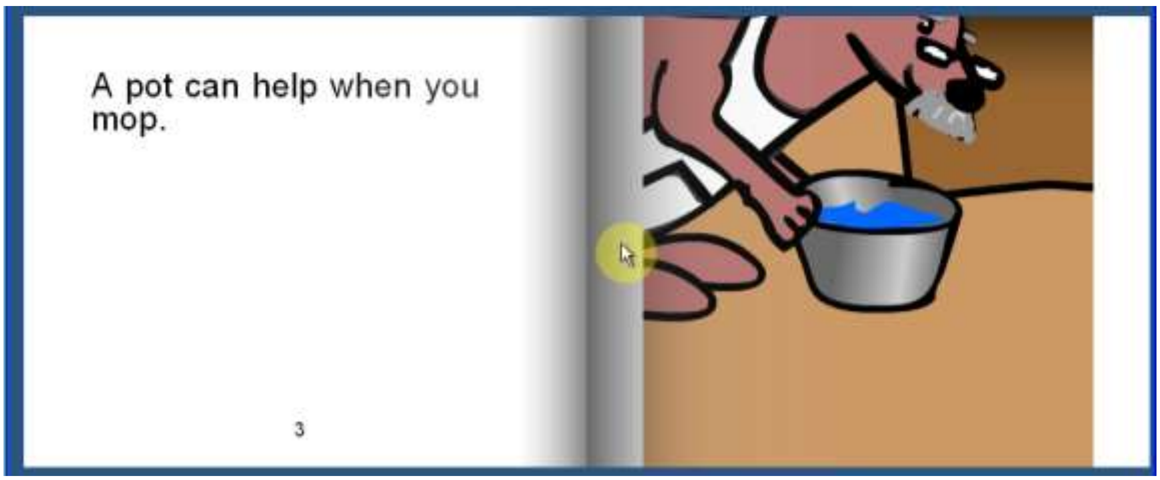

and sat...
6

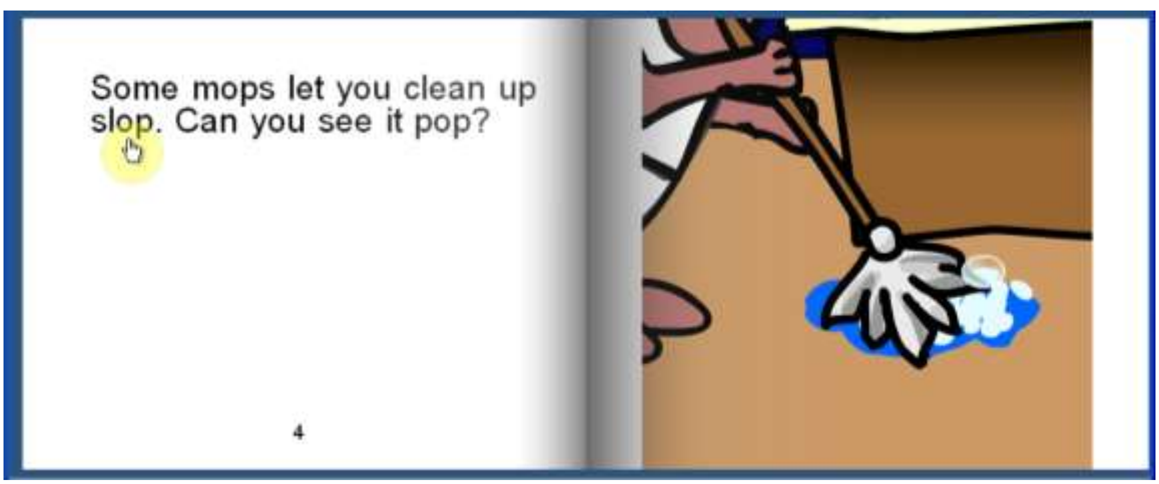

but at last he had to nap.
7

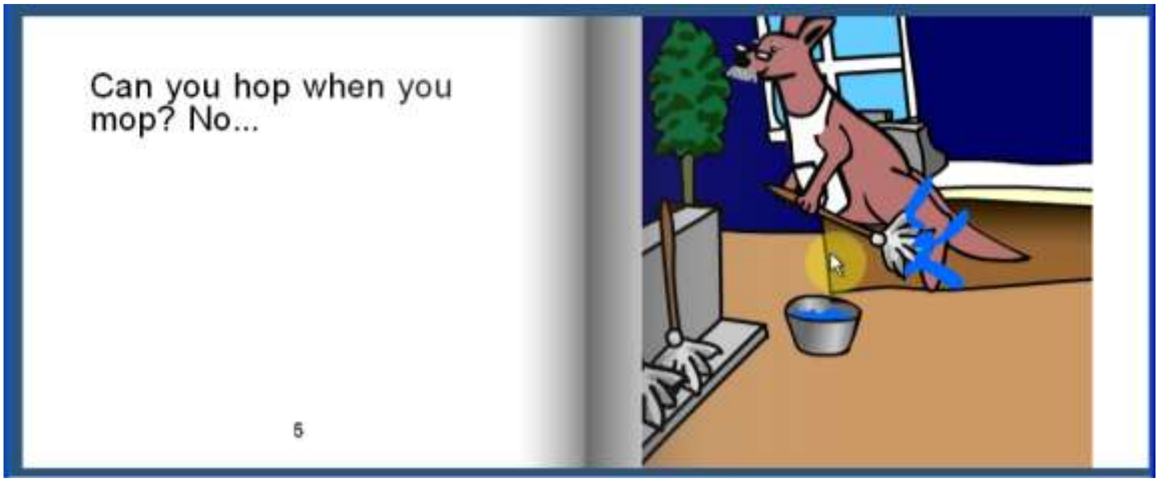

No rat for Cam. So that was that.
8

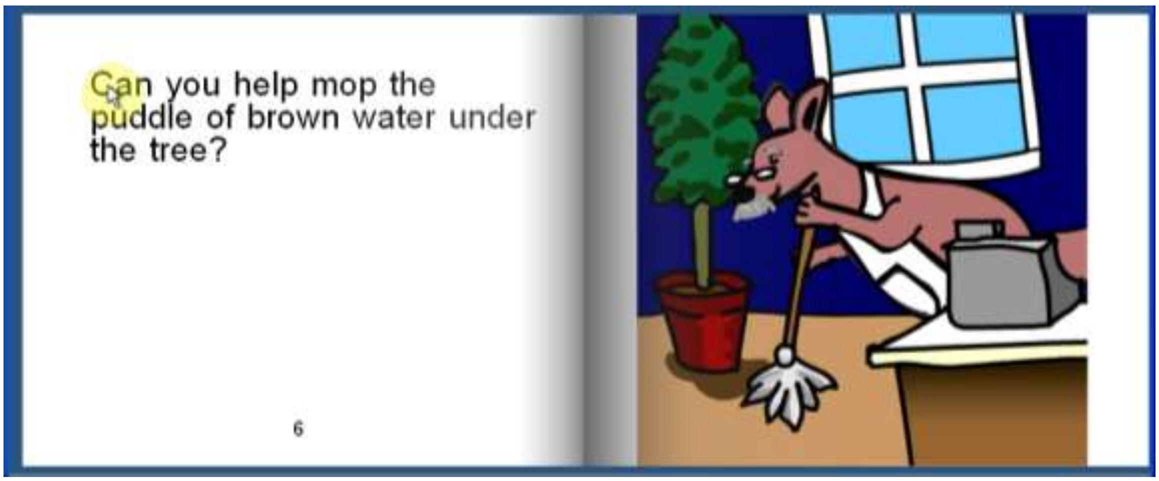

What can you get at a mop shop? Mops and more mops!
1
The Mop Shop

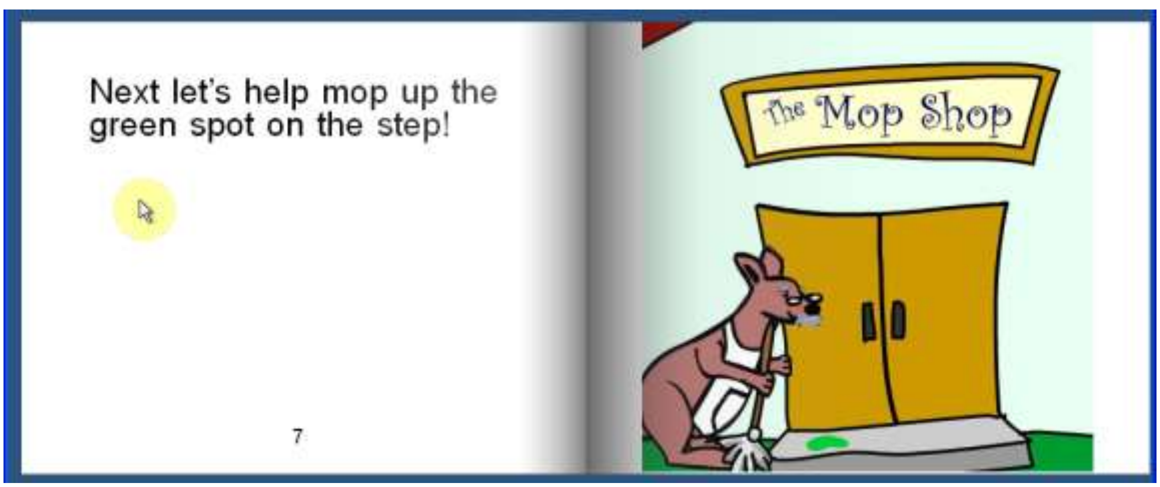

Mops are good when things get wet. Some mops you can hold on top.
2

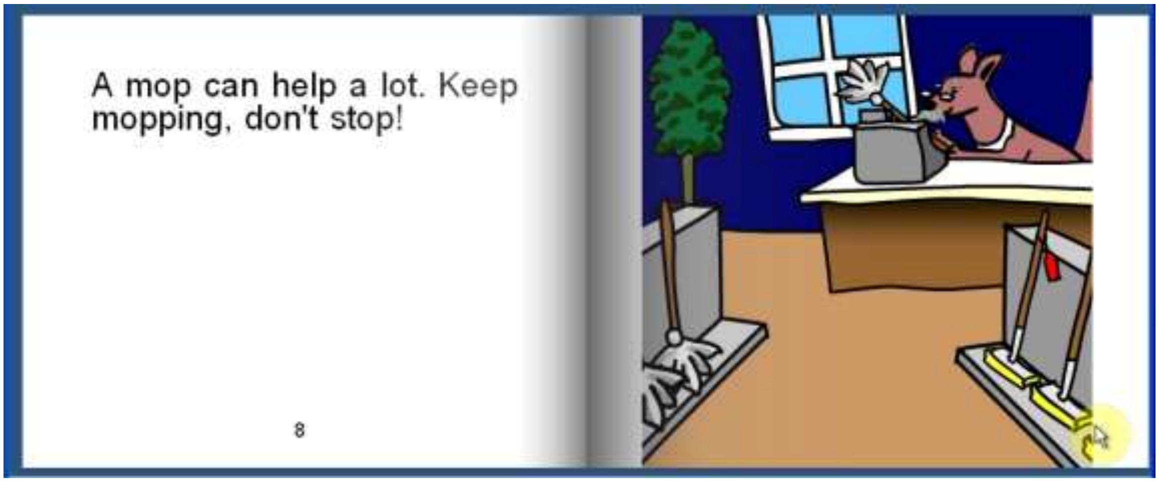

