Building Adaptation and Error Feedback in an Intelligent Tutoring System

for Reading Comprehension of English Language Learners

by

Audrey Wong

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved December 2016 by the
Graduate Supervisory Committee:

Erin Walker, Chair
Brian Nelson
Arthur Glenberg

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Many English Language Learner (ELL) children struggle with knowledge of vocabulary and syntax. Enhanced Moved by Reading to Accelerate Comprehension in English (EMBRACE) is an interactive storybook application that teaches children to read by moving pictures on the screen to act out the sentences in the text. However, EMBRACE presents the same level of text to all users, and it is limited in its ability to provide error feedback, as it can only determine whether a user action is right or wrong. EMBRACE could help readers learn more effectively if it personalized its instruction with texts that fit their current reading level and feedback that addresses ways to correct their mistakes. Improvements were made to the system by applying design principles of intelligent tutoring systems (ITSs). The new system added features to track the student’s reading comprehension skills, including vocabulary, syntax, and usability, based on various user actions, as well as features to adapt text complexity and provide more specific error feedback using the skills. A pilot study was conducted with 7 non-ELL students to evaluate the functionality and effectiveness of these features. The results revealed both strengths and weaknesses of the ITS. While skill updates appeared most accurate when users made particular kinds of vocabulary and syntax errors, it was not able to correctly identify other kinds of syntax errors or provide feedback when skill values became too high. Additionally, vocabulary error feedback and adapting the complexity of syntax were helpful, but syntax error feedback and adapting the complexity of vocabulary were not as helpful. Overall, children enjoy using EMBRACE, and building an intelligent tutoring system into the application presents a promising approach to make reading a both fun and effective experience.

ACKNOWLEDGMENTS

This research was supported by NSF grant #1324807.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

MOTIVATION

Reading is essential for success in school, work, and society in general. Without the ability to read, educational and economic opportunities are limited, which can decrease one's quality of life (August & Shanahan, 2006). Unfortunately, many English Language Learners (ELLs) struggle with reading. ELLs are children who speak a language other than English and are not proficient in English when they enter school. They tend to perform poorly in reading comprehension compared to their native English-speaking counterparts. This results in lower graduation rates and overall academic achievement--a bleak outlook for the increasing number of ELLs across the country, where 77.2 percent of ELLs speak Spanish at home based on a 2008-2009 report by the US Department of Education (Batalova & McHugh, 2010).

Knowledge of vocabulary and syntax contribute greatly to reading comprehension performance, and ELLs usually lack in these areas (Mokhtari & Niederhauser, 2013; Proctor, Carlo, August, & Snow, 2005). Recent studies have shown that the Moved by Reading (MbR) intervention is effective in improving the reading comprehension skills of monolingual, English-speaking children, and the same effects also apply to ELLs (Walker, Adams, Restrepo, Fialko & Glenberg, in press; Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004; Glenberg & Kaschak, 2002).

The MbR intervention consists of two stages called physical manipulation (PM) and imagine manipulation (IM). In PM, the child reads texts describing events in a particular scenario, such as a story taking place on a farm. When the child encounters a manipulation sentence, he or she must literally manipulate objects (e.g., toys or images) to act out that sentence. For example, if the child reads the sentence “The farmer picks up the hay”, then he or she must move the farmer object to the hay object to simulate the farmer picking up the hay. Understanding of the sentence is demonstrated by the correctness of the manipulation. After practicing with PM, the child begins IM where he or she is asked to imagine manipulating the objects instead (Glenberg, Willford, Gibson, Goldberg, & Zhu, 2011). PM and IM are designed to teach the skill of simulation, and using simulation, the child is forced to comprehend the text. Thus, mastering IM following PM enhances comprehension skills that are then transferred to situations where the child can read successfully without the help of any intervention (Glenberg, 2011).

Enhanced Moved by Reading to Accelerate Comprehension in English (EMBRACE) is an iPad application that focuses on MbR in the context of teaching ELL children, particularly those who come from Spanish-speaking families. PM is implemented by moving images across the screen, while IM is implemented by imagining moving those images. The advantage of using EMBRACE is that reading comprehension can be decomposed into a series of steps (i.e., manipulations) that make up tasks (i.e., sentences). However, these steps must be completed in a specific order, and the system is only concerned with whether the child’s attempt at the current step is right or wrong. If the child does not know why the manipulation was incorrect, then he or she is unlikely to learn from the mistake. The child may have to keep guessing until he or she happens to produce the correct answer. Or the child may become frustrated and give up entirely. On the other hand, if the child completes every step without error, then the text may be too easy, so he or she will get bored while receiving little benefit from the application. A solution to these problems is to build an intelligent tutoring system (ITS) into EMBRACE.

ITSs record various data about the user in the student model, such as skills mastered, learning styles, test scores, tasks completed, amount of time taken, and number of help requests or errors. The student model is used to inform the ITS of which task the student should complete next (VanLehn, 2006). Human tutors vary in their abilities to keep track of this information while adjusting the learning material to fit the student’s needs and diagnosing errors as misconceptions arise. However, ITSs have the capability to accomplish all of these functions simultaneously and in a consistent manner without ever becoming tired. ITS designers also have the freedom to change how the system should behave when teaching different materials to different populations of students (VanLehn, 2011). By keeping track of user knowledge in real time, rather than waiting until the completion of an assessment at the end of the text, an ITS in EMBRACE will be able to provide appropriate error feedback and complexity adaptations as needed while the user progresses through the application.

Overall, the goal of this project is to gain a clearer understanding of the skills and misconceptions of ELL children so that we can personalize their learning experience to help them become better readers. The research seeks to answer the following questions:

- How do different user actions relate to reading comprehension skills?
- How can reading comprehension skills be used to adapt text complexity?
- How can reading comprehension skills be used to provide appropriate error feedback?

The challenge with answering these research questions comes from the fact that reading comprehension requires a large set of skills that have a wide range of applications, and it is not clear how and which skills should be represented in an ITS. It is also difficult to determine the student’s initial reading level and when the system should present easier or harder texts. A final complication involves the type and frequency of error feedback that would best help students correct their mistakes.

I attempted to solve these challenges using the ITS I built for EMBRACE. To simplify the vast amount of possible reading comprehension skills, I chose to focus on two broad categories—vocabulary and syntax—plus an additional usability category to account for any system design issues. I also chose to start all students at a medium level complexity while increasing or decreasing the complexity of the next text to read based on their skills. Finally, I decided that the system should provide immediate error feedback to tell the student whether an action was correct or incorrect, as well as specific error feedback, if necessary, to give the student a hint on how to perform the correct action.

I will cover the background research, implementation details, and successes and failures of these design decisions in later sections. The structure for the remainder of this document is as follows: I will examine related work regarding the design principles of ITSs and a few examples of ITSs for reading comprehension. Then, I will describe EMBRACE in more detail along with how I was able to apply some of the design principles of ITSs. Afterwards, I will discuss how these design principles were implemented in the EMBRACE ITS. Finally, I will provide an analysis and discussion of the results from the study that was conducted to evaluate the system. This will be followed by my conclusion, project limitations, and future work.

CHAPTER 2

RELATED WORK

There is little research in the area of designing ITSs for reading comprehension of ELL children (Duchateau, Kong, Cleuren, Latacz, Roelens, Samir, . . . hamme, H. V., 2009). Most ITSs concentrate on developing skills in math or science rather than reading comprehension. This is because reading is considered an "ill-defined domain." Unlike problems in math or science, which may have only one correct answer, problems in reading may have multiple solutions, and there is a level of subjectivity in assessing comprehension skills (Jacovina & McNamara, 2016). Furthermore, the set of skills required in reading is huge and varied. Not only are there grammatical rules, but there are also exceptions to those rules, plus all of the vocabulary words in the language. Designing an ITS for reading comprehension that accurately and efficiently assesses all of these skills in different contexts poses an enormous challenge (Heilman & Eskenazi, 2006).

Design Principles of ITSs

Anderson, Corbett, Koedinger, and Pelletier (1995) devised eight principles to guide the design of ITSs. The first principle is to "represent student competence as a production set." This principle involves decomposing skills into components which allow the system to more accurately assess what the student knows based on his or her actions. The second principle is to "communicate the goal structure underlying the problem solving." Just as skills are decomposed into components, the process of solving a problem is decomposed into goals and sub-goals, and these goals should be made explicit to the student. The third principle is to "provide instruction in the problem-solving context." Applying this principle means providing instruction before new skills are introduced so that the student can refer back to that instruction when solving a problem. The fourth principle is to "promote an abstract understanding of the problem-solving knowledge." As the authors noted, "students will often develop overly specific knowledge from particular problem-solving examples," so production rules need to be sufficiently general. The fifth principle is to "minimize working memory load." This principle suggests the idea of only teaching a few new skills at a time because presenting too much information simultaneously can interfere with learning. The sixth principle is to "provide immediate feedback on errors." Not only does immediate feedback cut down on time spent in error states but it also makes it easier to interpret the student's actions. The seventh principle is to "adjust the grain size of instruction with learning." Finally, the eighth principle is to "facilitate successive approximations to the target skill." When a student is trying to learn a skill for the first time, the tutor may need to provide a lot of help, but as the student develops the skill, the system should provide less help over time. These principles can be found in three notable examples of ITSs for improving reading comprehension ability: Project LISTEN, REAP, and iSTART.

Project LISTEN

Project LISTEN's Reading Tutor targets oral reading skills by employing speech recognition software that listens to children read aloud. In each session, the child reads one or more stories obtained from a variety of sources selected based on level and genre. He or she may start with an introduction to the activity or end with a review of new vocabulary. As the Reading Tutor listens to the child read each sentence, it will provide help at the student’s request or when it detects situations in which the student may require assistance with a word (Mostow, Nelson-Taylor, & Beck, 2013; Mostow, 2012).

Different user actions, such as speech, delays, and help requests, are mapped to reading comprehension skills (Mostow, 2012; Hagen, Pellom, & Cole, 2007). The tutor assigns credit for each word that the student reads successfully because “failure to read the sentence fluently indicates that the child may not have understood it.” Credit is based on how long the student takes to identify each word and read it aloud (Mostow, 2012; Beck, Chang, Mostow, & Corbett, 2008). This information determines the student’s recommended reading level, which is used to help the tutor select the next best story to read (Poulsen, 2004).

Stories comes from a variety of sources, including children’s newspapers, public-domain materials, and texts written by project members. Each story is manually assigned a level and genre (Mostow, Nelson-Taylor, & Beck, 2013). After completing the initial tutorial, the student alternates with the tutor when selecting the next story to read. This method prevents the student from repeatedly selecting the same easy story. When it is the student’s turn, he or she is free to choose any story, although the tutor will suggest an appropriate reading level (Mostow & Beck, 2007; Poulsen, 2004; Mostow, Aist, Burkhead, Corbett, Cuneo, Eitelman, Huang, Junker, Sklar, & Tobin, 2003). When it is the tutor’s turn, it will choose a previously unread story that matches the student’s recommended reading level (Poulsen, 2004). The tutor will also purposely underestimate the student’s initial reading level based on age. The goal is to avoid frustrating students with stories that are too difficult for them to read (Mostow, Aist, Burkhead, Corbett, Cuneo, Eitelman, Huang, Junker, Sklar, & Tobin, 2003).

If the student is having difficulty with the reading, the Reading Tutor will use timing information from speech and silence to decide when and how to provide feedback. For example, the tutor can detect when the student requests help, hesitates, gets stuck, skips a word, misreads a word, or encounters a word that is likely to be misread. Feedback in these situations may be spoken, graphical, or both. Credited words are displayed in green, while uncredited words are displayed in black. Skipped words are underlined, and the tutor will cough to draw attention to it. The tutor can offer to read words or whole sentences aloud or give a hint on how to do so, such as by temporarily showing a rhyme or related words. It can also supply definitions when the student clicks a word for help (Mostow, Nelson-Taylor, & Beck, 2013; Mostow, 2012).

These methods of feedback are designed to be subtle to make the error less stressful for the student (Reeder, Shapiro, Wakefield, & D'Silva, 2015). If more than one type of feedback is appropriate in the situation, then the tutor will choose one randomly (Poulsen, 2004). If the student does not like the help provided, he or she can request the tutor to choose again (Beck, Chang, Mostow, & Corbett, 2008). Choosing randomly provides variety "for the sake of interest and to generate data on their relative efficacy" (Mostow, Aist, Burkhead, Corbett, Cuneo, Eitelman, Huang, Junker, Sklar, & Tobin, 2003).

Note that the Reading Tutor does not intervene unless the student reaches the end of the sentence, gets stuck, or requests help. One reason is because the tutor does not want to disrupt the flow of the story (Hagen, Pellom, & Cole, 2007). The other reason is that waiting until the end of the sentence prevents the tutor from having to decide on the fly what feedback to give. However, the tutor will still respond to all clicks rather than ignore the student, even if only to explain why it cannot perform the requested action (Mostow, 2012).

The effectiveness of Project LISTEN has been evaluated in multiple studies. In a study with 178 students, grades 1-4, significant fluency and comprehension gains were found in students instructed with the Reading Tutor compared to students instructed with Sustained Silent Reading (SSR). Fluency gains were also found among English as an Additional Language learners in grades 2-4 whose home language was Spanish using the tutor compared to SSR. Additionally, the tutor has been successful among ELLs in Ghana and India (Reeder, Shapiro, Wakefield, & D'Silva, 2015). These studies have found that students tend to read words more accurately immediately after receiving help, which may be attributed to recency effects. However, students request help when they do not know a word, and they are more likely to learn if help is provided (Beck, Chang, Mostow, & Corbett, 2008). Overall, the Reading Tutor appears to be more effective than SSR in teaching both native English speaking children and ELL children. It primarily raises fluency, which has a direct link to comprehension according to many studies. (Poulsen, 2004).

By comparison, an EMBRACE ITS, without speech recognition software, would not approach the task of improving reading comprehension by improving fluency. Nonetheless, an EMBRACE ITS would be similar to Project LISTEN's Reading Tutor in that both systems can provide auditory and visual feedback. For example, if the user taps on a word in a story, then the EMBRACE ITS will play the pronunciation of that word. If the user moves an image incorrectly, then the EMBRACE ITS will play an error noise and snap the image back to its original location. However, a major difference would be the way the two systems choose the type of feedback and when to provide it. The Reading Tutor cannot decide the type of feedback to provide until the student reaches the end of the sentence or the student asks for help, whereas an EMBRACE ITS would be able to make the decision immediately without the student's explicit request based on information from his or her current skill set. Thus, the user can correct errors right after they are made in the EMBRACE ITS. Additionally, both systems can keep track of the student's current reading level, but the EMBRACE ITS would automatically select texts at that level for the student to read in a specific order. Project LISTEN does have the advantage of using timing information to detect when a user is stuck in a sentence. While this information would be useful in an EMBRACE ITS, it is outside the scope of my project.

REAP

REAP's goal is to present real documents from the internet that are useful and appropriate for the student based on constraints, such as target vocabulary, document length, and reading level. Documents are selected automatically and are followed by a set of system-generated exercises related to the target vocabulary to review and practice. Afterwards, the system updates its student model and retrieves the next suitable reading material (Heilman, Zhao, Pino, & Eskenazi, 2008; Heilman, Collins-Thompson, Callan, & Eskenazi, 2006).

The user can constrain search results to contain at least some, but not necessarily all, words from his or her target vocabulary list. These words are highlighted if present in the document, and students can click on any word to receive definitions. Dictionary access is logged, so teachers can see which words students looked up (Heilman &

Eskenazi, 2006; Heilman, Zhao, Pino, & Eskenazi, 2008). The user also takes a pretest to determine which words in the document of the proper reading level have not been previously learned (Heilman & Eskenazi, 2006).