A pot can help when you mop.
3

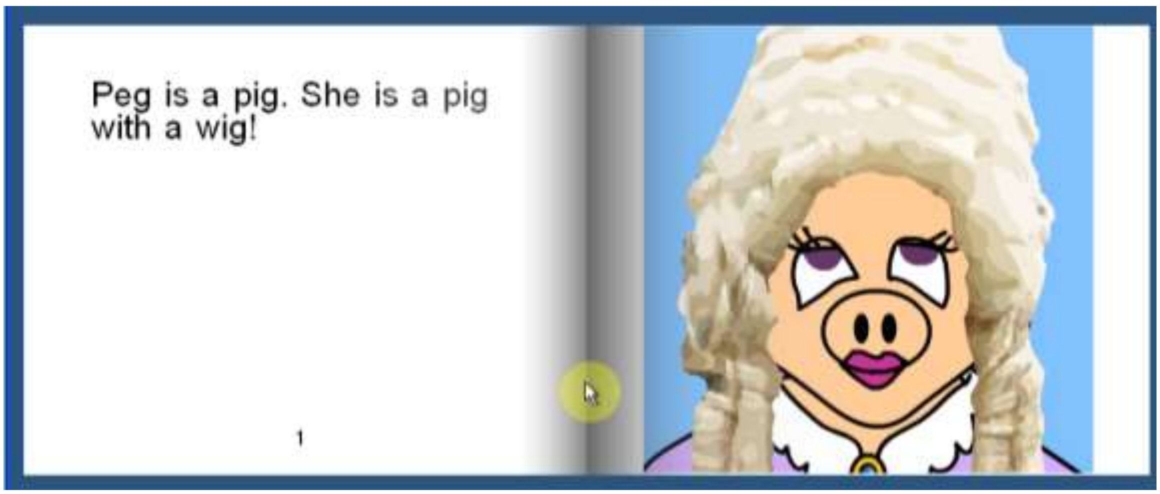

Some mops let you clean up slop. Can you see it pop?
4

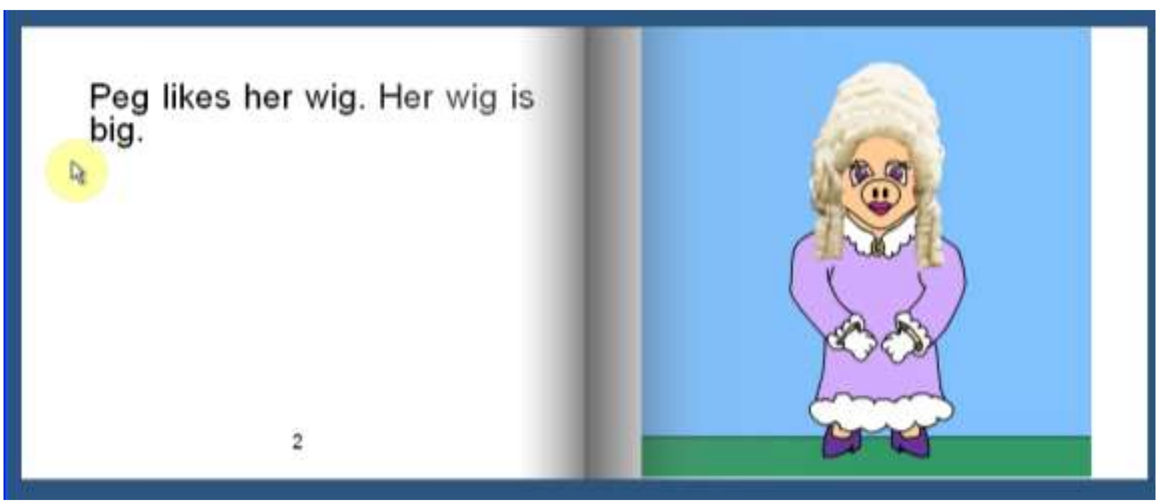

Can you hop when you mop? No...
5

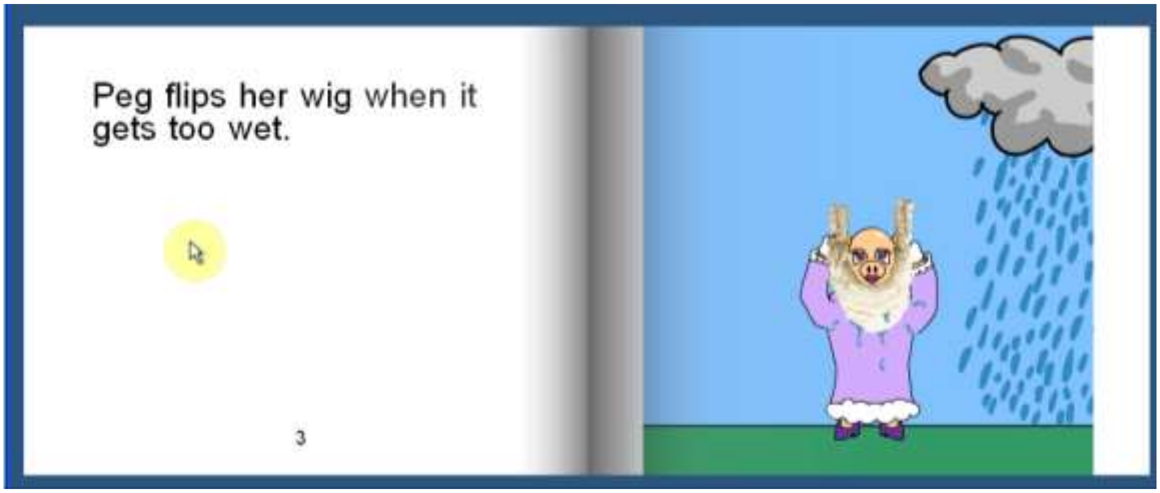

Can you help mop the puddle of brown water under the tree?
6

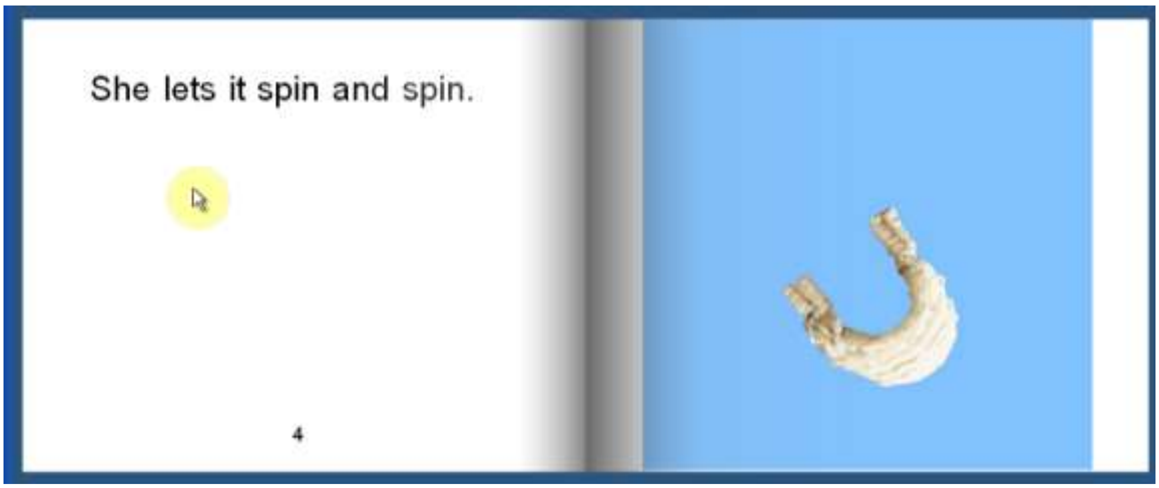

Next let's help mop up the green spot on the step!
7
The Mop Shop

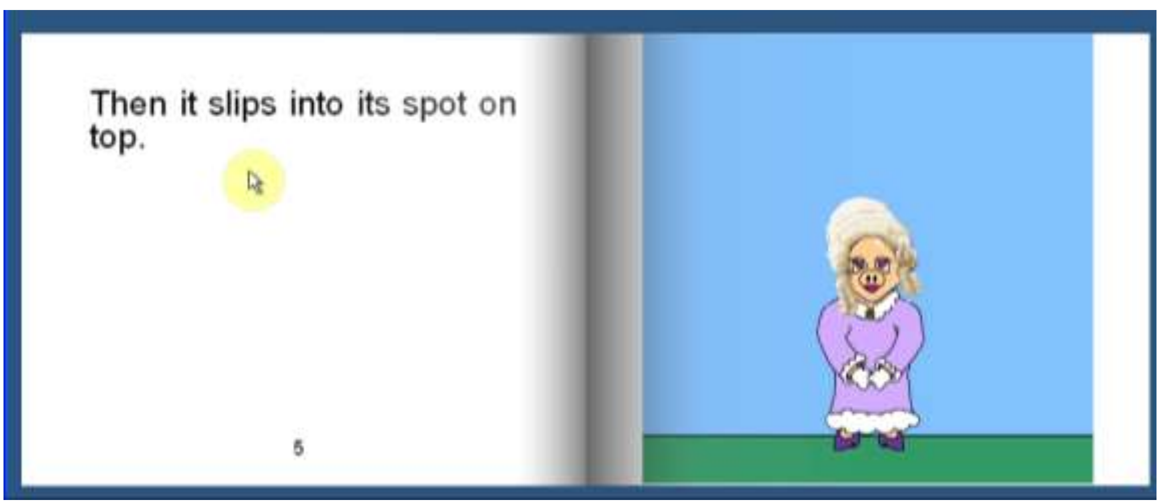

A mop can help a lot. Keep mopping, don't stop!