REAP uses statistical language modeling and information retrieval methods to model the student's knowledge and find useful, authentic reading materials for them. Documents are analyzed for syntactic features, readability, length, and occurrence of target vocabulary. REAP selects documents that satisfy a number of constraints using the student model and target reading level, which is sixth to eighth grade. After reading the text, the student completes multiple-choice cloze questions to practice the target words. These exercises provide immediate feedback and are also used to assess student knowledge. The system updates the student model after every reading so that it can choose the next best text to present (Heilman, Zhao, Pino, & Eskenazi, 2008; Heilman, Collins-Thompson, Callan, & Eskenazi, 2006; Heilman & Eskenazi, 2006).

Results from several studies using REAP indicate that students thought the system is easy to use and helped them learn both target and non-target vocabulary. (Heilman, Collins-Thompson, Callan, & Eskenazi, 2006). The system did not select any documents that are too difficult for them to understand. However, students did not find the documents particularly interesting to read (Heilman & Eskenazi, 2006).

Like REAP, the EMBRACE ITS would also include target vocabulary words with definitions to focus on in a text, although these words would be selected by the system, rather than the user, based on his or her skills. The EMBRACE ITS would similarly log help requests so that instructors can determine which vocabulary words users may be most confused about. However, an important distinction between the two systems is that the EMBRACE ITS would allow the user to practice words throughout the reading via manipulation of images that correspond to those words, while REAP must wait until after the user has finished the reading to practice words in the follow-up exercises. This means that students using the EMBRACE ITS can receive error feedback much sooner. Additionally, instead of selecting disjointed texts based on the desired constraints, the EMBRACE ITS would simply adapt the next chapter of the current story by increasing or decreasing the number of target vocabulary words and increasing or decreasing the syntactic complexity. This solves REAP's issue with finding documents that are not "static, difficult to produce, and very limited in quantity" as the EMBRACE ITS can automatically present different versions of the same texts (Heilman, Collins-Thompson, Callan, & Eskenazi, 2006). Thus, the system would not disrupt the flow of the text while maintaining the reader's interest.

iSTART

iSTART trains students to use different reading strategies. It is modeled after SERT, or self-explanation reading training, which improves reading comprehension by explaining texts to oneself (McNamara, Levinstein, & Boonthum, 2004). iSTART uses virtual agents to tutor students about self-explanation through lessons with quizzes, analysis of an agent's self-explanations with coaching, and the student's own self-explanations under an agent's guidance. iSTART has three modules: Introduction, Demonstration, and Practice. In the Introduction module, students watch the teacher-agent explain reading strategies to two student-agents. In the Demonstration module, students complete quizzes on various aspects of the strategies. And in the Practice module, students practice generating typed self-explanations while the program provides feedback on performance (McNamara & the CSEP Lab, 2006; Levinstein, Boonthum, Pillarisetti, Bell, & McNamara, 2007).

The quality of self-explanations is evaluated based on features such as length and number and type of content words. Then the system automatically assigns a score. Jacovina and McNamara explain that "higher scores are assigned to self-explanations that incorporate information from throughout the text and prior knowledge from outside the text, whereas lower scores are assigned to self-explanations that are short, irrelevant, or too similar to the target sentence." However, the system does not score on the accuracy of the content in the self-explanations (McNamara & the CSEP Lab, 2006).

The system provides adaptive feedback in the practice section based on student performance. Students answer multiple-choice questions after each reading strategy and receive immediate feedback. For example, the system may ask a student to add more information for "impoverished" self-explanations (McNamara & the CSEP Lab, 2006). Feedback includes both the score and suggestions on how to improve the student's answer (Jacovina & McNamara).

iSTART has been studied with over 1,000 middle school, high school, and college students. Users performed significantly better than students in the control group in terms of producing better quality self-explanations and higher comprehension scores. Studies have suggested that iSTART is effective in helping students use reading strategies to learn from texts and enhances comprehension, particularly among low-knowledge readers. However, it was noted that students found the repetitive activities boring (Levinstein, Boonthum, Pillarisetti, Bell, & McNamara, 2007; McNamara & the CSEP Lab, 2006).

Both iSTART and the EMBRACE ITS focus on teaching reading strategies. In the EMBRACE ITS, these reading strategies are physical manipulation and imagine manipulation. The presentation of these reading strategies is different, as the EMBRACE ITS does not use any virtual agents to tutor students. Instead, students complete an initial tutorial on physical manipulation, and then they must complete the remaining activities on their own, although brief instructions are provided at the beginning of each chapter. Performance on physical manipulation is largely based on the steps that the user makes and the number of errors per step. Thus, accuracy is important. The EMBRACE ITS can provide more immediate feedback tailored to fixing the user's specific error.

CHAPTER 3

EMBRACE

EMBRACE is an interactive storybook application on the iPad. It is written primarily in Objective-C and uses .epub files for each story, which contain .xhtml, .xml, and various audio and image files. The application has a library of both narrative and expository texts authored by project members (see Figure 1). Most of these texts are comprised of five to seven chapters (Walker, Adams, Restrepo, Fialko & Glenberg, in press). New users complete a short story with a single chapter to introduce the user to the application, providing instructions on how to request vocabulary help and how to move images for physical manipulation sentences.

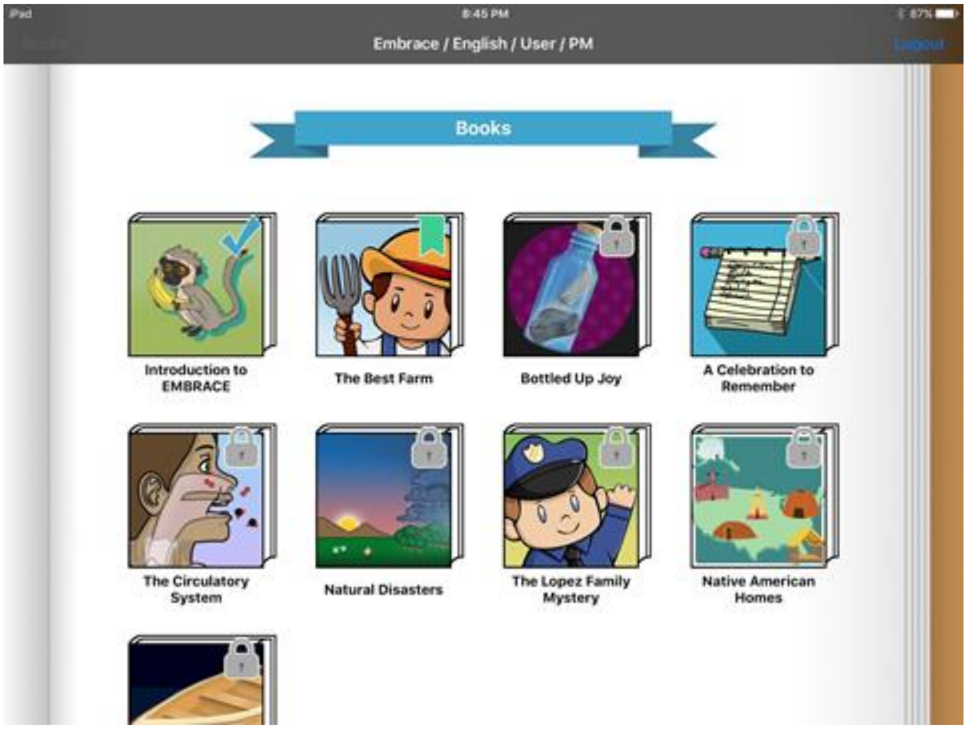


Figure 1. EMBRACE library of books.

After completing the tutorial, students usually start reading the stories in a predefined order as indicated by the system using special icons. These icons display the status of each story or chapter. A green bookmark icon marks which story or chapter to read next. A green checkmark icon marks that the story or chapter has already been read. And a gray lock icon marks that the story or chapter is not available for reading yet. A student will read all chapters of a story before being allowed to move onto the next story.

Each chapter begins with a list of target vocabulary words that are introduced in the text (see Figure 2). The user taps on each word to hear audio for the pronunciation and definition and to see the corresponding image, if available, highlighted on the screen. The user presses a “Next” button to advance to the next word.

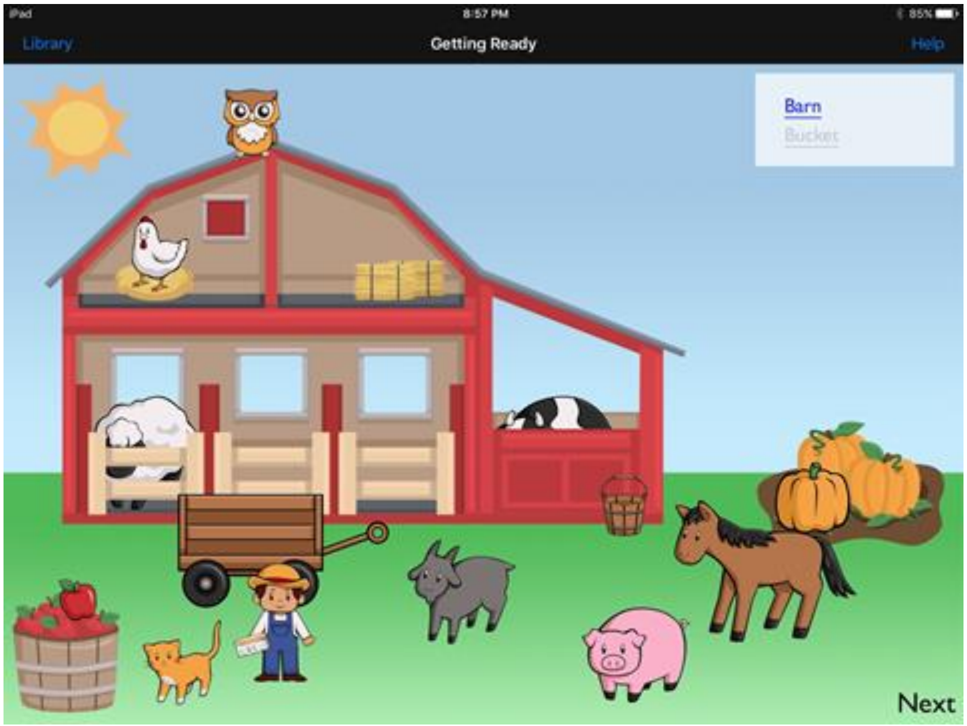
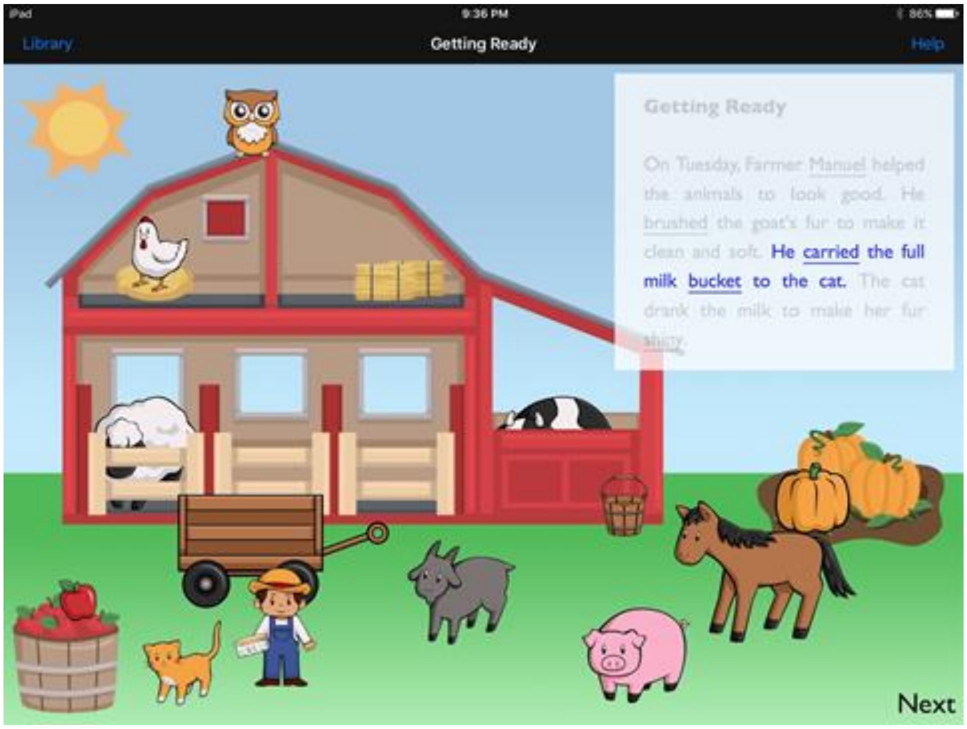


Figure 2. EMBRACE vocabulary list.

Following the vocabulary list is the chapter text (see Figure 3). Chapters consist of multiple pages like a regular storybook in which most of the page is filled with images depicting a scene in the chapter, and the sentences are displayed in a textbox in the upper left or upper right corner of the screen. The current sentence to read is displayed in either blue (manipulation sentence) or black (non-manipulation sentence). Previous and next sentences are dimmed out. Additionally, sentences contain underlined words that the user can tap to hear the pronunciation and see the image highlighted if available. The user advances the current sentence by completing all manipulations in the correct order for the sentence, if manipulation is required, and then pressing the "Next" button. If not all manipulations have been completed, the system will play an error noise (Walker, Adams, Restrepo, Fialko & Glenberg, in press).



Getting Ready

On Tuesday, Farmer Manuel helped the animals to look good. He brushed the goat's fur to make it clean and soft. He carried the full milk bucket to the cat. The cat drank the milk to make her fur shiny.

Figure 3. EMBRACE manipulation activity.

Physical manipulation sentences can contain one or more steps. Most manipulations involve moving one object to another object or location. For example, Figure 3 highlights the current sentence "He carried the full milk bucket to the cat." This sentence requires two steps: (1) move the farmer object to the bucket object, and (2) move the farmer and bucket objects to the cat object. In some cases, a menu with images depicting different possible interactions is shown to disambiguate the manipulation. Moving the farmer object to the cat object can result in two possible interactions: (1) the farmer picks up the cat, or (2) the cat stands on the farmer's head. The user must select the correct option for the current step. He or she cannot dismiss the menu until they choose the correct option. These menus can also appear in imagine manipulation activities without the user moving any images. In these situations, the user must choose the option that best represents what they imagined happening in the sentence.

When the user makes an error, an error noise is played and moved objects are reset to their original positions (Walker, Adams, Restrepo, Fialko & Glenberg, in press). Making five errors on the same step will induce the system to display a message that lets the user know the system will show him or her how to complete the current step. This is because students become too frustrated if they are not given the answer after a certain number of tries (Lepper & Chabay, 1985). When the user presses "OK" to dismiss this message, he or she will watch as the system temporarily highlights the correct object to move, animates moving this object to the other correct object or location, and temporarily highlights this object again.

At the end of each chapter is a series of multiple-choice reading comprehension assessment questions (Walker, Adams, Restrepo, Fialko & Glenberg, in press). The user can press buttons next to the question and each answer option to hear the text read aloud. Selecting a correct answer option highlights it in blue, while selecting an incorrect answer option grays it out. The user cannot press the "Next" button to move on until they choose the correct option.

The EMBRACE system logs all user and computer actions. User actions include help requests and the objects that were moved. Computer actions include which audio was played. Log files are generated in .xml format. This data can be analyzed to determine what features of the application are effective or ineffective.