8

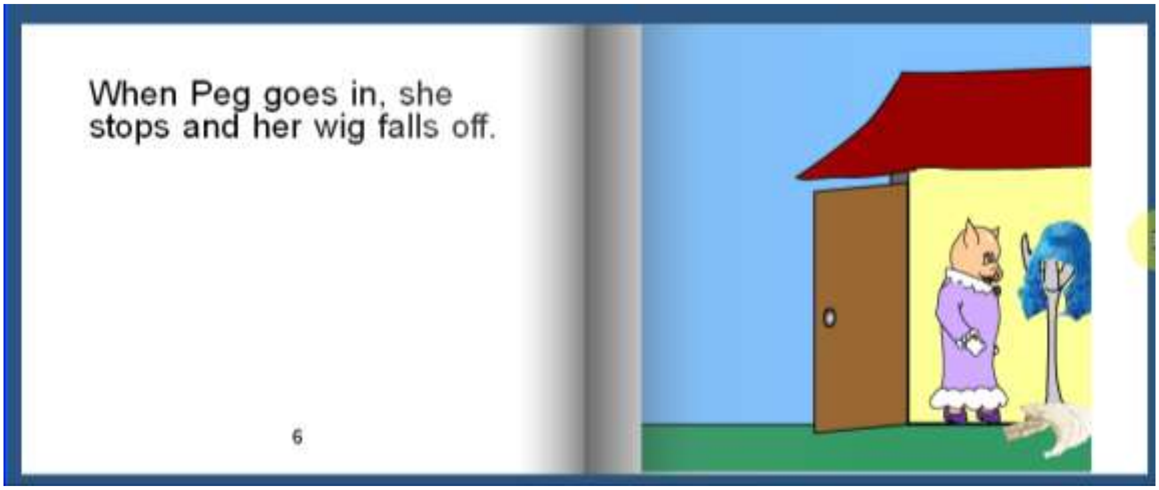

Peg is a pig. She is a pig with a wig!
1

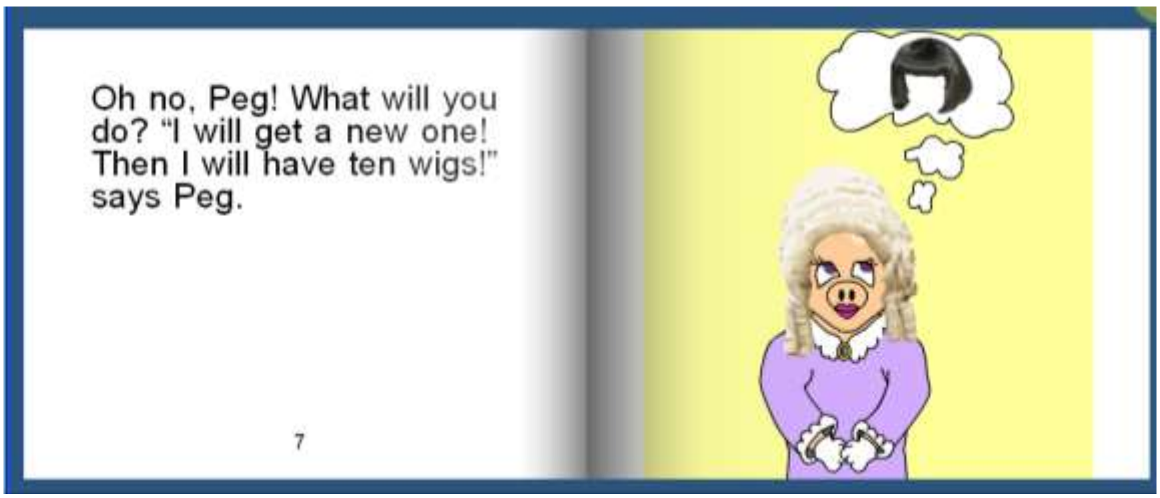

Peg likes her wig. Her wig is big.
2


Peg flips her wig when it gets too wet.
3


She lets it spin and spin.
4


Then it slips into its spot on top.
5


When Peg goes in, she stops and her wig falls off.
6

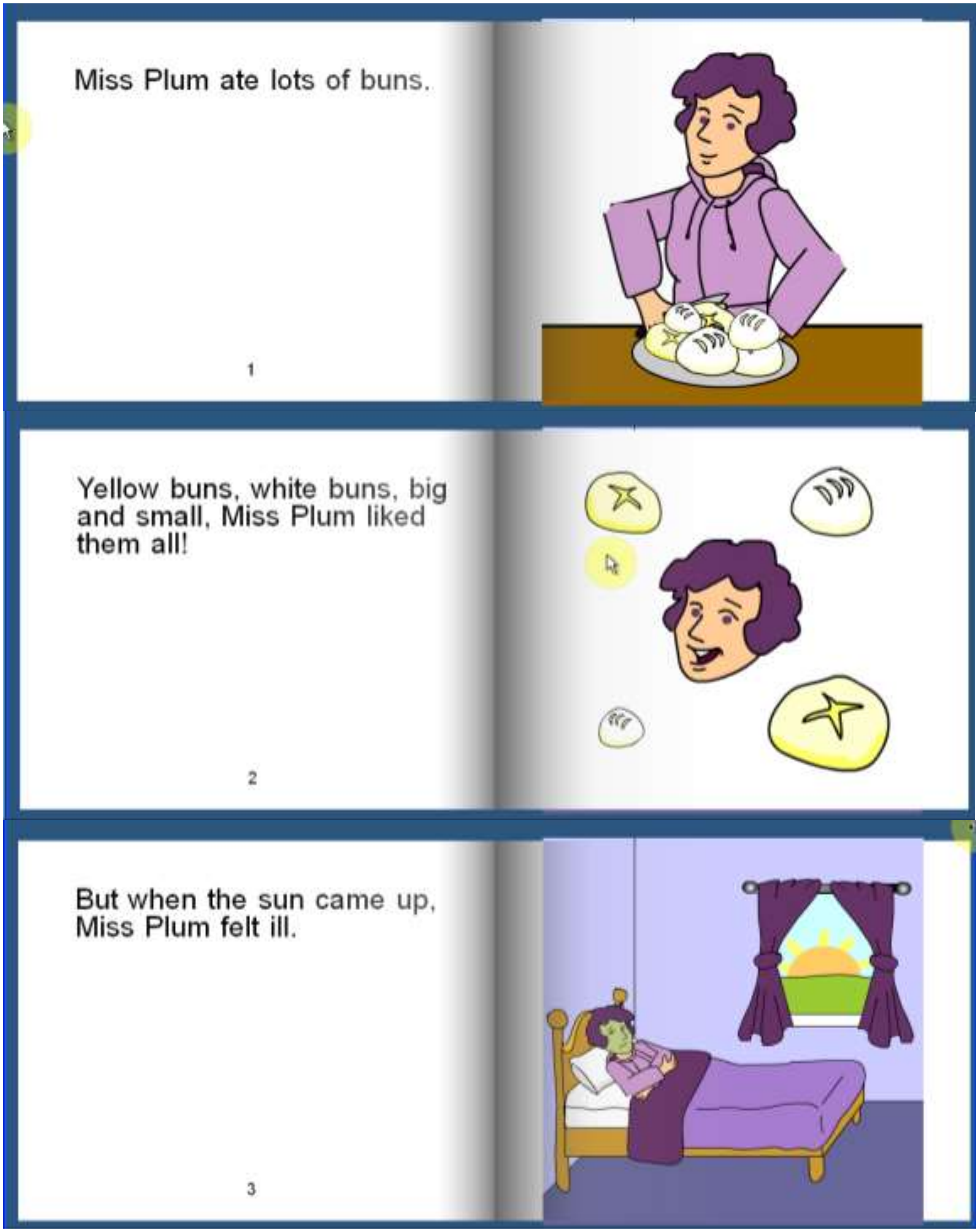

Oh no, Peg! What will you do? "I will get a new one! Then I will have ten wigs!" says Peg.
7

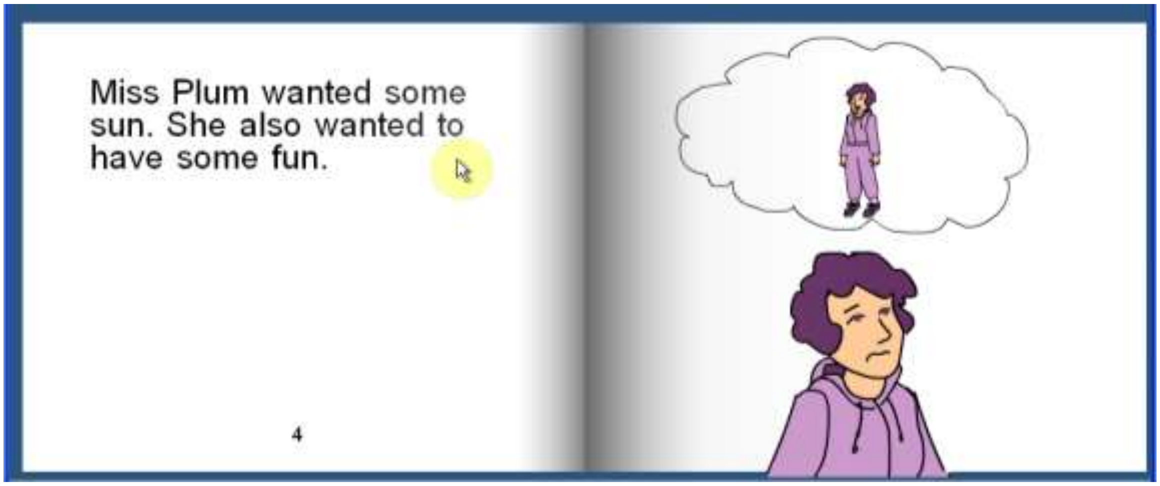

Red wigs, blue wigs, short and tall. Can you help her find them all?
8

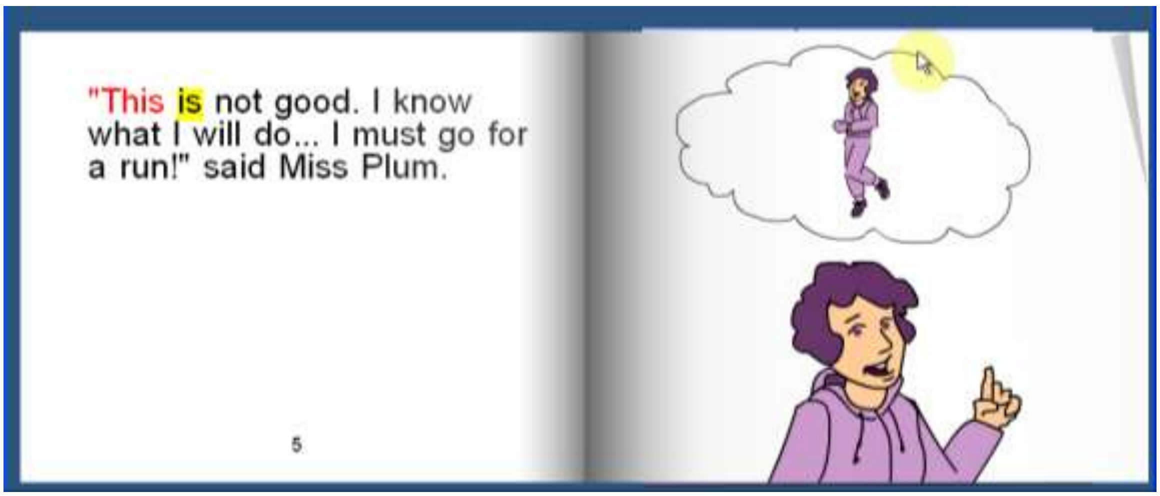
APPENDIX H

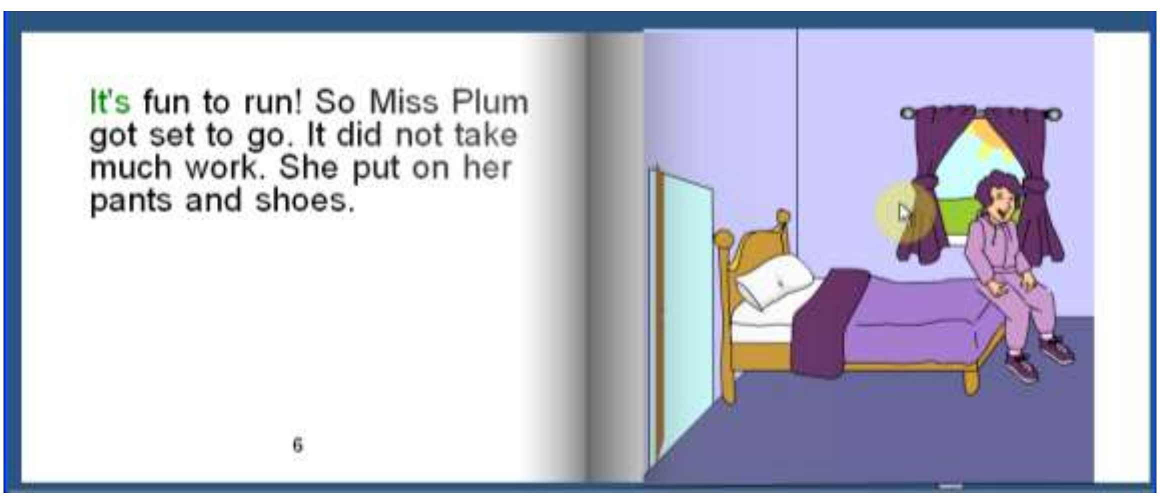
# ASR Test Case E-Book Screens

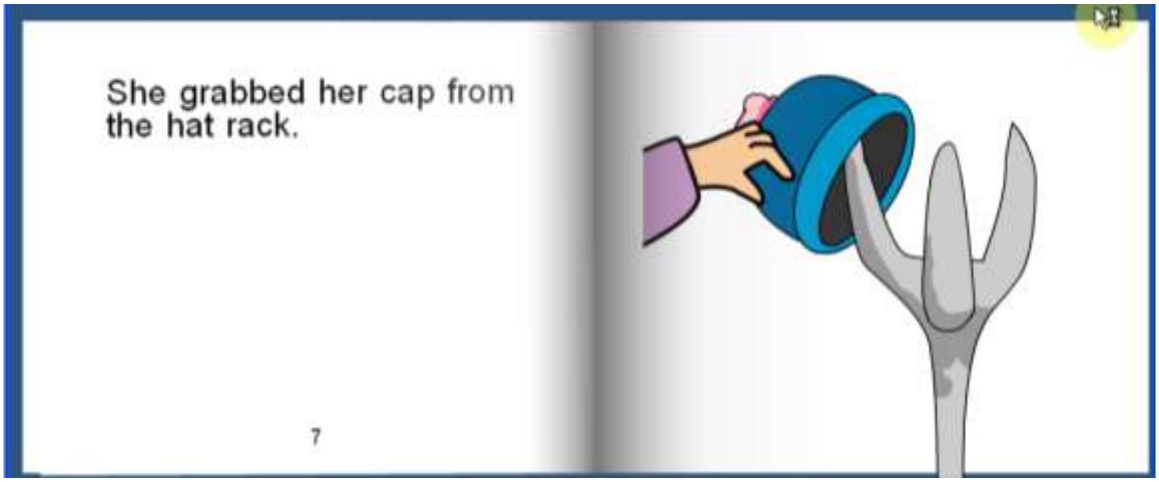
**ASR Test Case E-Book Screens**

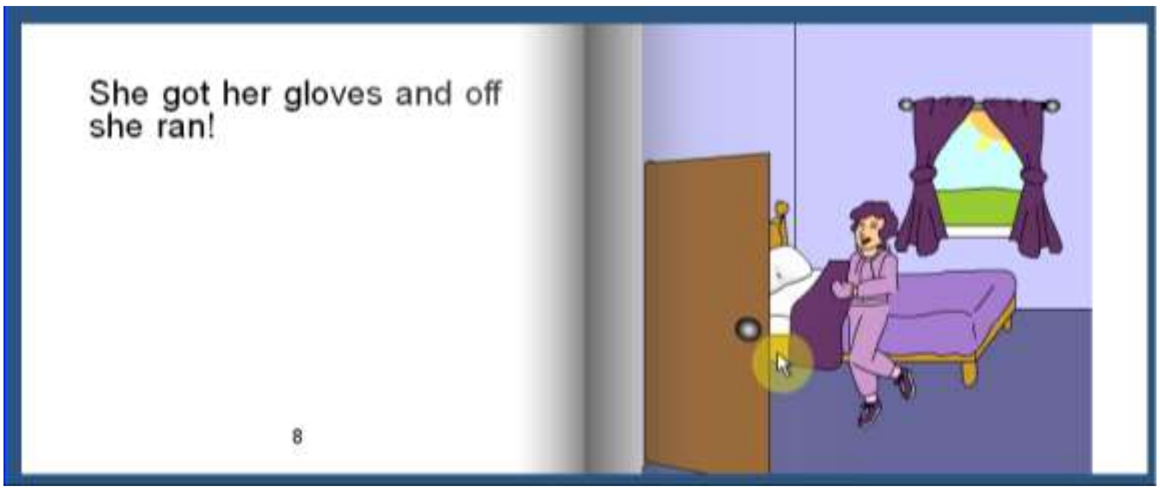


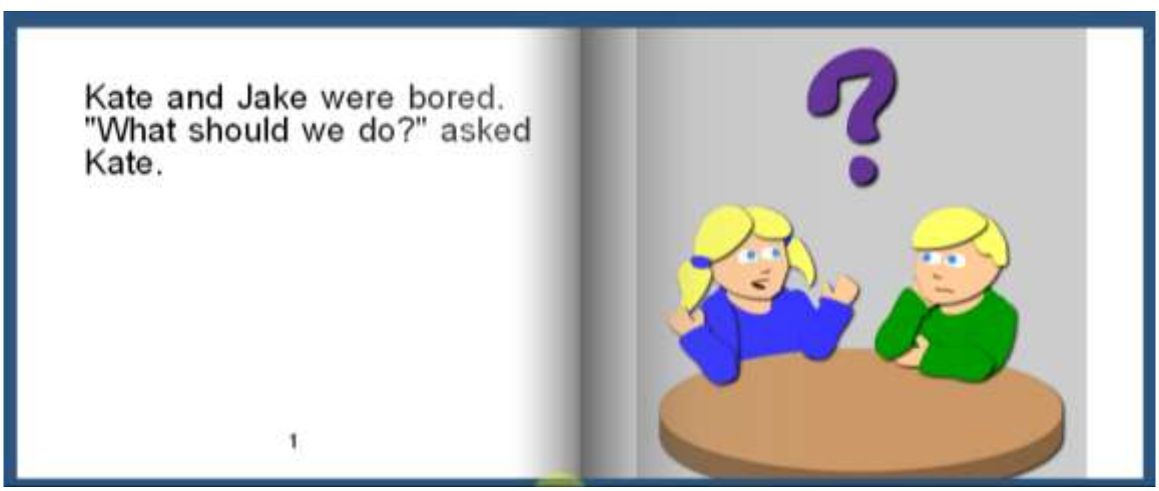

Miss Plum wanted some sun. She also wanted to have some fun.
4

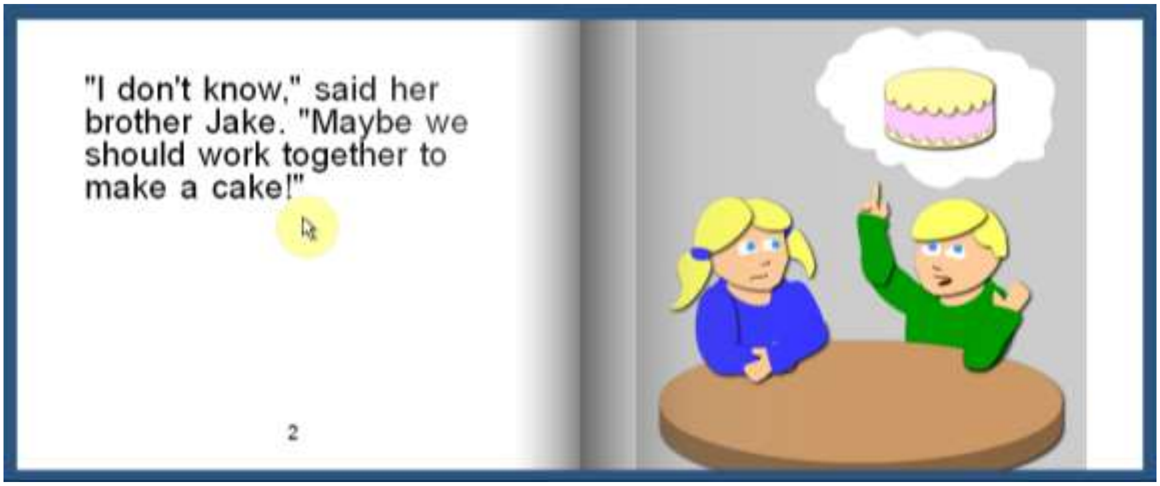

"This is not good. I know what I will do... I must go for a run!" said Miss Plum.
5