To assess the effectiveness of EMBRACE, a study was conducted with 93 Latino Dual Language Learner students assigned to read both a narrative and expository text in one of four conditions: English or Spanish support and simulation or no simulation. The study showed benefits of simulation and additional benefits of simulation with Spanish support for some children. The study suggests that simulation improves reading comprehension on narrative texts, but for more challenging expository texts, simulation does not help poor decoders as much as it helps good decoders. Another finding was that simulation with native language support improves performance for good decoders or students with strong English language skills. Results indicated that poor decoders tended to request more help than good decoders, and help requests increased as reading comprehension scores decreased. However, simulation and Spanish support appeared to provide enough assistance that students did not need to request as much help from the application. This makes EMBRACE a promising approach to teaching reading comprehension (Walker, Adams, Restrepo, Fialko & Glenberg, in press).

CHAPTER 4

APPLYING DESIGN PRINCIPLES OF ITSS

This section will explain how I applied the design principles of ITSs in EMBRACE to tackle the challenges of teaching reading comprehension. I used these principles to guide my design and implementation of various system features, including the classification of student skills into vocabulary, syntax, or usability types and a mapping of different user actions to these skills to facilitate complexity adaptations and error feedback.

One of the earliest design principles that I applied was related to my implementation of immediate error feedback. EMBRACE can already be considered a "limited intelligent tutoring system, in that feedback is provided to children on each step" (Walker, Adams, Restrepo, Fialko & Glenberg, in press). For example, if the current step requires the child to move the image of the farmer to the image of the hay, then the system is able to determine whether the child moves these images correctly or incorrectly. If the movement is correct, then the system will update the images by grouping them together to represent the action in the step. Otherwise, the system will play an error noise and snap the images back to their original positions. The child is unable to move onto the next step until the current step is completed correctly.

Although the feedback described above is minimal for an ITS, the solution checking model plays a vital role in the EMBRACE ITS as a whole. It allows EMBRACE to apply the first principle of designing ITSs by decomposing reading comprehension into a series of steps to complete. This solution checking model was built in collaboration with project members. Solution steps are encoded for each chapter of a story in .xml files. Each sentence in the chapter is assigned an idea number, where an idea is a concept or meaning conveyed by that sentence. Each idea is then mapped to zero or more steps. (Non-manipulation sentences have zero steps.) The most common step types are “group” and “check.” Group involves moving two objects together to group them. This step type is encoded by specifying the subject (object to move), the object (object to move to), and the action (an ID describing the interaction between the two objects that corresponds to the x-y location on each image where the two objects should group together). The sentence “The farmer picked up the hay” would be encoded using a single group step that specifies the farmer as the subject, hay as the object, and “pick up” as the action. On the other hand, check step types involve moving an object to a location or area on the screen. Encoding this step requires specifying the object to move, the location, and the action. The sentence “The farmer climbed up to the hayloft” might be encoded using a single check step that specifies the farmer as the object, hayloft as the location, and “climb” as the action. Using the solution checking model, the system verifies whether the user’s step matches the correct step after every attempt.

This immediate error feedback also allows EMBRACE to apply the sixth principle of designing ITSs. However, the solution checking model by itself simply tells the user whether his or her attempt was correct or incorrect. It does not explain why, which is helpful for understanding how to fix an error (Anderson, Corbett, Koedinger & Pelletier, 1995). Fortunately, by decomposing each sentence into multiple steps, the system is able to determine how well the student knows each skill at every step.

Since research indicates that effective reading comprehension requires knowledge of both vocabulary and syntax, I have designed the EMBRACE ITS to categorize skills into the following types: vocabulary, syntax, and usability (Mostow, Gates, Ellison, & Goutam, 2015; Mokhtari & Niederhauser, 2013; Wessels, 2013). Usability is not necessarily part of the literature on ITSs for reading comprehension, but it was added as a skill category primarily to catch errors made by students who have initial trouble learning to use the system or students who have essentially made the correct movement for a step but were not as precise as the system expected.

Different user actions within the application may cause skill values to either increase or decrease. Using the affected skills, other project members and I have made improvements to the immediate error feedback provided by the EMBRACE ITS. These improvements are concerned with detecting the most probable type of error that the user made and tailoring the feedback to help fix that type of error. Thus, feedback specifically addresses vocabulary, syntax, or usability errors. Many students may be reluctant to ask for help, or they may not even realize they need help, so I have also built the EMBRACE ITS to decide when it is appropriate to provide feedback to the user (Mostow, Aist, Burkhead, Corbett, Cuneo, Eitelman, Huang, Junker, Sklar, & Tobin, 2003; Lepper & Chabay, 1985). As a result, this allows EMBRACE to apply the eighth principle of designing ITSs, facilitating successive approximations to the target skill. More feedback may be provided near the beginning of a story when skill values are low, while less feedback may be provided near the end of a story when skill values are presumably higher. This means a student will receive less help from the EMBRACE ITS as he or she demonstrates improvement to his or her reading comprehension ability.

Finally, EMBRACE can apply the fifth design principle of ITSs, minimizing working memory load, by ensuring that students have mastered particular skills before teaching new ones. I have designed the EMBRACE ITS to adapt the complexity of its vocabulary and syntax based on the current skill values in the student model. Students learn in different ways at different rates, so it does not make sense to require all users to read at the same level throughout every point in the texts (Lepper & Chabay, 1985). Thus, the EMBRACE ITS may add extra words to the chapter vocabulary lists if the user needs more practice and later remove those words once the user has mastered them. Drawing focus to these words is important, as the student may not learn them otherwise (Heilman & Eskenazi, 2006). Additionally, project members and I have worked on adding functionality to the EMBRACE ITS to dynamically switch chapters to use simpler or more complex sentences because reading comprehension is facilitated when the student reads text at the appropriate level (Brown & Eskenazi, 2004). As the student reads through a story, he or she will ideally master reading simple sentences before moving onto medium sentences and finally complex sentences.

Various characteristics of the EMBRACE system align with the principles of designing an ITS. The next section will describe in more detail how each of the features of the EMBRACE ITS are implemented.

CHAPTER 5

SYSTEM DESCRIPTION

Overview

The EMBRACE ITS implements Bayesian knowledge tracing techniques to keep track of and update the user's vocabulary, syntax, and usability skills in the student model. It also includes features for adapting the complexity of chapter vocabulary and syntax, as well as features for providing error feedback for each of the three skill types. While I designed the behavior of these features, note that since EMBRACE is a large project funded by the National Science Foundation and has many members, several aspects of these features were implemented in collaboration with others. These collaborations are specified for each feature where appropriate.

Figure 4 shows an overview of how the system works. In this system, the ITS makes decisions not only after every step in a sentence but also after every chapter in a story. All skills are set to their initial values at the beginning of a story. No adaptations are made to vocabulary or syntax for the first chapter, so the text is completely comprised of medium complexity sentences. Every time the user correctly completes a step in a manipulation sentence, the EMBRACE ITS increases all three skill types. If the user makes an incorrect step, then the system analyzes the error to find the related skills to increase or decrease. Then, from the related skills, it determines the most probable error to give immediate feedback if necessary. After the user completes a chapter, the ITS determines which low skill words to add to the vocabulary list of the next chapter along with the level to set for the syntactic complexity. This process repeats throughout the rest of the story.

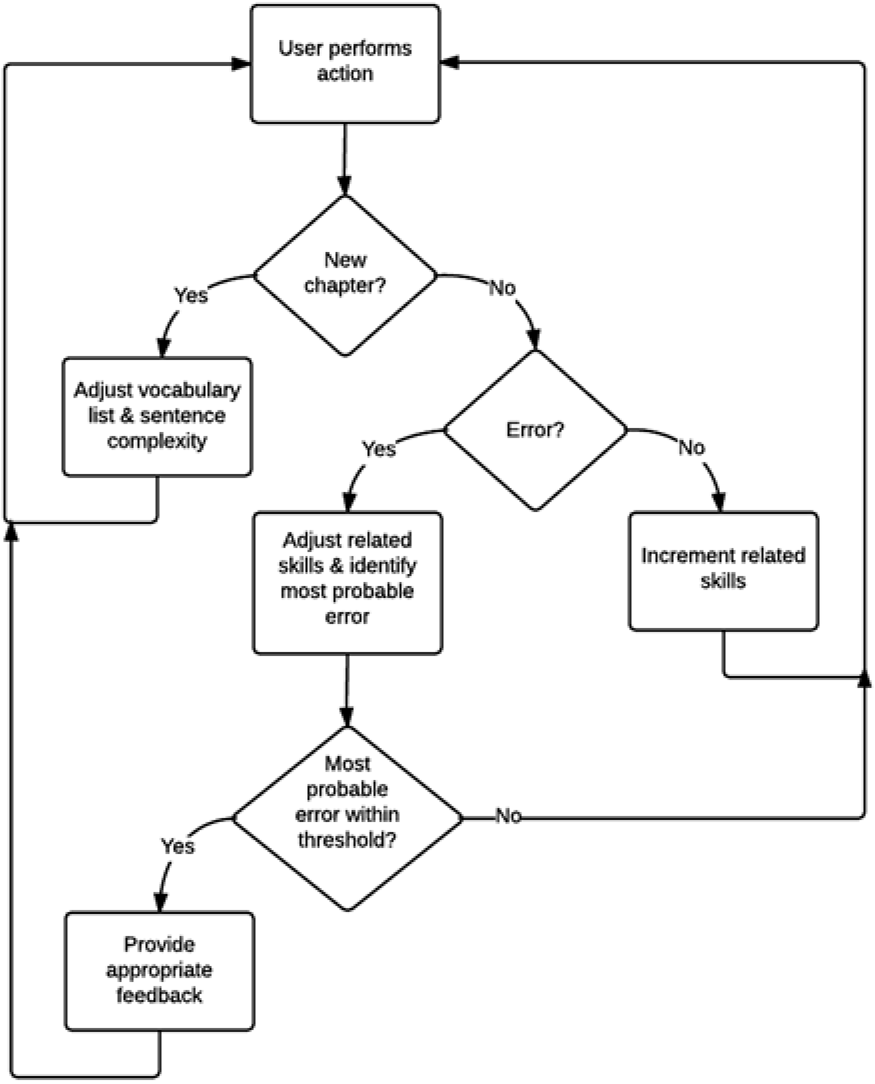

User performs action
New chapter?
Yes
No
Adjust vocabulary list & sentence complexity
Error?
Yes
No
Adjust related skills & identify most probable error
Increment related skills
Most probable error within threshold?
Yes
No
Provide appropriate feedback

Figure 4. Overview of the EMBRACE ITS.

Skill Types

Skills are categorized into three types: vocabulary, syntax, and usability. These skills represent various aspects of the student's knowledge of reading comprehension and using the application. Vocabulary skills are divided into specific words (e.g., farmer, hay, etc.), while syntax skills are divided into simple, medium, and complex to reflect the three possible levels of text in which the student may read. The usability skill assesses how well the student is able to use the system.

Each skill is assigned a value from 0 to 0.99, which represents the probability that the student knows the skill. All skills start at a default value of 0.15 to assume that the student may begin with at least some knowledge of the skill already. The only exception is the simple syntax skill, which starts at 0.99 since chapter complexity begins at medium syntax.

Vocabulary and syntax are the two primary skills because knowledge of these two areas is important to reading comprehension. Usability was added in cases where the student's attempt may be close to the solution but due to possible system design issues, the attempt was marked as incorrect. Sometimes students, particularly those who are new to using the application, may not move objects as precisely as system developers had intended. For example, if images are small, and the correct object to move is next to an incorrect object, it may be more reasonable to assume that the student moved the incorrect object due to an accident rather than a lack of knowledge. Usability errors alert system developers of confusing areas to fix in the application. However, as EMBRACE has been tested with over a hundred students already, these issues should be small in number.

Adjusting Related Skills

Different user actions map to each skill depending on whether the action was correct or incorrect. The EMBRACE ITS currently only considers actions that involve the user tapping on a word in either the vocabulary list or within a chapter and the user moving objects for physical manipulation sentences. It does not process other actions such as tapping on a menu option (in physical manipulation and imagine manipulation), pressing the “Next” button before all steps have been completed, or selecting an assessment answer.

When the user taps on a word in the vocabulary list, this action is considered correct for the purposes of the ITS. Thus, the user’s vocabulary skill corresponding to that word is increased. There are a few reasons behind this approach. One reason is that students are required to tap on these words, so they will likely learn the word via direct instruction after receiving the pronunciation, definition, and highlighted image. The other reason is that not all words have corresponding images (e.g., abstract words like “contest”) so the user would not have any opportunity to manipulate images, or increase their vocabulary skills, for those words. Including a word in the vocabulary list is a limited way for the student to practice that word. By this reasoning, the EMBRACE ITS increases the vocabulary skill of words tapped in the introduction list.

On the other hand, the EMBRACE ITS treats taps on underlined words in the chapter text as incorrect actions. Users generally do not tap on underlined words unless they do not know the words, so it seems reasonable for the system to decrease vocabulary skills in this case.

User attempts at physical manipulation steps can affect all skill types. Figure 5 shows an overview of how skills are updated, and a continuation of the process is shown in Figure 6. Suppose the student performs a manipulation for the sentence “The farmer picked up the hay.” If the action was correct (i.e., the student correctly moves the image of the farmer to the image of the hay), then the vocabulary skills for words corresponding to all the objects that were correctly moved (i.e., farmer) and the vocabulary skill for the word corresponding to the correct destination (i.e., hay) are increased. (Note that a destination describes either another object or a location or area.) Additionally, the syntax skill corresponding to the complexity of the current sentence (e.g., medium) and the usability skill are also increased.

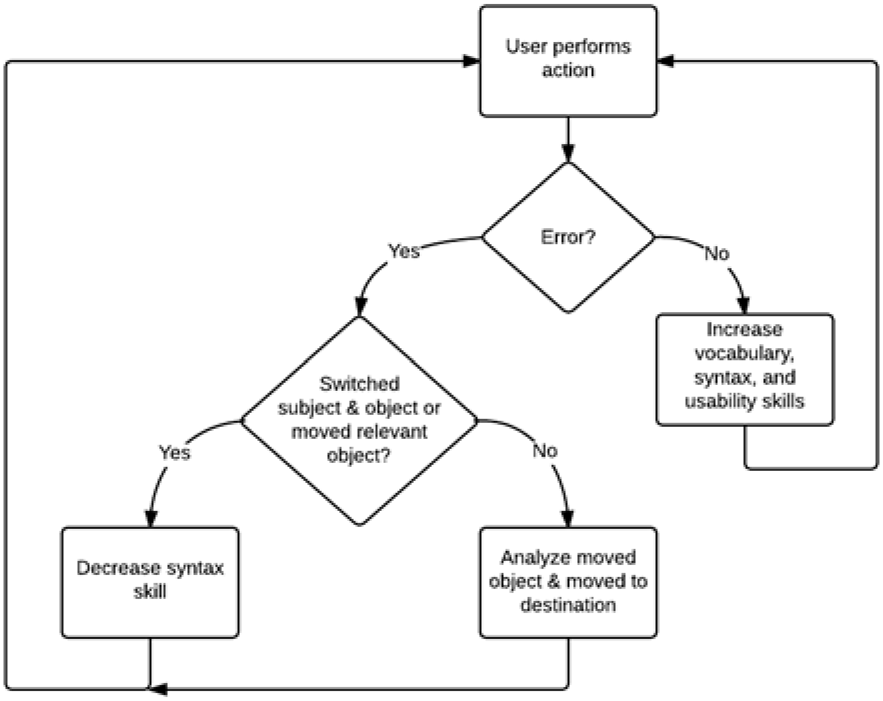

User performs action
Error?
Yes
No
Increase vocabulary, syntax, and usability skills
Switched subject & object or moved relevant object?
Yes
No
Decrease syntax skill
Analyze moved object & moved to destination

Figure 5. Overview of adjusting related skills. Continuation of the process is shown in Figure 6.

On the other hand, if the action was incorrect, then the system analyzes the moved objects and destinations of the moved objects while comparing them to the solution. Depending on how and which items were involved, the system determines the relevant skills to update.

First, the system checks for syntax-only errors. These errors indicate that only the syntax skill should be updated and nothing else. If the user mixed up the order of the subject and object in the step (i.e., the student moved the image of the hay to the image of the farmer instead), then the system will decrease the syntax skill. If the user action only involved items relevant to any step in the current sentence (i.e., the student only moved the farmer or hay but no other objects), then the system will also decrease the syntax skill. This second condition covers cases where the student may incorrectly skip ahead of the current step in a multi-step sentence.

If the system did not detect any syntax-only errors, then it will proceed to check for all three error types--syntax, vocabulary, and usability. It considers skills related to the moved objects followed by skills related to the destinations of the moved objects.

Figure 6 shows an overview of the following process: If the user moved the wrong subject (e.g., the student moved the image of the cow instead of the image of the farmer), then the system will check if the distance between the wrong subject (i.e., cow) and the correct subject (i.e., farmer) is above a certain threshold (90 pixels in the EMBRACE ITS). If the distance exceeds the threshold, then the system will make a decision on whether to update vocabulary or syntax skills. It will decrease the syntax skill if the vocabulary skill for the correct subject is above 0.9. A high vocabulary skill value indicates that the student likely knows the vocabulary word corresponding to the correct subject, so he or she is unlikely to have made a vocabulary error. A low vocabulary skill value indicates the opposite, so the system will decrease the vocabulary skills for the correct subject, and each object in the group containing the wrong subject. Suppose the student had moved a group of objects containing the cow and chicken instead of the farmer by himself--then the EMBRACE ITS would decrease the vocabulary skills for farmer, cow, and chicken. On the other hand, if the distance is below the threshold, then the user probably picked up the wrong subject by mistake because it is so close to the correct subject. Thus, the system will simply decrease the usability skill. One last case to consider is if the user moved the correct subject (although he or she moved the correct subject to the incorrect destination). The system would increase the vocabulary skill for the correct subject.

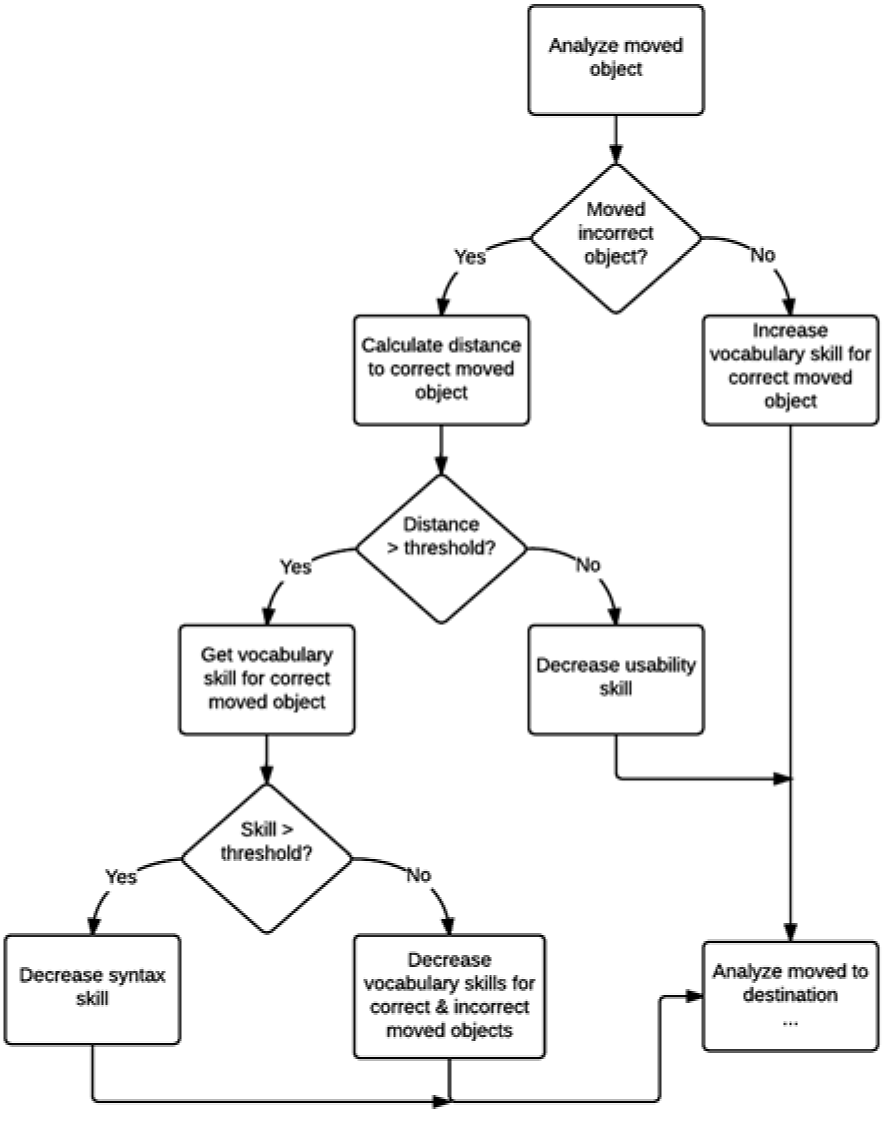

Analyze moved object
Moved incorrect object?
Yes
No
Calculate distance to correct moved object
Increase vocabulary skill for correct moved object
Distance > threshold?
Yes
No
Get vocabulary skill for correct moved object
Decrease usability skill
Skill > threshold?
Yes
No
Decrease syntax skill
Decrease vocabulary skills for correct & incorrect moved objects
Analyze moved to destination
...

Figure 6. Overview of analyzing moved objects and destinations in an error.

The EMBRACE ITS follows a similar process when analyzing the destinations of the moved objects. If the user moved to the wrong destination (e.g., the student moved the image of the farmer to the image of the pumpkin instead of the image of the hay), then the system will check if the distance between the wrong destination (i.e., pumpkin) and the correct destination (i.e., hay) is above the threshold. If this is true, then the system will decrease the syntax skill if the vocabulary skill for the correct destination is above 0.9. Otherwise, it will decrease the vocabulary skill for both the correct destination and the wrong destination. If the distance is below the threshold, then the system will simply decrease the usability skill.

Skills are increased or decreased using Bayesian knowledge tracing, which was implemented in collaboration with project team members. I created the mapping between different user actions and each skill type; other members helped me increase or decrease the skill values based on the action, and I later refined the methods. The Bayesian knowledge tracing algorithm calculates the probability that the student knows a particular skill according to his or her previous actions. These calculations involve parameters called guess, slip, and transition. Guess is the probability that the student can correctly apply a skill without actually knowing it; slip is the probability that the student can incorrectly apply a skill despite actually knowing it; and transition is the probability that the student’s knowledge of a skill transitions from a not-known state to a known state after applying the skill (Baker, Corbett, & Aleven, 2008). As described earlier, the EMBRACE ITS uses the student’s vocabulary skills from previous actions to determine whether the current error is related to vocabulary or syntax. It also keeps track of whether the student has previously attempted the step or not. Table 1 shows the equations used by the EMBRACE ITS to calculate new skill values, while Table 2 lists the separate guess, slip, and transition values for each skill type along with a single dampening value to lessen the amount of change in skill value. If the student made a correct action, meaning the value for a skill should increase, then the system will calculate the new skill value using the equation for the evaluated skill value for correct actions. Otherwise, it will use the equation for the evaluated skill value for incorrect actions.

Value	**Equation**
No slip	$\frac{1 - slip}{dampen}$
No guess	$\frac{1 - guess}{dampen}$
Evaluated skill value (correct action)	$\frac{previous\ skill\ value * no\ slip}{previous\ skill\ value * no\ slip + (1 - previous\ skill\ value) * guess}$
Evaluated skill value (incorrect action)	$\frac{previous\ skill\ value * slip}{previous\ skill\ value * slip + (1 - previous\ skill\ value) * no\ guess}$
New skill value	$evaluated\ skill\ value + (1 - evaluated\ skill\ value) * transition$

Table 1. EMBRACE ITS knowledge tracing equations.

	Vocabulary	**Syntax**	**Usability**
Guess	0.40	0.50	0.40
Slip	0.20	0.30	0.50
Transition	0.10	0.05	0.01
Dampen	10	10	10

Table 2. Guess, slip, transition, and dampen values for each skill type.

I obtained the values in Table 2 after experimenting with various numbers and seeing whether the change in skill values appeared reasonable. The guess and slip values for syntax skills are higher than the guess and slip values for vocabulary skills because user actions do not map to syntax skills as clearly as they map to vocabulary skills. Thus, when students made correct or incorrect actions, I wanted to avoid attributing too much credit to knowledge of syntax. This is also the reason why the transition value for syntax skills is lower than the transition value for vocabulary skills. On another note, I made the slip value for usability skills relatively high because I wanted to give students more leniency when making usability errors. Additionally, dampening was added for skill updates if the user made an error following his or her first attempt on the same step. This prevents the user's skill value from dropping too drastically for repeated errors.

Using these skill values, the EMBRACE ITS can adapt the complexity of vocabulary and syntax and provide error feedback for all three types of skills.

Adapting Complexity of Vocabulary

Vocabulary is adapted by adjusting the list of words that appear in the beginning of each chapter (see Figure 7). At most, eight words can appear in this list. (More than eight words may overwhelm or frustrate the user.) The list always starts with the new words and definitions that are introduced in the chapter. After that, the EMBRACE ITS sorts the remaining possible words by their vocabulary skill value in ascending order. Words with the lowest vocabulary skills are added to the list if they (a) appeared in a previous chapter, (b) appear in the following chapter, and (c) have a skill value below a threshold of 0.80. Note that the threshold was set to 0.80, which is a bit lower than the 0.90 used to determine high skill vocabulary. This reduces the number of potential extra words that the user may already know. These words can include words from previous vocabulary lists, words corresponding to objects involved in previous manipulations, and words that the user requested by tapping on them. When the user taps on these words, some may play audio for both pronunciation and definition, while others may only play audio for pronunciation twice. In either case, the corresponding image, if available, is highlighted. All words that appear in this list are also underlined in the chapter, so that the user can tap on them again to hear the pronunciation and see an image highlighted if present.

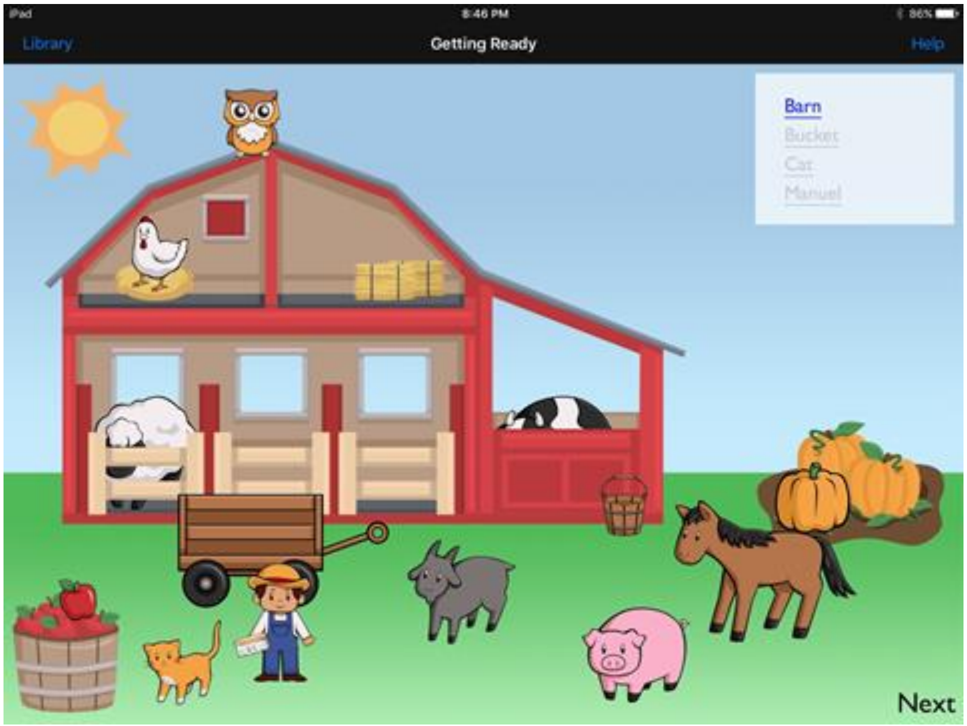
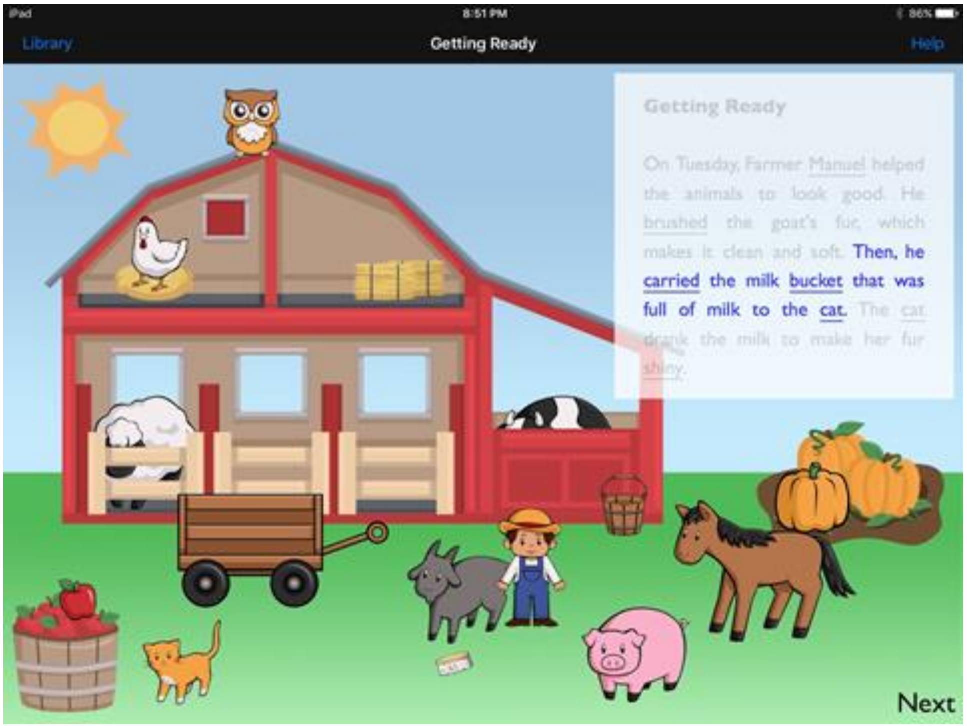


Figure 7. Vocabulary complexity adaptation.

I completed most of the work on this feature on my own, although I received some help generating audio for words that did not already have pronunciations in the system.

Adapting Complexity of Syntax

Syntax is adapted by adjusting the complexity of sentences at the beginning of a chapter. This feature was also implemented in collaboration with project team members who updated the function I had initially created to adapt syntax by page rather than by chapter. In the previous EMBRACE application without the ITS, all text was considered medium level. However, in the EMBRACE ITS, simple and complex versions were written for the texts. Figure 8 shows a complex version of a chapter. The medium version of the highlighted sentence was “He carried the full milk bucket to the cat.” After adapting the syntax, the complex version is shown: “Then, he carried the milk bucket, that was full of milk to the cat.”