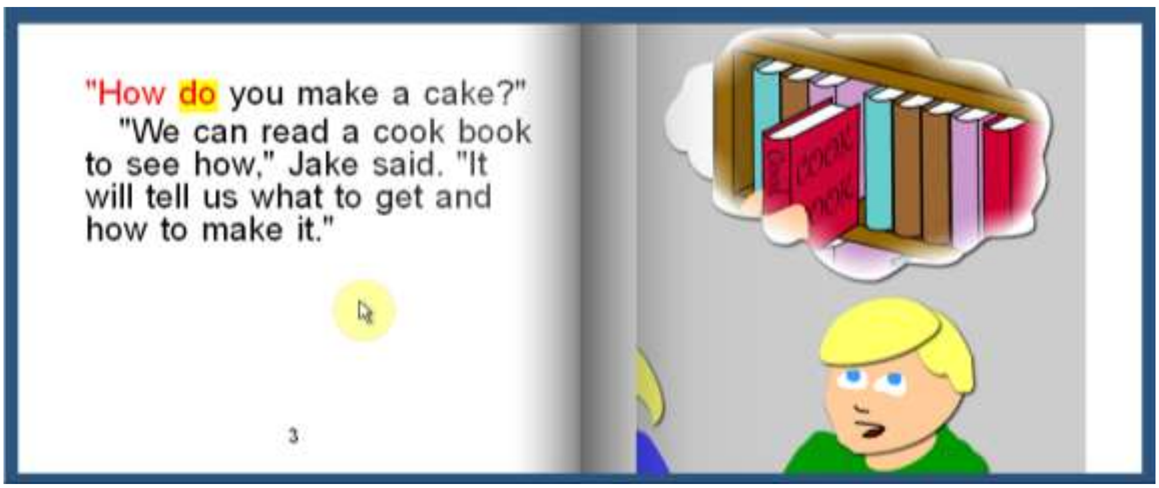

It's fun to run! So Miss Plum got set to go. It did not take much work. She put on her pants and shoes.
6

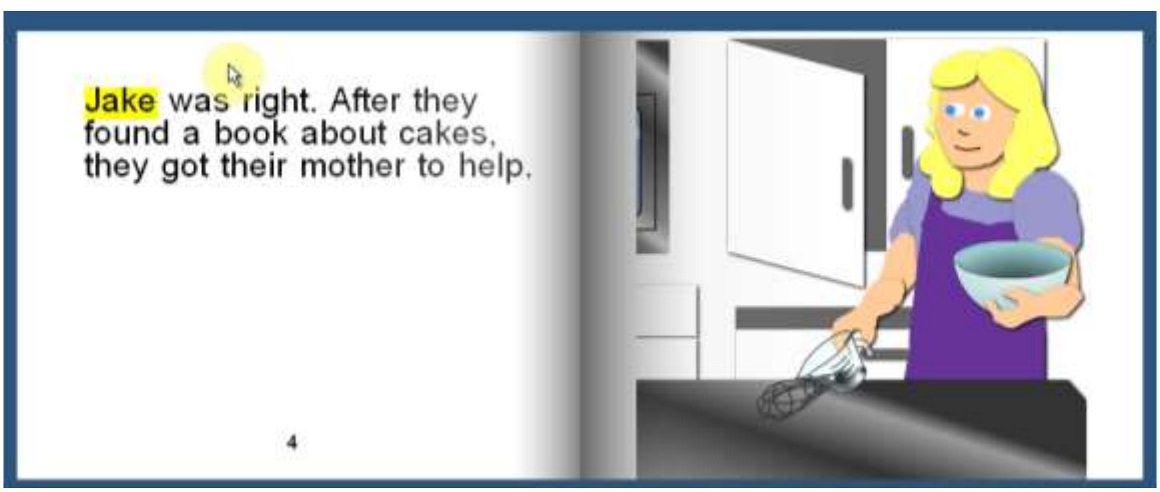

She grabbed her cap from
the hat rack.
7

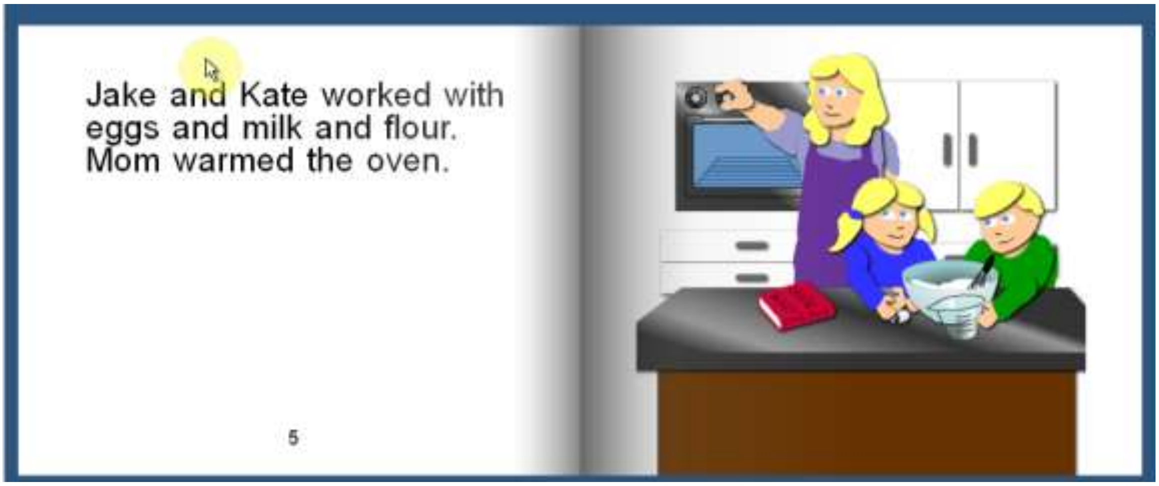

She got her gloves and off
she ran!
8

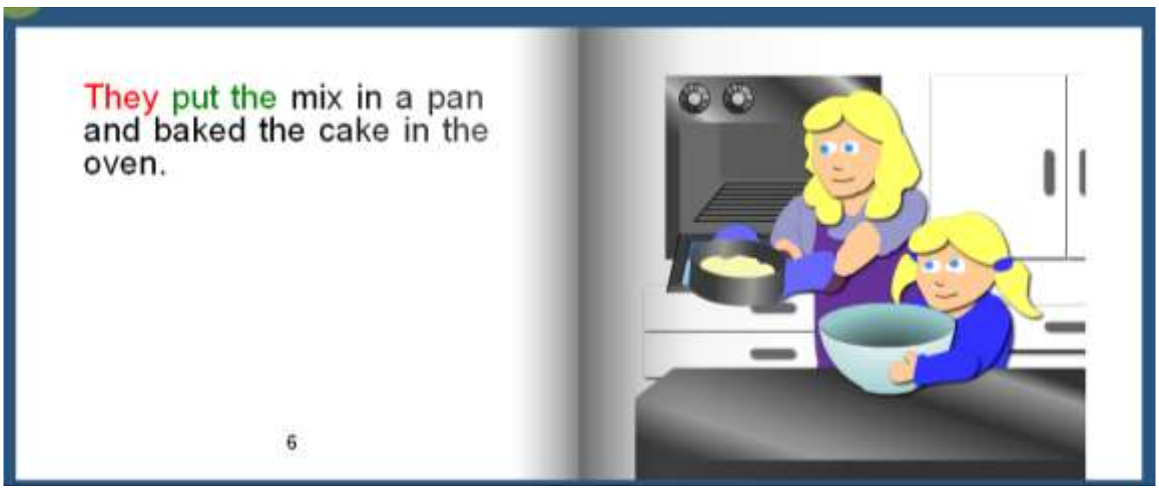

Kate and Jake were bored.
"What should we do?" asked
Kate.
1

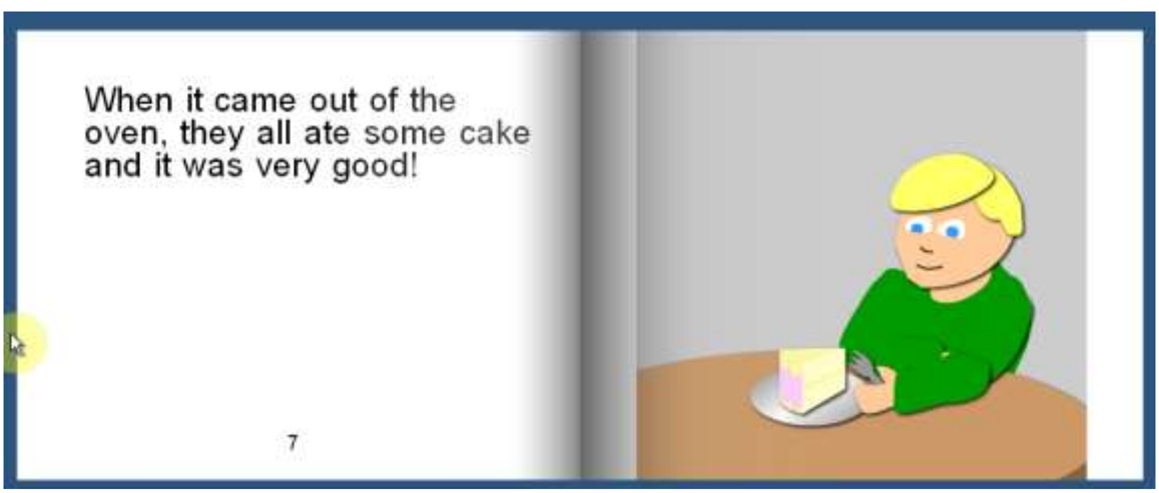

"I don't know," said her brother Jake. "Maybe we should work together to make a cake!"
2