Getting Ready

On Tuesday, Farmer Manuel helped the animals to look good. He brushed the goat's fur, which makes it clean and soft. Then, he carried the milk bucket that was full of milk to the cat. The cat drank the milk to make her fur shiny.

Figure 8. Syntax complexity adaptation.

By default, the user will start at medium complexity for the first chapter of a story. Afterwards, if his or her simple syntax skill is below 0.9 or his or her medium syntax skill is below 0.4, then the chapter will switch to simple sentences. If his or her medium syntax skill is below 0.9 or his or her complex syntax skill is below 0.4, then the chapter will switch to medium sentences. Otherwise, the chapter will switch to complex sentences. This decision-flow is illustrated in Figure 9.

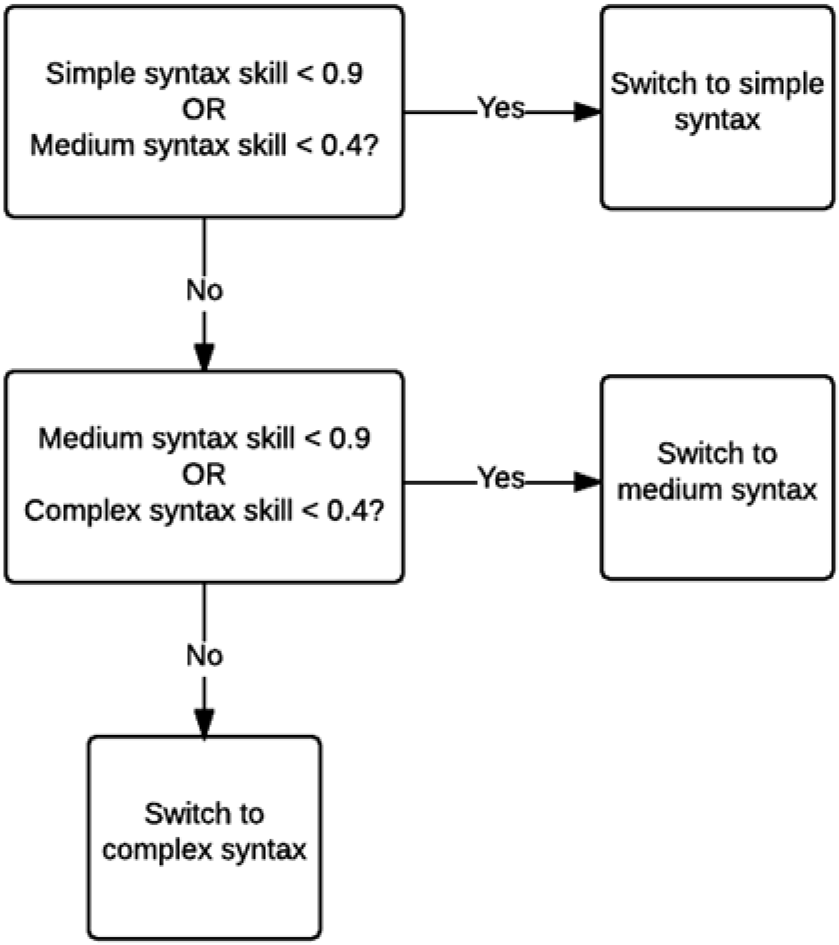

Simple syntax skill < 0.9
OR
Medium syntax skill < 0.4?
Yes
Switch to simple syntax
No
Medium syntax skill < 0.9
OR
Complex syntax skill < 0.4?
Yes
Switch to medium syntax
No
Switch to complex syntax

Figure 9. Overview of adapting complexity of syntax.

Mastery of a particular syntax complexity level is set to 0.9. (Low skill is indicated by values below 0.4.) The idea is that students cannot move onto a higher syntax complexity level until the current one has been mastered. Additionally, students cannot skip syntax complexity levels. The EMBRACE ITS will only ever switch between simple and medium syntax or between medium and complex syntax.

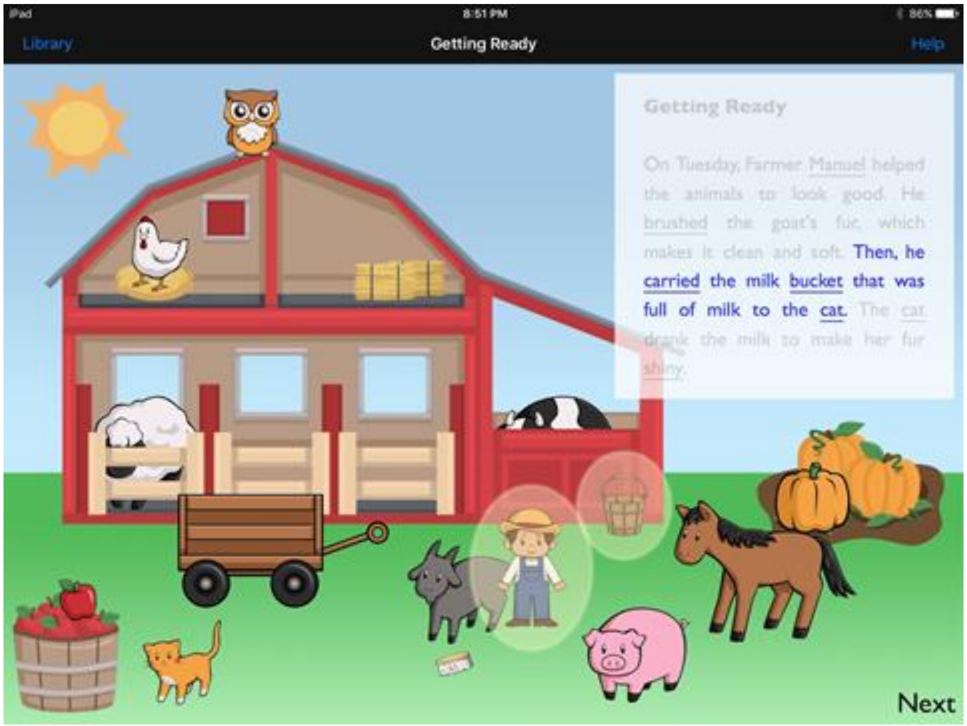
Since syntax complexity is medium by default, the system converts medium sentences to either complex or simple sentences using the corresponding idea numbers and solution steps. Complex sentences often merge multiple idea numbers associated with two or more medium sentences, while simple sentences often split idea numbers associated with a single medium sentence. Consider the two medium sentences: “Manuel first put the horse into the corral, and then he put the cow into the corral” and “He can clean the barn better with big animals in the barn.” Suppose the first sentence is associated with idea number 1, while the second sentence is associated with idea number 2. The complex version combines both ideas into a single sentence: “Manuel first put the horse and then the cow into the corral, because he can clean the barn better with the big animals in the corral.” However, the simpler version consists of three sentences: “Manuel put the horse into the corral”; “Then Manuel put the cow into the corral”; and “He can clean the barn better with big animals in the corral.” The first two simple sentences share idea number 1, while the third simple sentence directly maps to idea number 2. If a sentence with the desired complexity cannot be found to express a particular idea or step, then the system will choose the medium sentence instead. This decision ensures that the same ideas and steps are used in the same order across all complexities.

Determining Error Feedback Type

The system determines appropriate error feedback to provide each time skills are updated when the user makes an error. From the list of updated skills, the system selects the type corresponding to the lowest skill value that falls under a threshold of 0.5. This threshold is the same across all error types. Once skill values are above 0.5, students are either unlikely to require error feedback, or they have the opportunity to correct errors on their own before receiving help. Note that for any error, the system only updates a single syntax skill or a single usability skill, but it may update multiple vocabulary skills at the same time. Thus, an average of the updated vocabulary skill values is compared to the threshold instead. If no updated skill values fall below the threshold, then the system does not provide any additional error feedback. Since skill values start low at the beginning of a story, error feedback is provided more frequently than later in the story as the child increases in skills throughout the texts.

Providing Feedback for Vocabulary Errors

The system provides vocabulary error feedback by playing a feedback noise and temporarily highlighting the correct objects or object and location involved in the current step. Figure 10 shows an example: If the user made a vocabulary error in the sentence "Then, he carried the milk bucket that was full of milk to the cat" by moving the farmer to the sheep instead of the bucket, then the system would highlight the farmer and the bucket images. In another example, if the user made a vocabulary error in the sentence "The farmer put the pig into its pen" by moving the group containing the farmer and pig to the corral instead of the pen, then the system would highlight the pig image and pen location. The entities involved are highlighted simultaneously, so there is no indication of which object should be moved to the other in the case of two objects being highlighted at the same time. This feedback simply lets the user know exactly which objects and locations are involved in the step.



Getting Ready

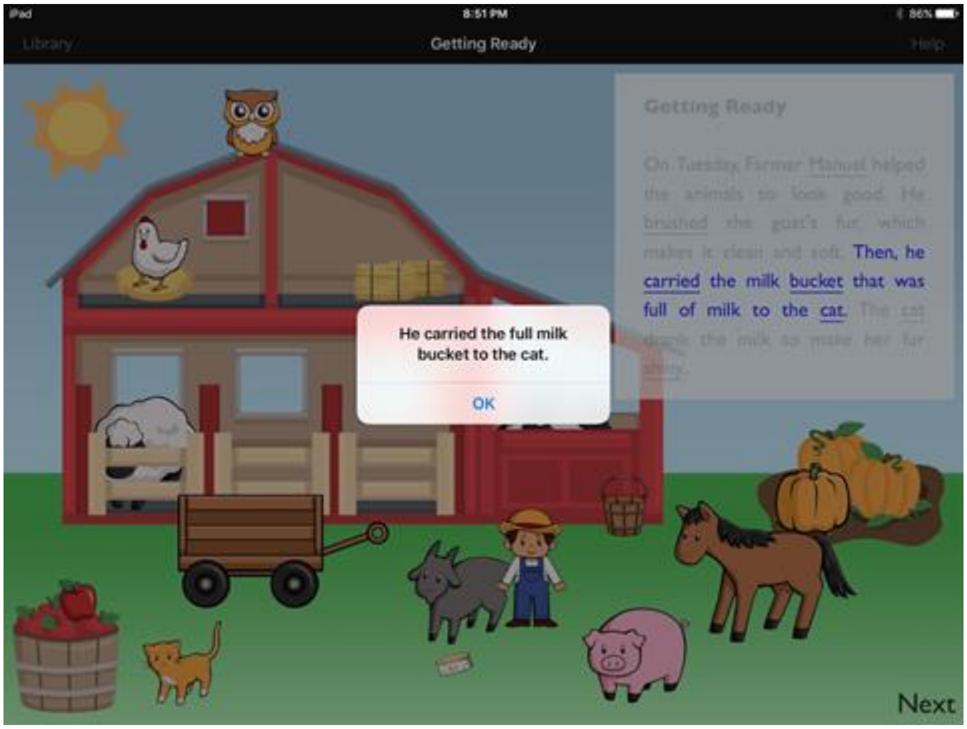
On Tuesday, Farmer Manuel helped the animals to look good. He brushed the goat's fur, which makes it clean and soft. Then, he carried the milk bucket that was full of milk to the cat. The cat drank the milk to make her fur shiny.

Figure 10. Vocabulary error feedback.

Work on this feature was split between highlighting objects only, which I implemented, and highlighting an object and a location, which other project members implemented.

Providing Feedback for Syntax Errors

The system provides syntax error feedback by playing a feedback noise and displaying a popup message with the simpler version of the current sentence that contains the current step. Once the user dismisses this popup message by pressing the "OK" button, he or she can resume manipulation (see Figure 11). For multi-step sentences, if the user makes an error on the first step, for example, then the popup message will only contain the simpler version of the current sentence for the first step. If the user makes an error on the second step, then the popup message will only contain the simpler version of the current sentence for the second step. If no simpler version is available, then the system simply defaults to providing usability error feedback instead. This is because a bottom-out hint may be the best form of feedback for users having trouble understanding simple sentences. For example, if the student made a syntax error in the sentence “Then, he carried the milk bucket that was full of milk to the cat” by moving the cart to the barn before having the farmer put a pumpkin inside the cart, the system would show a popup with the simpler version: “The farmer put a pumpkin in the cart.” Suppose the student correctly advances the steps to the point where the pumpkin is in the cart, but he or she then proceeds to move the pumpkin and cart to the barn without the farmer. The system would then show the popup with the message “Then the farmer pulled the cart back to the barn.”



Getting Ready

On Tuesday, Farmer Manuel helped the animals to look good. He brushed the goat's fur, which makes it clean and soft. Then, he carried the milk bucket that was full of milk to the cat. The cat drank the milk to make her fur shiny.

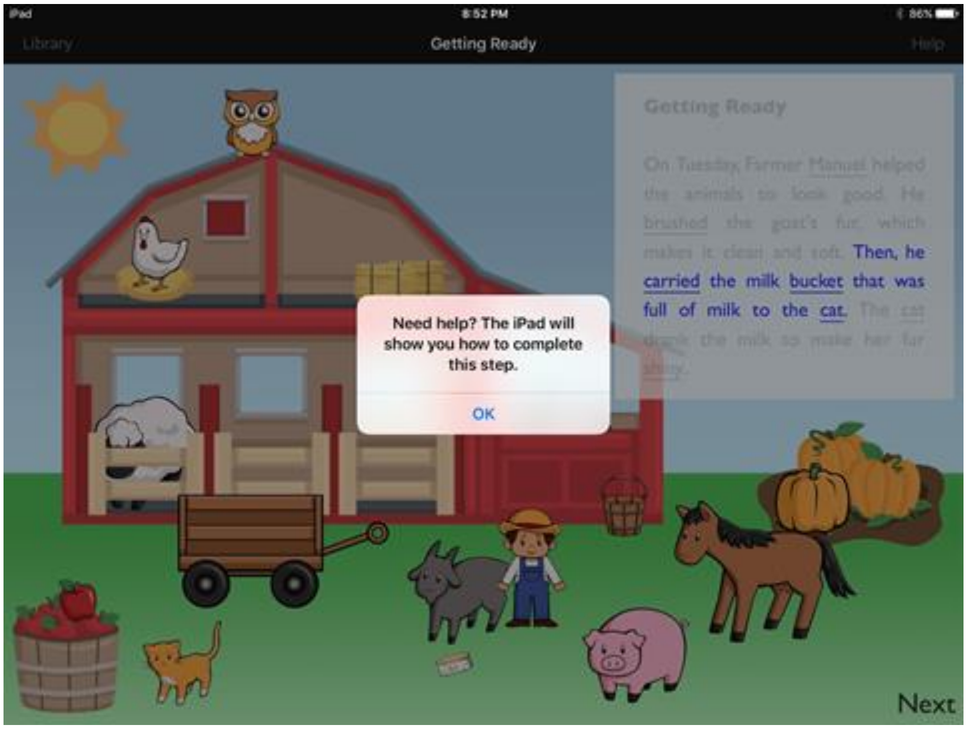
He carried the full milk bucket to the cat.

Figure 11. Syntax error feedback.

The user can view this message for as long as they require. However, once the message is dismissed, only the original sentence is displayed in the textbox. This makes the feedback temporary, similar to the feedback for vocabulary errors.

Providing Feedback for Usability Errors

The system provides usability error feedback by playing a feedback noise and displaying a popup message that lets the user know that the system will show him or her how to complete the current step (see Figure 12). Once the user dismisses this message by pressing the "OK" button, the system will temporarily highlight the correct object, animate the correct object being moved to the other correct object or correct location, and then temporarily highlight the correct object again. This is the same type of feedback that the user receives when he or she makes five errors on the same step. It is essentially a bottom-out hint that allows the user to move on with the sentence.



Getting Ready

On Tuesday, Farmer Manuel helped the animals to look good. He brushed the goat's fur, which makes it clean and soft. Then, he carried the milk bucket that was full of milk to the cat. The cat drank the milk to make her fur shiny.

Need help? The iPad will show you how to complete this step.

Figure 12. Usability error feedback.

Animation for certain step types was implemented in part by other project members and then later refined by me.

Summary of System Description

In this section, I have provided an overview of how the EMBRACE ITS system works, including the classification of skills into vocabulary, syntax, and usability types and how different user actions can increase or decrease these skills based on the level of correctness. I also discussed how the system adapts the complexity of vocabulary by adding words that the user struggled with previously and how the system adapts the complexity of syntax by setting the next chapter to comprise of either simple, medium, or complex sentences. Then I discussed how the system chooses the most probable error type to provide feedback. I showed vocabulary error feedback, which highlights the correct images involved in the step, syntax error feedback, which shows a simpler version of the current sentence in a popup message, and usability error feedback, which animates completing the step for the user.

CHAPTER 6

STUDY

The goals of the study were to evaluate how well each of the features--skill updates, complexity adaptations, and error feedback—functioned in the EMBRACE ITS and whether these features made reading with the EMBRACE ITS more helpful than reading with EMBRACE alone.

The study was conducted with 7 non-English Language Learner students from an after school program. Their ages ranged from 7 to 11 years old, and there were 4 boys and 3 girls. The study took one hour a day for two days spread out over two weeks. On Day 1, students completed the introduction to EMBRACE to learn how physical manipulation works. Then they completed all chapters of a narrative story, either Bottled Up Joy, which contains 5 chapters, or The Lopez Family Mystery, which contains 6 chapters. The first chapter is read aloud to them, while they have to read the remaining chapters aloud by themselves. Each chapter begins with the vocabulary introduction list and ends with assessment questions. Using a within-group study design, students were randomly assigned to either EMBRACE or EMBRACE-ITS. Depending on the condition of his or her group, the student reads the first book using EMBRACE alone or EMBRACE ITS. On Day 2, students completed the remaining book using EMBRACE alone or EMBRACE ITS. Then they completed exit interviews to discuss their experiences reading using the system. The exit interview included questions about the participants' reading habits, opinions using the iPad application, and whether or not they thought each type of adaptation and feedback was helpful. All user interactions with the system were logged, and the answers to the group exit interview were audio recorded.

Study data can be extracted from both the log data and exit interview answers. From the log data, I can obtain the types of error feedback participants received, extra words listed in vocabulary adaptations, words that participants requested, change in skill values, syntax complexity per chapter, errors that participants made, and assessment scores. I can also contextualize most of this information using the log data that records which objects participants moved and where. From the exit interview answers, I can gain a general sense of the kinds of readers that participated and their reactions to using the system.

CHAPTER 7

RESULTS

Of the 7 participants, only 5 completed both stories, and only 4 participated in the exit interviews. Because the study had few participants with a lot of incomplete data, I will not be performing any statistical analysis of the results. Instead, I will be analyzing individual participants' behaviors. The results of the study will be used to help answer the following questions in alignment with the study goals:

1. How many of each type of error feedback did participants receive? Which type did participants receive the most? The least?
2. How many extra vocabulary words did participants receive?
3. How many words did participants request in the EMBRACE condition compared to the EMBRACE ITS condition?
4. How did vocabulary skills change over time? How did error feedback and adaptation affect vocabulary skills?
5. How did syntax complexity change over time?
6. How did syntax skills change over time? How did error feedback and adaptation affect syntax skills?
7. How does the usability skill compare between the EMBRACE condition and the EMBRACE ITS condition?
8. How many errors per user step did participants make in the EMBRACE condition compared to the EMBRACE ITS condition?
9. How many errors per assessment question did participants make in the EMBRACE condition compared to the EMBRACE ITS condition?

This section contains tables summarizing the data from the study. Some cells are either blacked out or left blank due to incomplete data. The section also contains line graphs where EMBRACE participants are represented by solid lines, while EMBRACE ITS participants are represented by dashed lines.

Overall Error Feedback

How many of each type of error feedback did participants receive? Which type did participants receive the most? The least?

Table 2 totals the number of each type of feedback that each participant received while reading in the EMBRACE ITS condition. Participant itsp05 did not receive any feedback, as he or she made few errors throughout the story, and skill values were too high to warrant receiving feedback for those errors. On the other hand, participant itsp04 received the most feedback--nearly double the next highest amount of feedback, and the types of feedback received are relatively even in amount. Overall, participants received syntax error feedback the most, followed by usability error feedback, and then vocabulary error feedback.

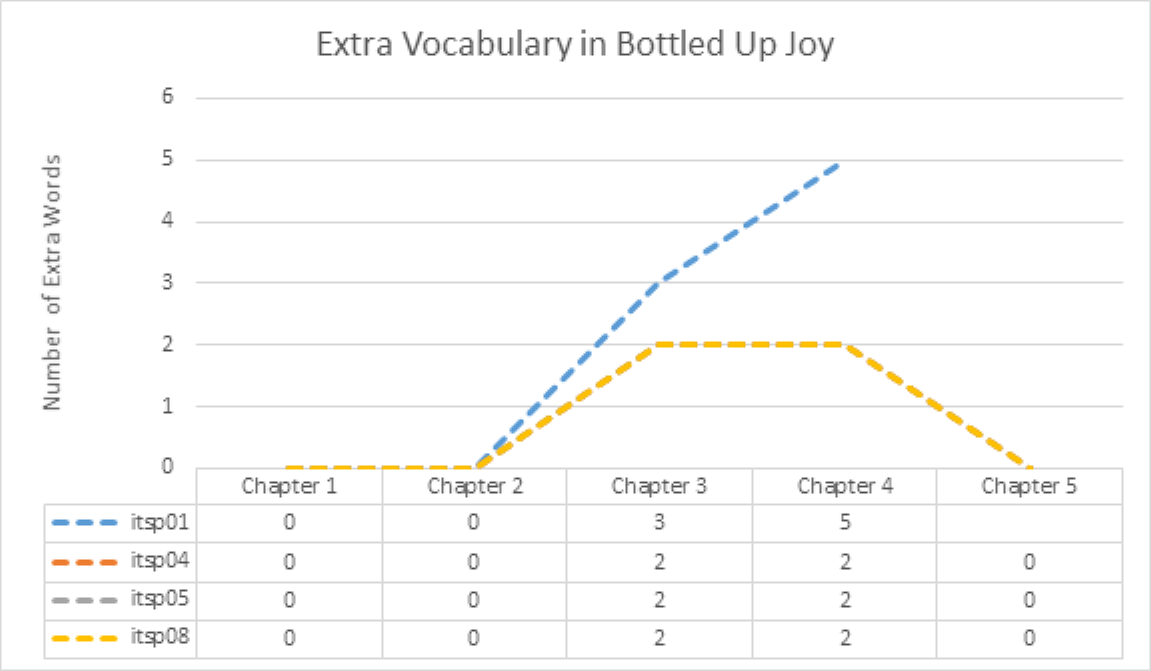
Participant	Vocabulary	Syntax	Usability	Total
itsp01	2	0	1	**3**
itsp03	[redacted]	[redacted]	[redacted]	[redacted]
itsp04	3	4	4	**11**
itsp05	0	0	0	**0**
itsp06	2	0	4	**6**
itsp07	2	4	0	**6**
itsp08	0	3	2	**5**
Total	**9**	**12**	**10**	**31**
AVG	**2.571**	**3.286**	**2.857**	**8.857**
STDEV	**1.225**	**2.041**	**1.835**	**3.656**

Table 3. Number of each type of error feedback received.

Vocabulary Skills, Adaptations, and Feedback

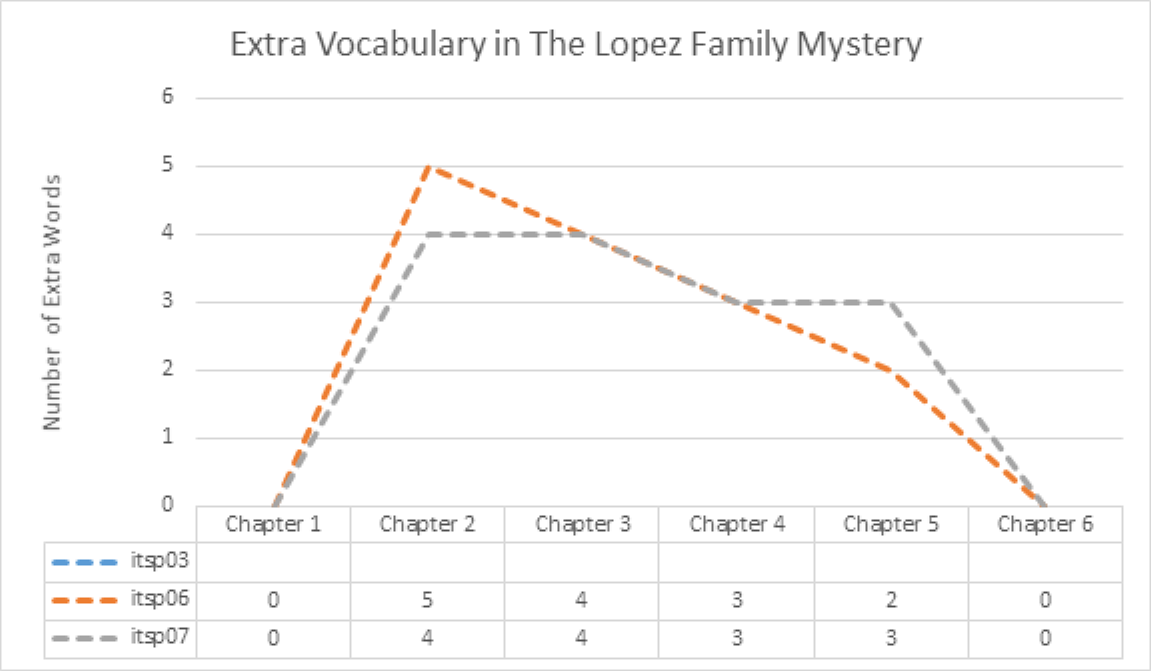
How many extra vocabulary words did participants receive?

Figures 13 and 14 show the number of extra vocabulary per chapter that participants received when the vocabulary lists were adapted while reading Bottled Up Joy or The Lopez Family Mystery, respectively, using the EMBRACE ITS:



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
itsp01	0	0	3	5	
itsp04	0	0	2	2	0
itsp05	0	0	2	2	0
itsp08	0	0	2	2	0

Figure 13. Number of extra vocabulary per chapter in Bottled Up Joy. Note that participant itsp08 overlaps participants itsp04 and itsp05.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
itsp03						
itsp06	0	5	4	3	2	0
itsp07	0	4	4	3	3	0

Figure 14. Number of extra vocabulary per chapter in The Lopez Family Mystery.

Participants received no extra vocabulary in the last chapter of both stories. Bottled Up Joy readers also received no extra vocabulary in the second chapter, and they received fewer words than The Lopez Family Mystery readers overall. When extra vocabulary was added, participants received at least two words per chapter. Also of note is that participant itsp01 received double the amount that each of the other participants received while reading the same story. In The Lopez Family Mystery, participants appear to receive more words in the earlier chapters compared to the later chapters.

How many words did participants request in the EMBRACE condition compared to the EMBRACE ITS condition?

Table 3 totals the number of tapped vocabulary in the chapter texts for each participant in each condition.

Participant	**EMBRACE**	**EMBRACE ITS**	**Total**	**AVG**	**STDEV**
itsp01	18	16	**34**	**17.000**	**1.414**
itsp03	2	■	**2**	**2.000**	■
itsp04	1	1	**2**	**1.000**	**0.000**
itsp05	1	0	**1**	**0.500**	**0.707**
itsp06	■	1	**1**	**1.000**	■
itsp07	9	0	**9**	**4.500**	**6.364**
itsp08	0	4	**4**	**2.00**	**2.823**
Total	**31**	**22**	**53**	**26.500**	**6.364**
AVG	**5.167**	**3.667**	**8.833**	**4.417**	**1.061**
STDEV	**7.083**	**6.218**	**11.984**	**5.881**	**0.611**

Table 4. Number of tapped vocabulary.

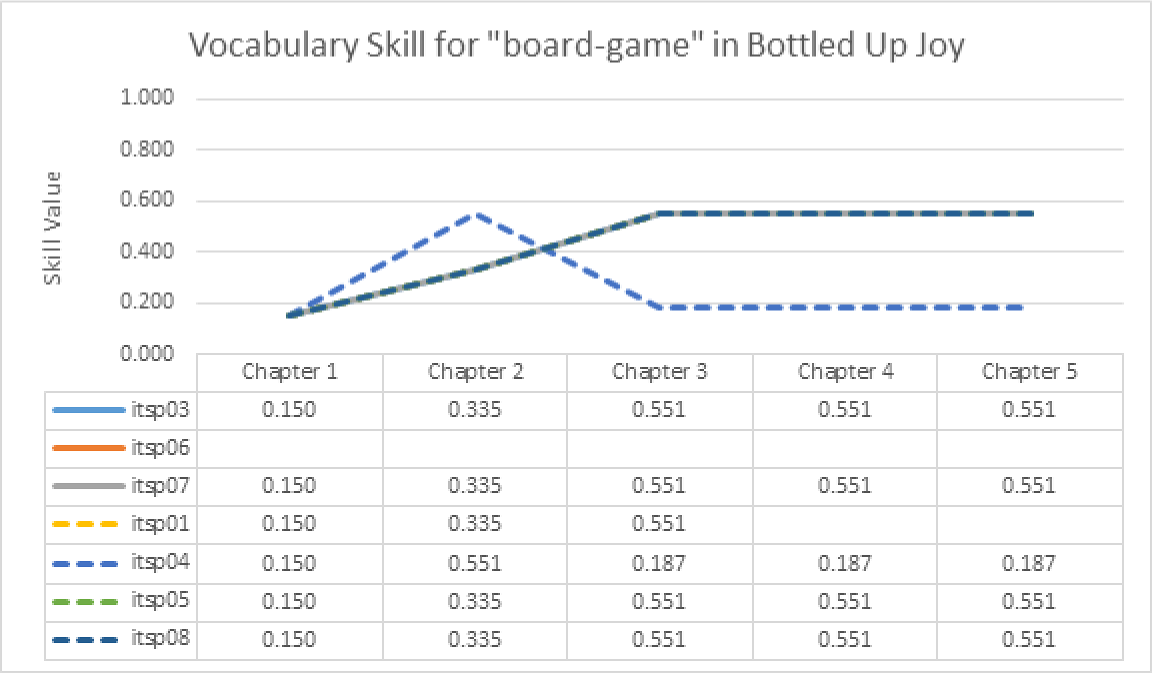
Participant itsp01 requested the most vocabulary in both conditions. Participant itsp07 also requested a large number of vocabulary compared to the other participants in the EMBRACE condition. However, he or she did not request any vocabulary while using the EMBRACE ITS. In general, most participants did not request much vocabulary with some participants who requested no vocabulary throughout an entire story. Given how little users tapped on underlined words in the chapters, any words that were added to the vocabulary introduction lists would mostly comprise of words related to the solution steps.

How did vocabulary skills change over time? How did error feedback and adaptation affect vocabulary skills?

Since there are many different vocabulary skills pertaining to the solution steps for each story, I chose to focus the next analysis on the change in skill value for two words corresponding to objects that appear in extra vocabulary or vocabulary error feedback and are manipulated frequently enough for users to have multiple opportunities to practice them. These words are “board-game” in Bottled Up Joy and “baby” in The Lopez Family Mystery.