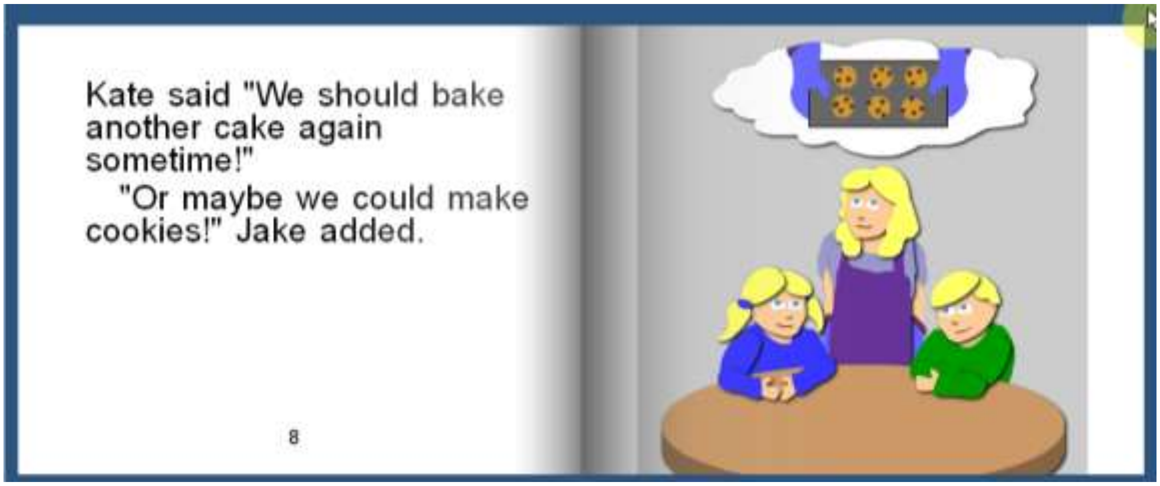

"How do you make a cake?"
"We can read a cook book to see how," Jake said. "It will tell us what to get and how to make it."
3


Jake was right. After they found a book about cakes, they got their mother to help.
4


Jake and Kate worked with
eggs and milk and flour.
Mom warmed the oven.
5


They put the mix in a pan
and baked the cake in the
oven.
6

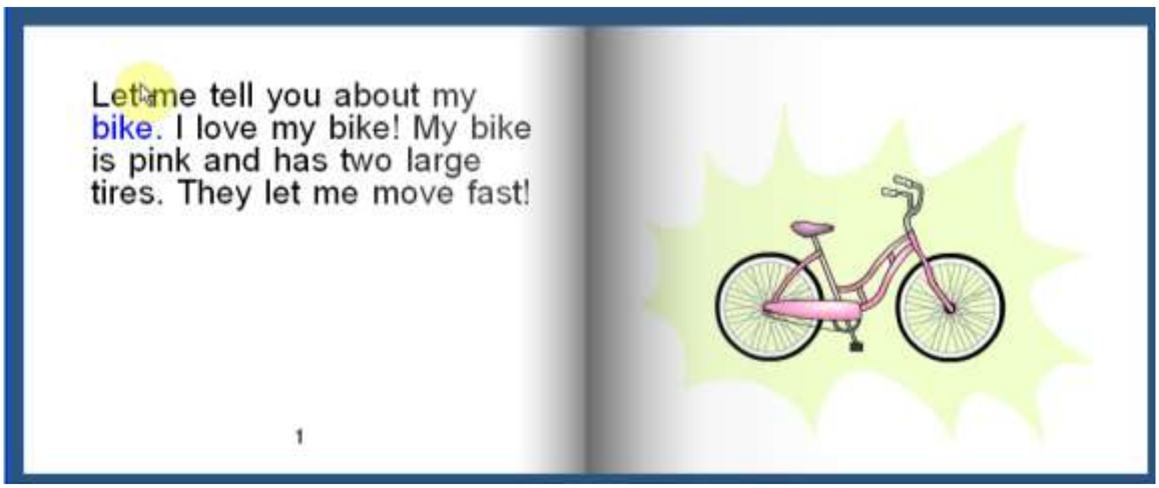

When it came out of the
oven, they all ate some cake
and it was very good!
7



Kate said "We should bake another cake again sometime!"

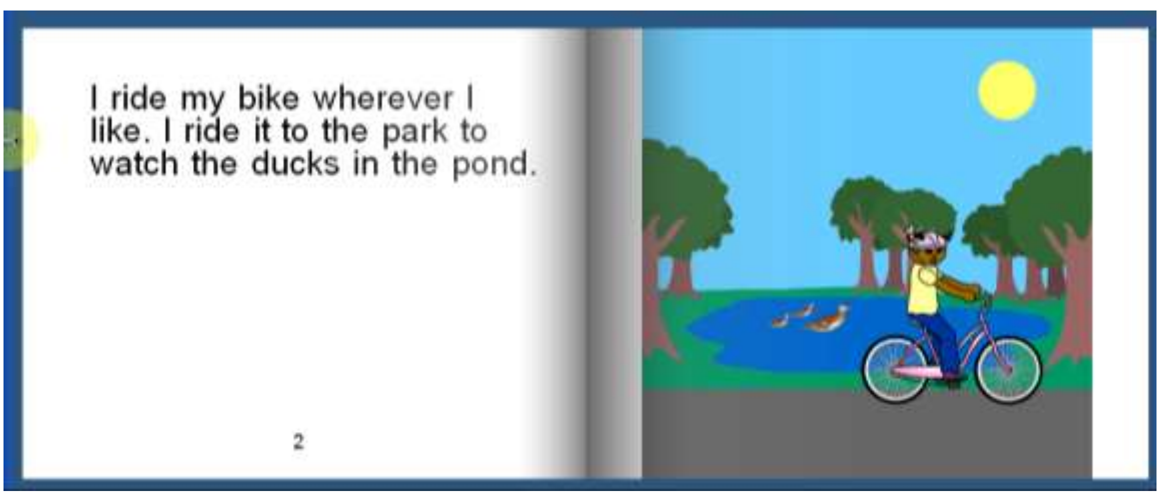
"Or maybe we could make cookies!" Jake added.

8



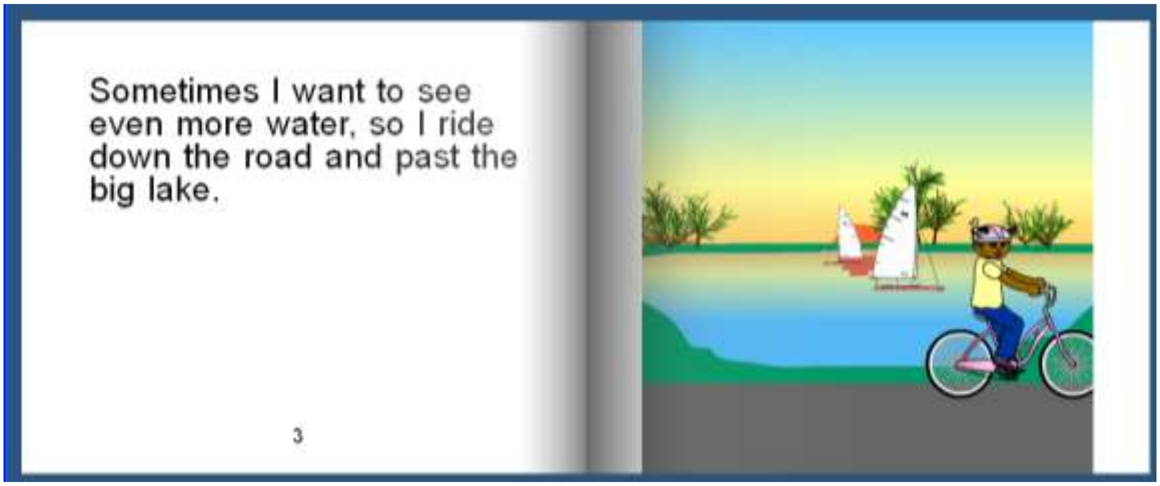
Let me tell you about my bike. I love my bike! My bike is pink and has two large tires. They let me move fast!

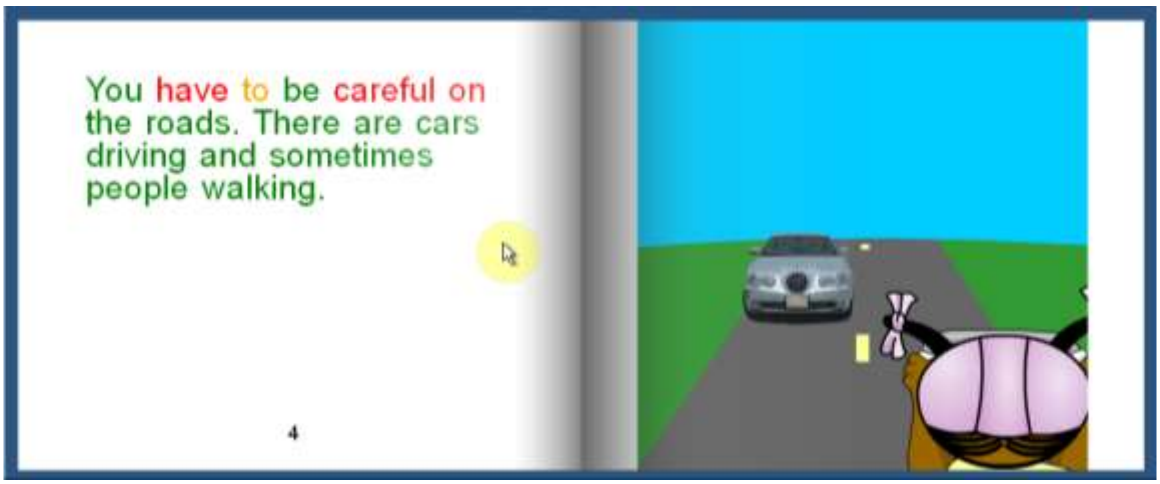
1

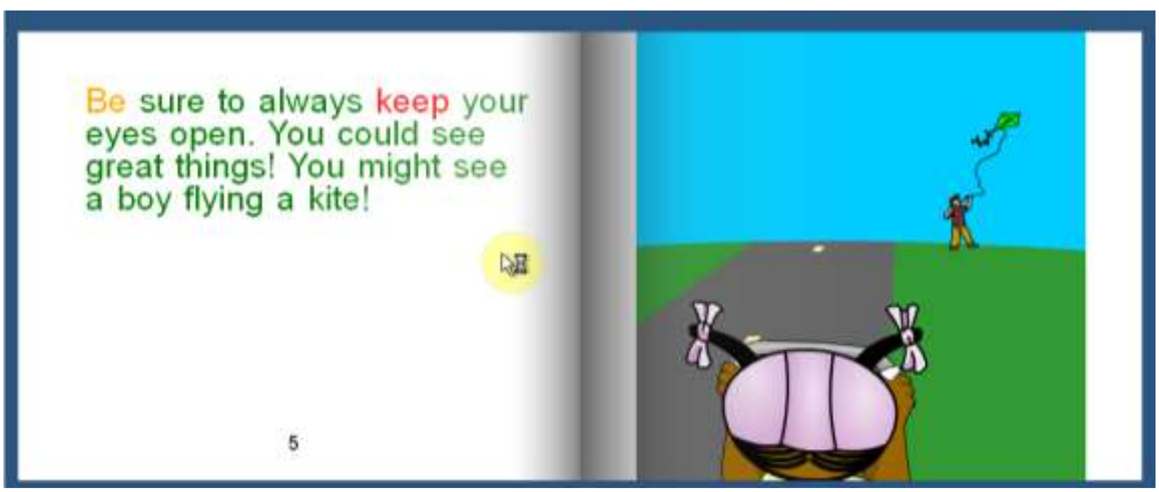


I ride my bike wherever I like. I ride it to the park to watch the ducks in the pond.