The following figures will show the change in vocabulary skill for each word throughout the chapters for both conditions. The skill value shown is the final value at the end of the chapter.

Figure 15 examines the word “board-game” in Bottled Up Joy.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
itsp03	0.150	0.335	0.551	0.551	0.551
itsp06					
itsp07	0.150	0.335	0.551	0.551	0.551
itsp01	0.150	0.335	0.551		
itsp04	0.150	0.551	0.187	0.187	0.187
itsp05	0.150	0.335	0.551	0.551	0.551
itsp08	0.150	0.335	0.551	0.551	0.551

Figure 15. Final vocabulary skill per chapter for “board-game” in Bottled Up Joy. Solid lines represent EMBRACE participants, while dashed lines represent EMBRACE ITS participants. Note that participants itsp03 and itsp07 are overlapping, as well as itsp01 (up to chapter 3) with itsp05 and itsp08.

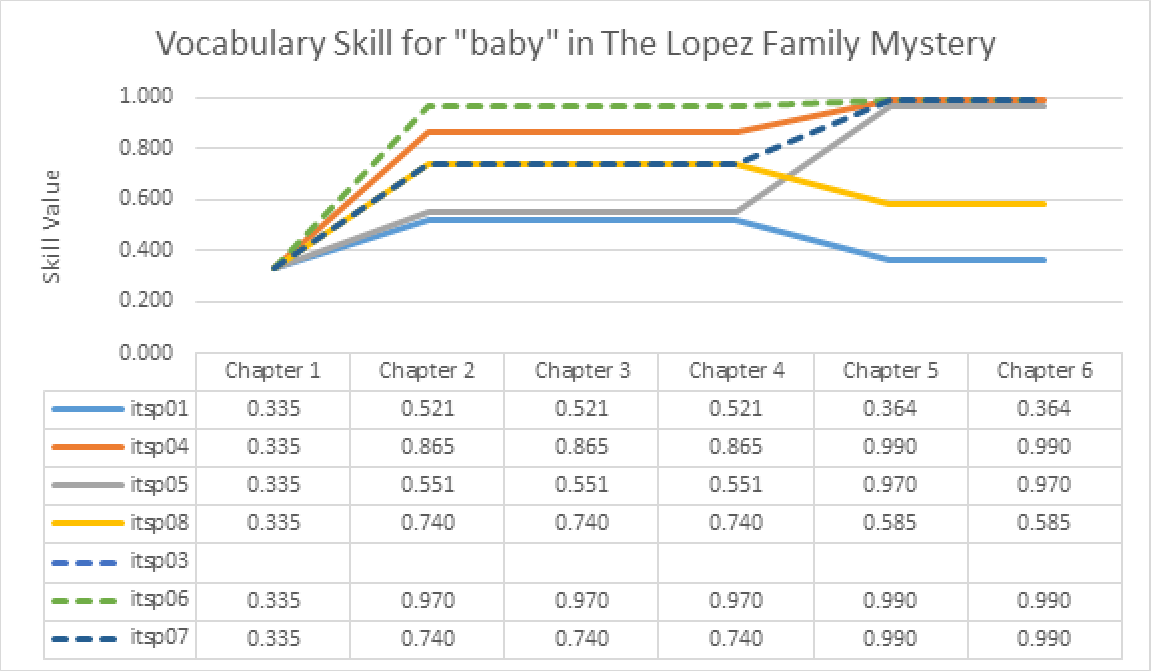
The object corresponding to “board-game” is only manipulated in chapters 2 and 3. However, it appears in the vocabulary list of chapter 2 by default, and it was included as extra vocabulary for all EMBRACE ITS participants in chapter 3. This should have increased the skill value above that of the EMBRACE only condition, but the system did not correctly map the “board-game” word to the object encoded as “boardgame” in the solution steps.

Participant itsp01 was the only user in both conditions to request this word in chapter 2 during the medium complexity sentence “Then, he opened the board-game instead." However, the request did not decrease his or her skill value, so he or she was still able to reach the same skill value as most of the other participants.

Participant itsp04 was actually able to increase his or her skill value to 0.551 in chapter 2 due to an error. This error occurred in the same sentence mentioned previously for itsp01. The sentence required moving a character to the "board-game" object, and while the participant moved the wrong subject, he or she still moved to the correct destination, indicating correct knowledge of the "board-game" word. Thus, itsp04 received an extra opportunity to increase the skill value, while other participants had only one opportunity. However, his or her skill drastically decreased in the next chapter due to the same type of step. The participant continuously tried moving the "board-game" object by itself, which resulted in multiple vocabulary error feedbacks that highlighted the character and the "board-game" object simultaneously. In this situation, the system should have ideally presented the participant with syntax error feedback instead. It did not detect the syntax error as it saw the user moving the relevant "board-game" object to an irrelevant destination.

Overall, participants in both conditions did not appear to have much trouble with this word, so adding the word as extra vocabulary in chapter 3 probably made little difference.

Figure 16 examines the word "baby" in The Lopez Family Mystery.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
itsp01	0.335	0.521	0.521	0.521	0.364	0.364
itsp04	0.335	0.865	0.865	0.865	0.990	0.990
itsp05	0.335	0.551	0.551	0.551	0.970	0.970
itsp08	0.335	0.740	0.740	0.740	0.585	0.585
itsp03						
itsp06	0.335	0.970	0.970	0.970	0.990	0.990
itsp07	0.335	0.740	0.740	0.740	0.990	0.990

Figure 16. Final vocabulary skill per chapter for “baby” in The Lopez Family Mystery. Solid lines represent EMBRACE participants, while dashed lines represent EMBRACE ITS participants.

The object corresponding to “baby” is manipulated in chapters 1, 2, and 5. Participant itsp01 requested this word once in chapter 5 during the medium complexity sentence “Later in the day, the baby wanted to play with her rattle.” The word was not involved in any vocabulary error feedback, but it was part of the extra vocabulary for itsp06 in chapter 2 and for itsp07 in chapters 2 and 5.

Just like the previous word, both groups end at 0.335 in chapter 1, and differences start appearing in chapter 2. Participant itsp01 made a lot of errors using the “baby” object incorrectly; most of these errors involve moving a different object rather than the correct “baby” object. An example sentence with medium complexity was “She looked in the living room, but she couldn't find her rattle there either.” The “she” in this sentence refers to the “baby” object, but the participant moved a different female character instead, which might indicate a need for a pragmatics category within the syntax skill. In chapter 5, the skill decreased a lot. This may be partly due to the system incorrectly identifying an error instead of a usability issue.

On the other hand, participants itsp04 and itsp05 made few errors using the “baby” object throughout the texts, and when they did make an error, it still demonstrated correct knowledge of the word. Participant itsp08 made few errors as well, but his or her errors did not end up increasing the skill value. In chapter 5, the skill value decreased in a step that required moving an object to the “baby” object, which the participant appeared to perform correctly, but the system kept wrongly identifying it as a vocabulary error.

As for the performance of the EMBRACE ITS participants, both increased their averages throughout the story. The appearance of the word as extra vocabulary boosted their skill values in chapters 2. However, participant itsp06 had a higher average as a result of making multiple errors that still demonstrated correct knowledge of the word. Both participants maxed out the skill in chapter 5.

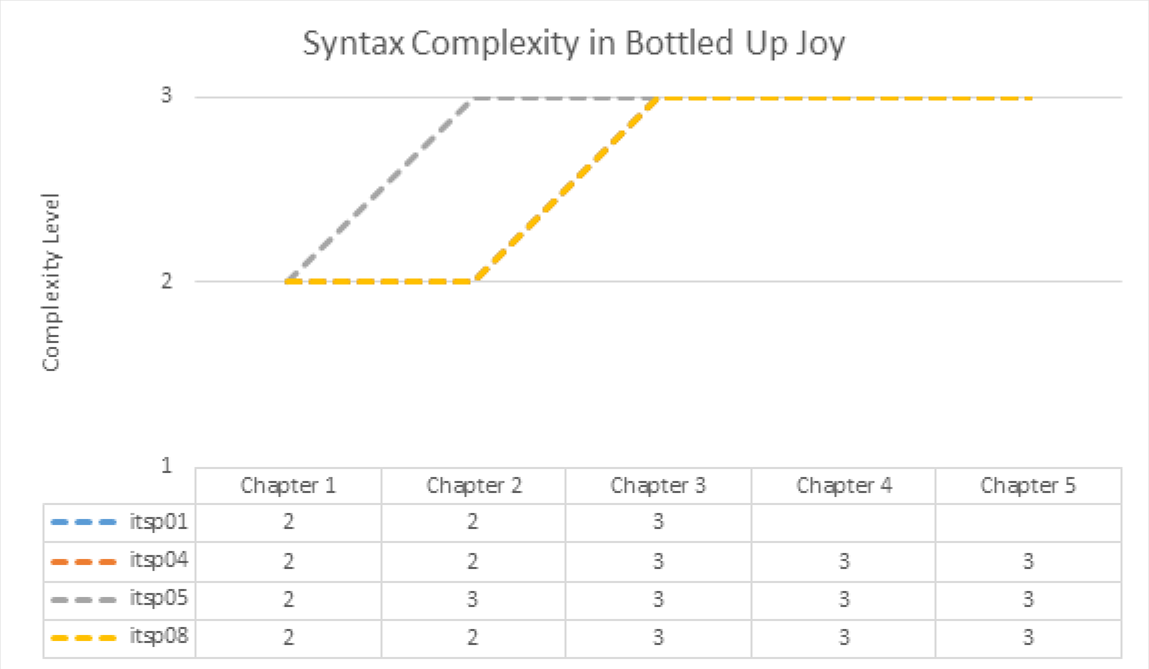
Overall, most participants across both groups appeared to understand the word “baby” well, but EMBRACE ITS users were able to max out the skill slightly faster.

Similar analyses were conducted for other words across both stories. Overall, vocabulary adaptations mainly helped to boost EMBRACE ITS users’ vocabulary skills, but without that boost, there was little difference between the skill values of EMBRACE ITS and EMBRACE users. Users did seem to respond well to vocabulary error feedback, as most users were able to perform the correct step after receiving the feedback. However, the system ran into a few bugs where users were sometimes not credited for correct actions. Additionally, users were often able to increase their vocabulary skills multiple times despite making errors because these errors still demonstrated correct knowledge of the word; as a result, these users often had higher skill values than users that did not make any errors using the word.

Syntax Skills, Adaptation, and Feedback

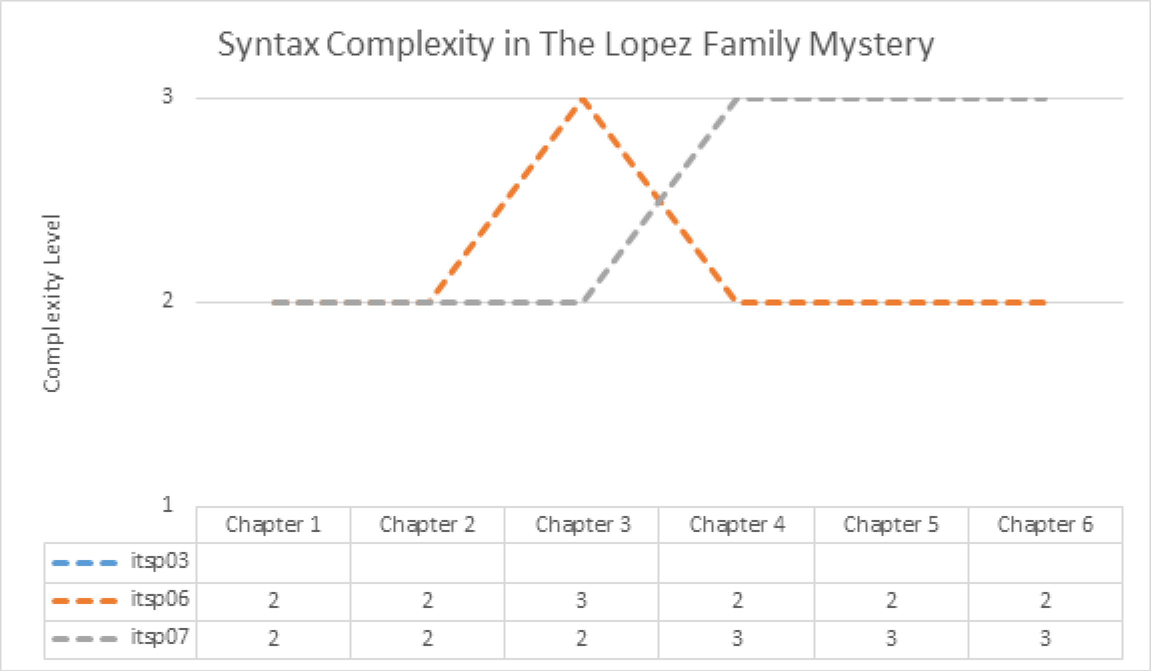
How did syntax complexity change over time?

Figures 17 and 18 show the syntax complexity per chapter for participants using the EMBRACE ITS to read either Bottled Up Joy or The Lopez Family Mystery, respectively. A 1 indicates simple syntax, 2 indicates medium syntax, and 3 indicates complex syntax.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
itsp01	2	2	3		
itsp04	2	2	3	3	3
itsp05	2	3	3	3	3
itsp08	2	2	3	3	3

Figure 17. Syntax complexity per chapter in Bottled Up Joy. Note that participant itsp08 overlaps participants itsp04 and itsp01 (up to chapter 3).



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
itsp03						
itsp06	2	2	3	2	2	2
itsp07	2	2	2	3	3	3

Figure 18. Syntax complexity per chapter in The Lopez Family Mystery.

In both stories, most participants reached the complex syntax level by the third chapter. All participants except itsp06 remained at the complex syntax level for the remainder of the story. No participants ever dropped to the simple syntax level.

How did syntax skills change over time? How did error feedback and adaptation affect syntax skills?

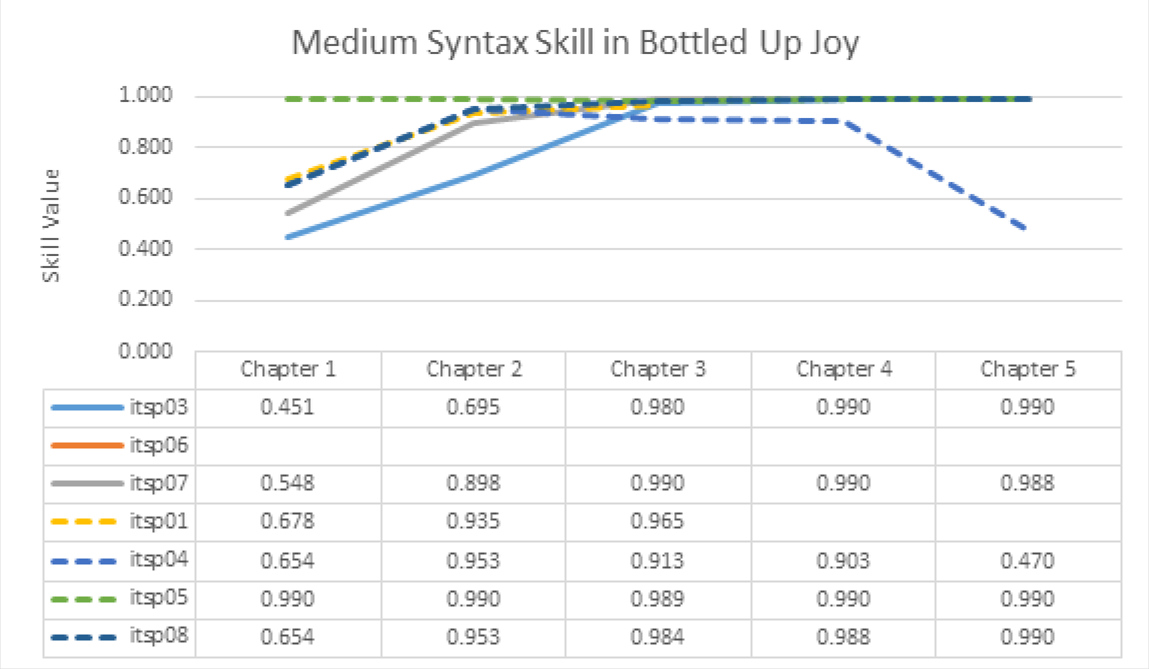
Table 4 summarizes the average medium syntax skill for each participant in both conditions. The averages were calculated using the medium syntax skill value at the end of each chapter in the story.