2



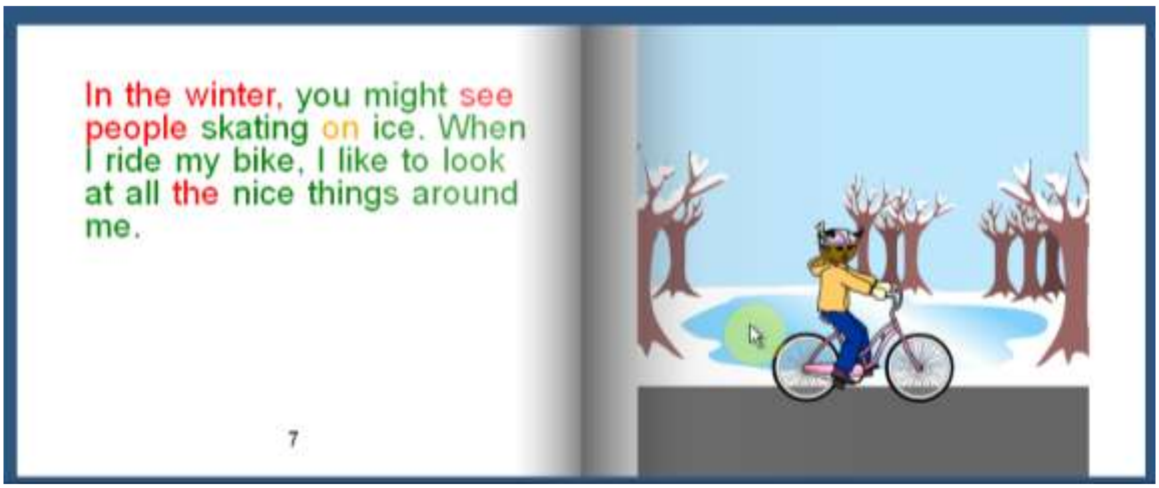



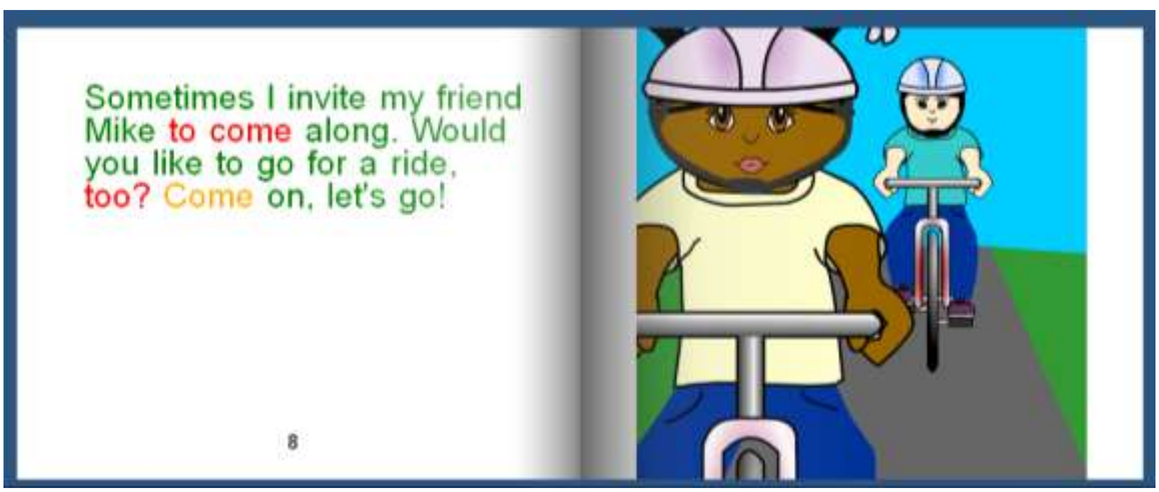






(Jaguar S-Type image by Sfoskett: http://commons.wikimedia.org/wiki/User:Sfoskett )


In the fall, you can look at all the leaves. Red leaves, orange leaves, brown leaves and yellow leaves. So many colors!
6


In the winter, you might see people skating on ice. When I ride my bike, I like to look at all the nice things around me.
7


Sometimes I invite my friend Mike to come along. Would you like to go for a ride, too? Come on, let's go!
8

APPENDIX I

# Tutorial Screens

**Tutorial Screens**



APPENDIX J

# Word Identification and Synthesis Practice

**Word Identification and Synthesis Practice**

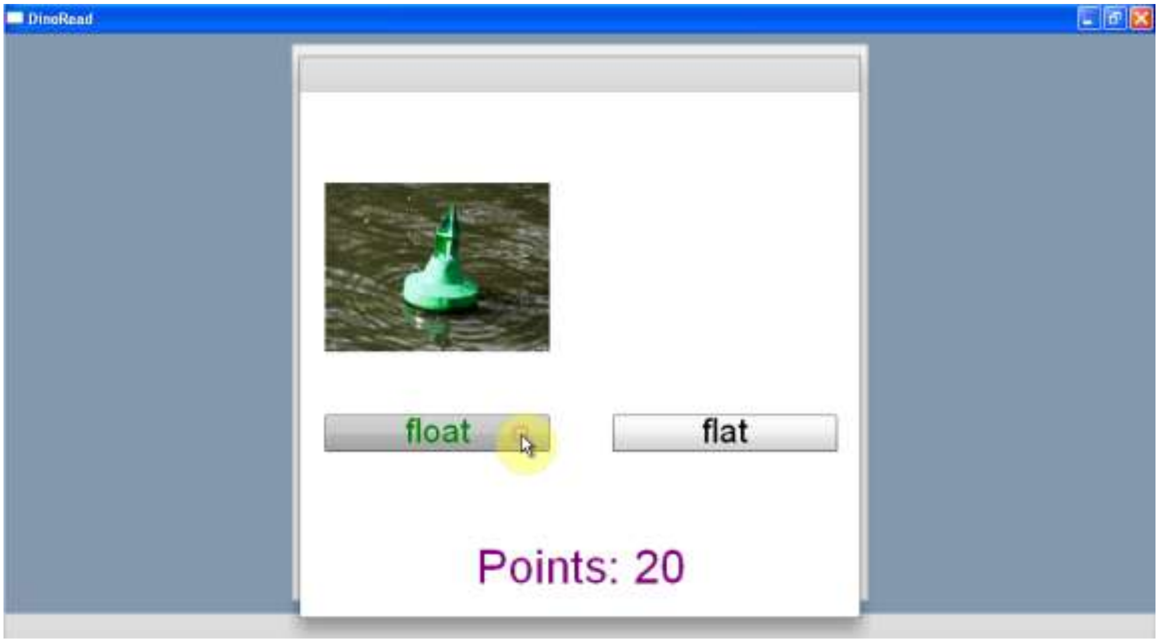



**Figure 4. Format of word identification practice screen.**

Photo by Adamantios (http://commons.wikimedia.org/wiki/File:Green-river-buoy.jpg);

Creative Commons Attribution – Share Alike

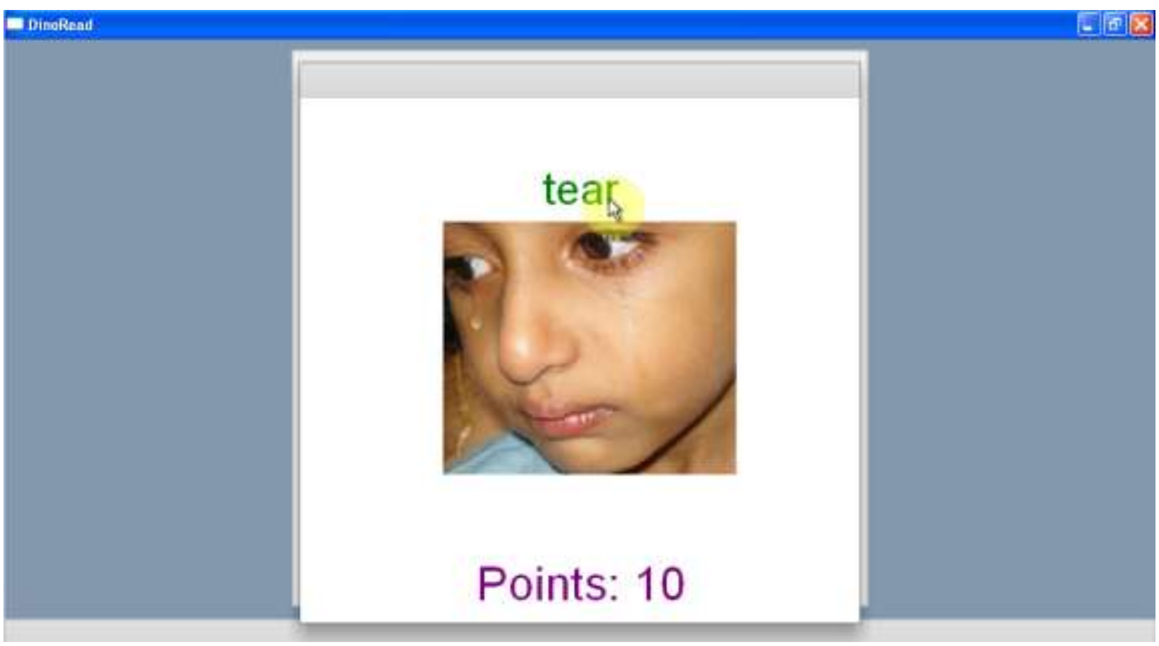



**Figure 5. Word synthesis practice screen.**

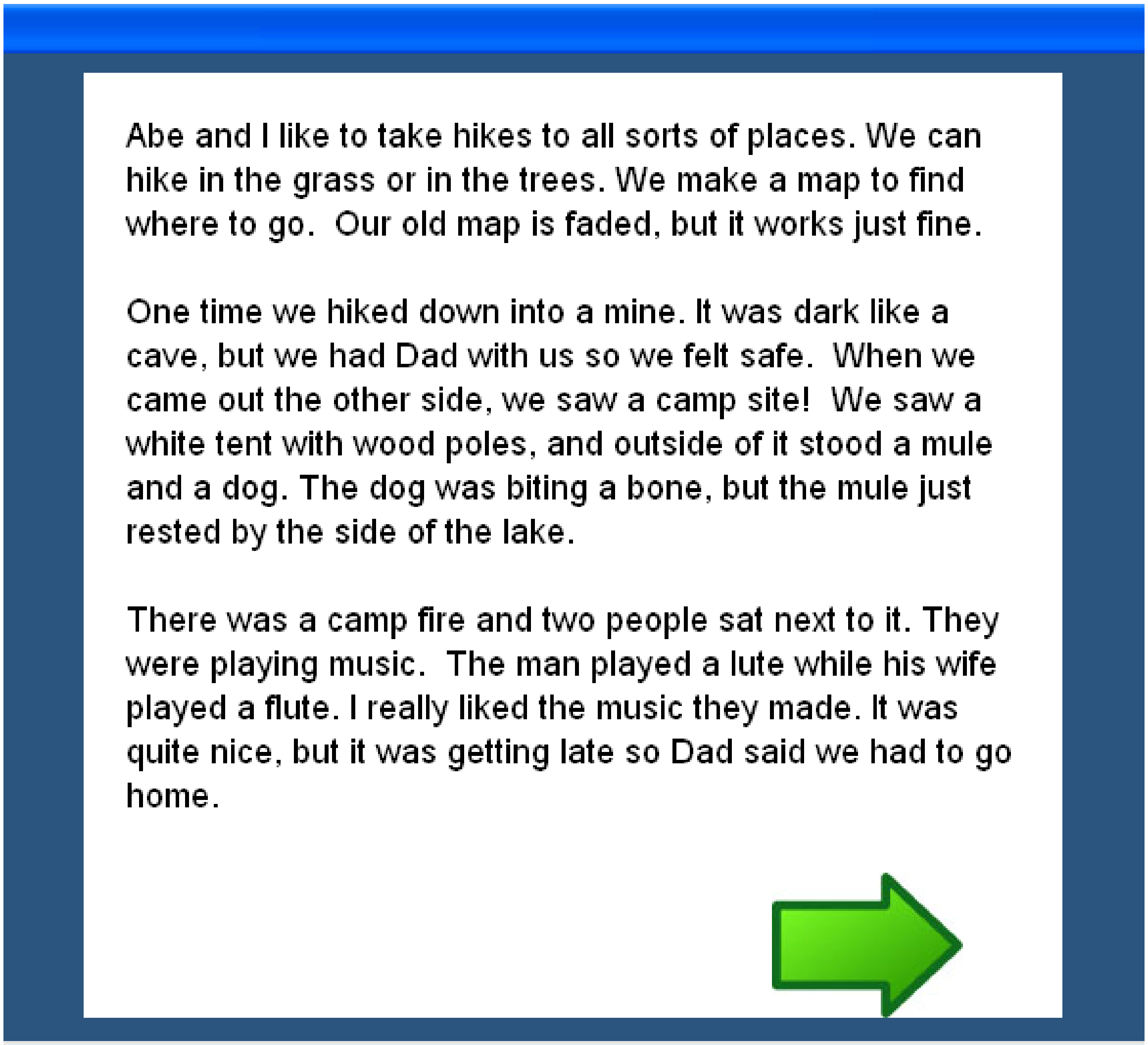
Photo by Reji Jacob (http://commons.wikimedia.org/wiki/File:Tears_.jpg);  Creative Commons Attribution – Share Alike

APPENDIX K

# Running Record Texts

## Running Record Texts

**Text 1: Silent e / long-vowel (cvce)**



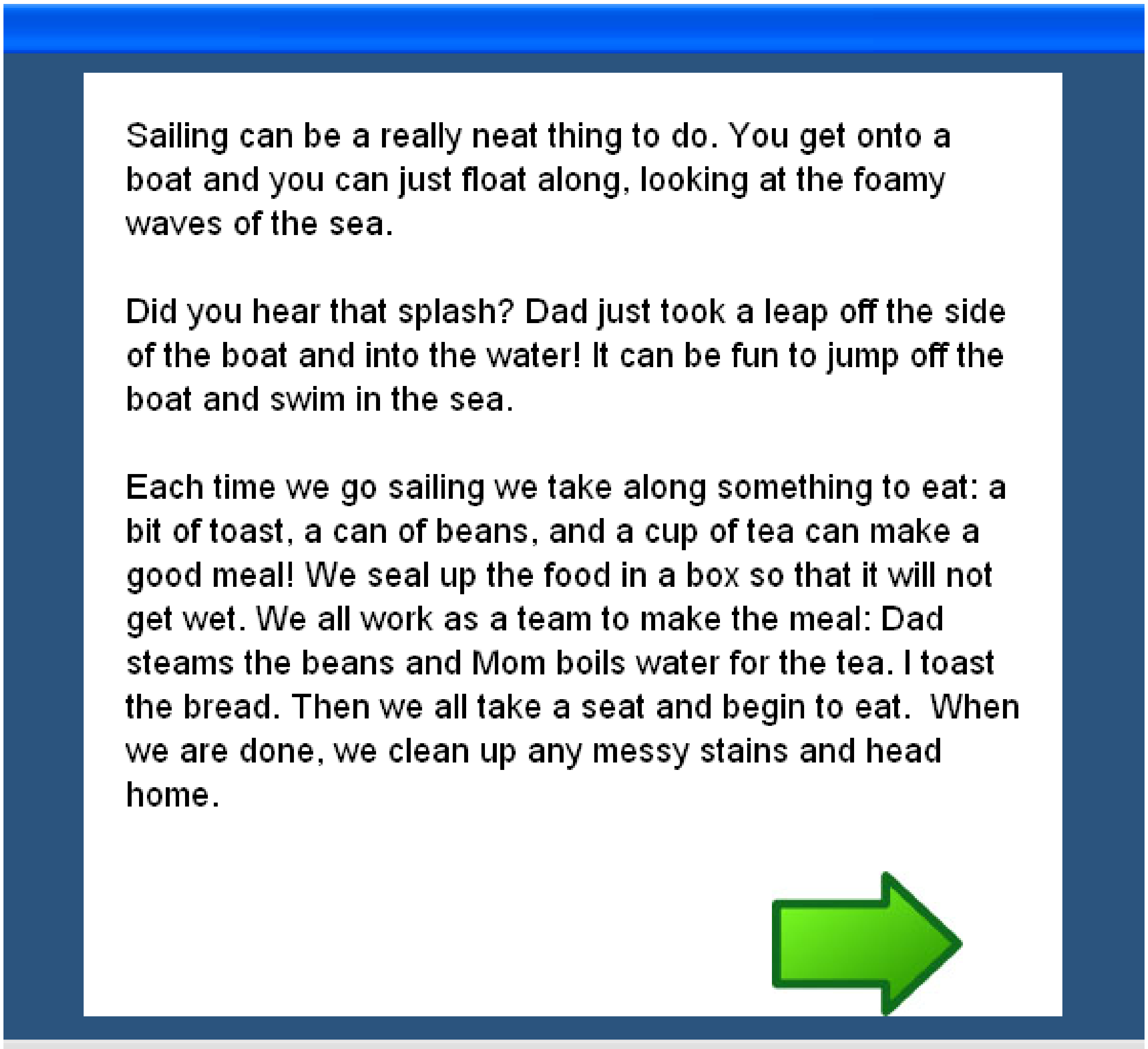
Abe and I like to take hikes to all sorts of places. We can hike in the grass or in the trees. We make a map to find where to go. Our old map is faded, but it works just fine.

One time we hiked down into a mine. It was dark like a cave, but we had Dad with us so we felt safe. When we came out the other side, we saw a camp site! We saw a white tent with wood poles, and outside of it stood a mule and a dog. The dog was biting a bone, but the mule just rested by the side of the lake.

There was a camp fire and two people sat next to it. They were playing music. The man played a lute while his wife played a flute. I really liked the music they made. It was quite nice, but it was getting late so Dad said we had to go home.

**163 total words; 36 cvce words (22%)**

**Text 2: Long vowel digraph (oa, ea, ai)**




**162 total words; 32 vowel-digraph words (19.8%)**

**Text 3: r-controlled (ar, er, ir, or, ur)**

Living on a farm is hard work! You wake up early in the morning to the sounds of birds chirping, and that is when you start your day. Father often works in the fields, tilling and turning the dirt for the plants. Mother sometimes works in the barn, milking the cows or churning butter from their milk.  Other times you can find her on the porch, rocking in her chair as she makes a scarf out of yarn. She just sits and rocks, working on her scarf as she listens to our dog Barney bark as he chases rabbits around the yard.

My job is to help take care of the horses. I bring them food and brush their hair. The large horse is named Marge and she is mostly brown but has some white marks on her head. She just had a a baby we named Rocker. I am extra careful with him to make sure he doesn't get hurt.

**161 total words; 32 r-controlled words (19.9%)**

APPENDIX L

# Development and Testing Timeline

**Development and Testing Timeline**

| | |
|---|---|
| March 24 | Begin development of *DinoRead* software and conditions |
| September 8 | Send out participation invites for study |
| September 8-20 | Develop Phase 1 content (audio recordings) |
| September 20-27 | Phase 1 – Digitized Audio testing |
| Sept. 20 – Oct. 4 | Develop Phase 2 content (text, graphics, animations) |
| October 4-13 | Phase 2 – Graphics Format testing |
| Oct. 13 – Nov. 8 | Develop Phase 3 content (text + animations) |
| November 8-18 | Phase 3 – ASR testing |
| Nov. 8 – Dec. 6 | Develop Phase 4 content (tutorial animations, games) |
| December 6-15 | Phase 4 – Tutorial/Game testing |