Participant	EMBRACE	EMBRACE ITS	AVG	STDEV
itsp01	0.594		**0.594**	
itsp03	0.821		**0.821**	
itsp04	0.936	0.779	**0.857**	**0.111**
itsp05	0.884	0.990	**0.937**	**0.075**
itsp06		0.872	**0.872**	
itsp07	0.883	0.947	**0.915**	**0.046**
itsp08	0.990	0.914	**0.952**	**0.054**
AVG	**0.851**	**0.900**	**0.876**	**0.035**
STDEV	**0.138**	**0.081**	**0.122**	**0.041**

Table 5. Average medium syntax skill.

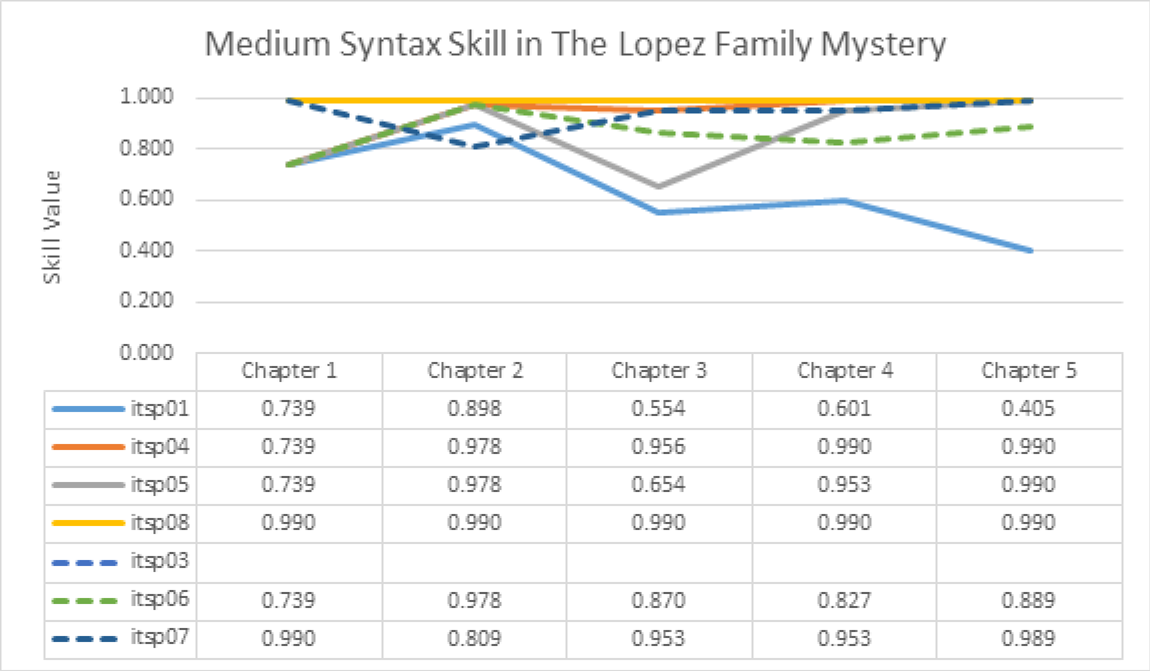
Except for participant itsp01, all participants for which there are data reached above 0.800 in the EMBRACE condition. Participant itsp08 maxed out the medium syntax skill, while itsp05 and itsp07 nearly reached mastery level. Participant itsp04 almost maxed out the medium syntax skill in the EMBRACE condition, yet his or her medium syntax skill is significantly lower in the EMBRACE ITS condition, though not low enough to switch back from complex syntax to medium syntax. However, the remaining participants stayed relatively the same or increased. Overall, participants had high medium syntax skills in both the EMBRACE and EMBRACE ITS conditions, but the skills are slightly higher in EMBRACE ITS condition. The high averages in the EMBRACE ITS condition supports the fact that most participants advanced to the complex syntax level.

Figures 19 and 20 show the change in medium syntax complexity per chapter for each story. The skill value shown is the final value by the end of the chapter.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
itsp03	0.451	0.695	0.980	0.990	0.990
itsp06					
itsp07	0.548	0.898	0.990	0.990	0.988
itsp01	0.678	0.935	0.965		
itsp04	0.654	0.953	0.913	0.903	0.470
itsp05	0.990	0.990	0.989	0.990	0.990
itsp08	0.654	0.953	0.984	0.988	0.990

Figure 19. Final medium syntax skill per chapter in Bottled Up Joy. Solid lines represent EMBRACE participants, while dashed lines represent EMBRACE ITS participants.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
— itsp01	0.739	0.898	0.554	0.601	0.405
— itsp04	0.739	0.978	0.956	0.990	0.990
— itsp05	0.739	0.978	0.654	0.953	0.990
— itsp08	0.990	0.990	0.990	0.990	0.990
- - - itsp03					
- - - itsp06	0.739	0.978	0.870	0.827	0.889
- - - itsp07	0.990	0.809	0.953	0.953	0.989

Figure 20. Final medium syntax skill per chapter in The Lopez Family Mystery. Solid lines represent EMBRACE participants, while dashed lines represent EMBRACE ITS participants.

In both stories, most participants in both conditions generally increased in medium syntax skill value throughout the chapters.

Figure 21 shows the change in medium syntax skill per chapter for participant itsp01 in both conditions. Participant itsp01 appeared to decrease in skill value while reading in the EMBRACE condition but increase in skill value while reading in the EMBRACE ITS condition. In chapter 3 of the EMBRACE condition, the participant repeatedly made syntax errors by moving the object of a step instead of the subject. One sentence was “He barked and barked, and suddenly, the keys dropped onto the ground.” For this sentence, the participant kept moving the “keys” object by itself. Similar errors were made in the other chapters. When the participant started reading in the EMBRACE

ITS condition, he or she continued to make these kinds of mistakes but received no syntax error feedback because the skill value was still above the threshold. If the participant had read The Lopez Family Mystery using the EMBRACE ITS, then his or her skill value would have been low enough to trigger syntax error feedback.

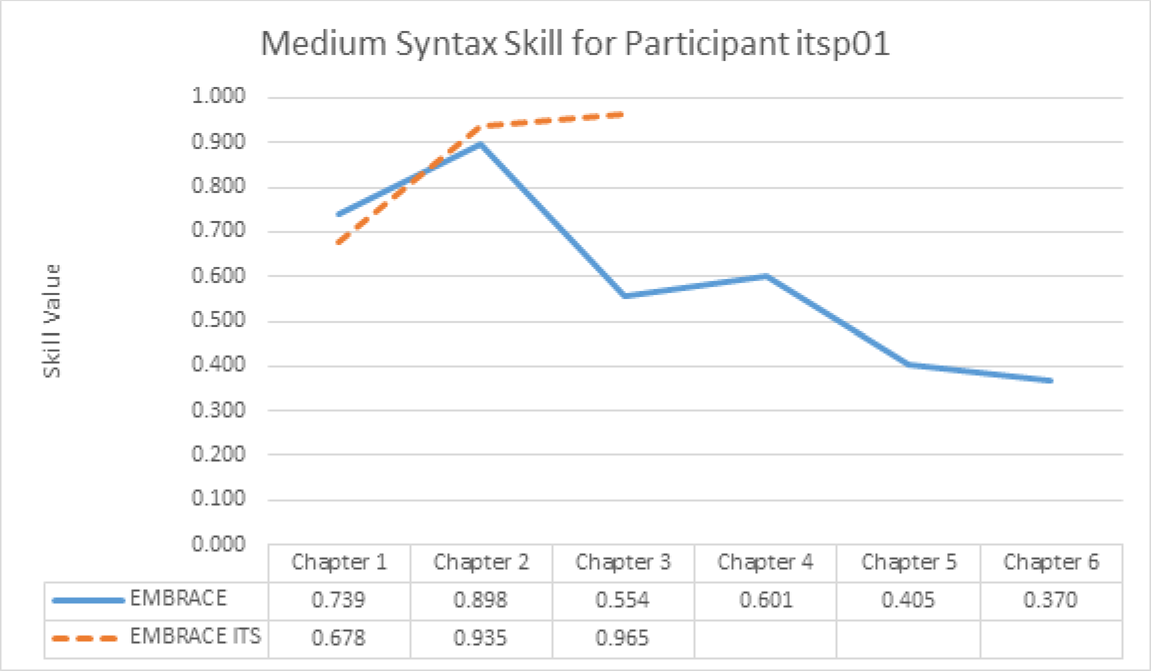
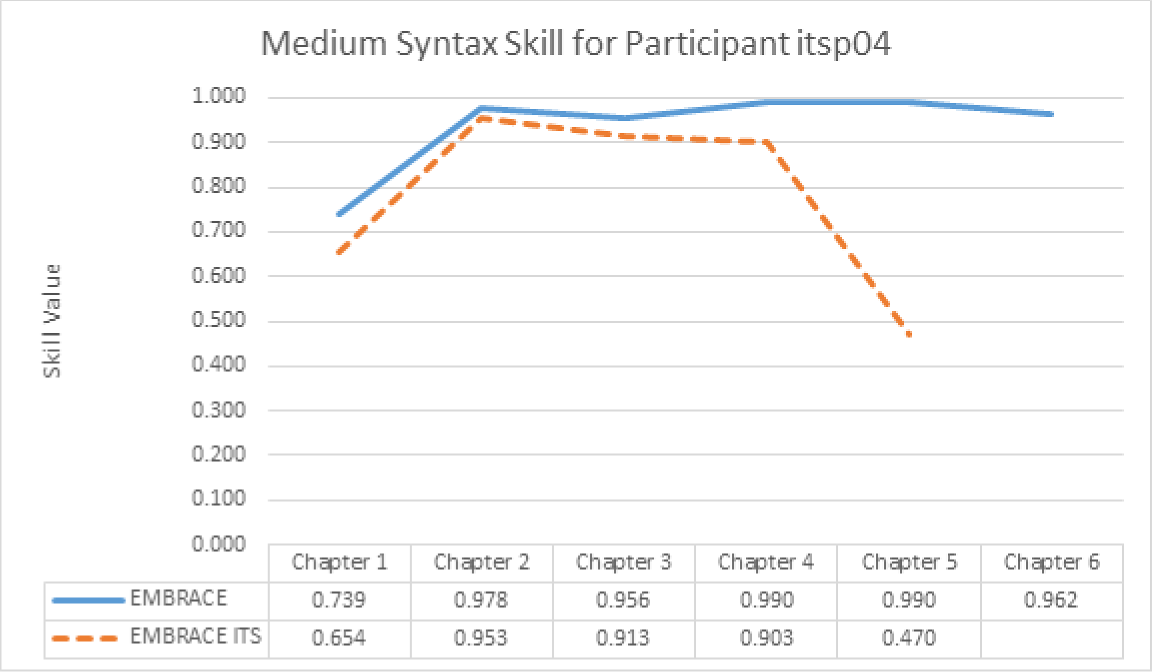


Figure 21. Final medium syntax skill per chapter for participant itsp01.

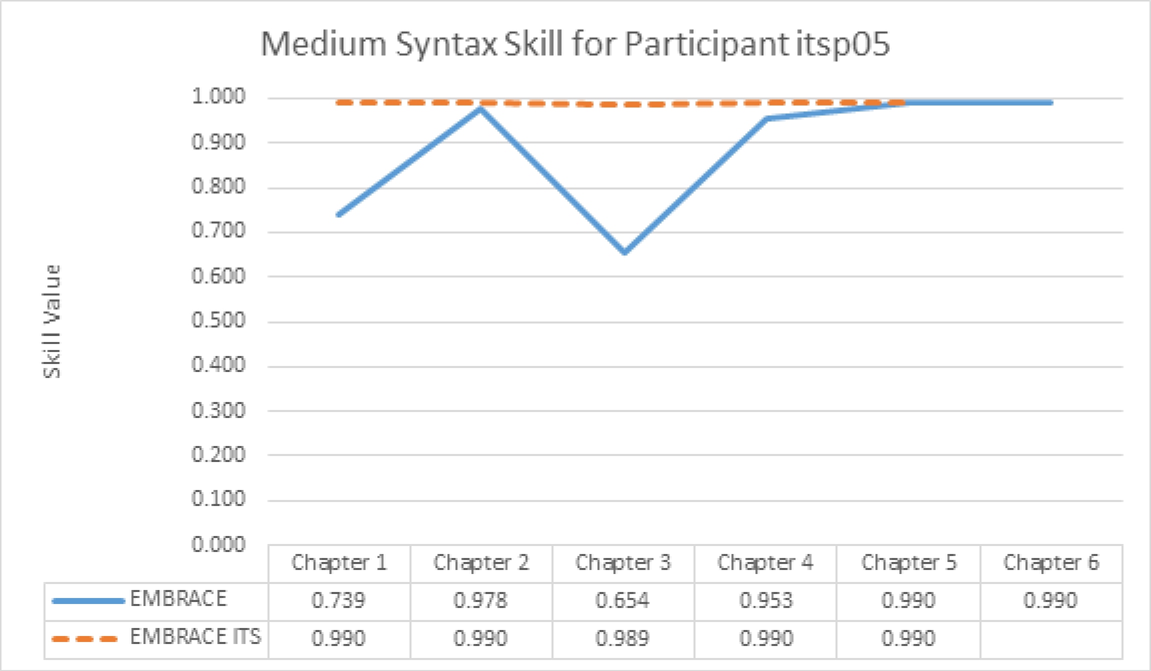
Figure 22 shows the change in medium syntax skill per chapter for participant itsp04 in both conditions. Participant itsp04 seemed to increase in skill value in both conditions but decreased drastically in the last chapter of the EMBRACE ITS condition. He or she repeatedly made errors that still involved objects relevant to the sentence, which drove the syntax skill value low enough to trigger syntax error feedback. One such sentence was of complex syntax: "'Can we just be friends?' asked the boy while picking up the bottle and handing it to her." In this sentence, the participant would move the "bottle" object itself. The system showed the simpler version as "'Can we just be friends?' Lucas picked up the bottle and handed it to her." However, the participant continued to make mistakes until the system automatically completed the step instead.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
EMBRACE	0.739	0.978	0.956	0.990	0.990	0.962
EMBRACE ITS	0.654	0.953	0.913	0.903	0.470	

Figure 22. Final medium syntax skill per chapter for participant itsp04.

Figure 23 shows the change in medium syntax skill per chapter for participant itsp05 in both conditions. Participant itsp05 generally seemed to increase in skill value in the EMBRACE condition, but they decreased significantly in chapter 3. This appears to be caused by a system crash which reset his or her skill values.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
EMBRACE	0.739	0.978	0.654	0.953	0.990	0.990
EMBRACE ITS	0.990	0.990	0.989	0.990	0.990	

Figure 23. Final medium syntax skill per chapter for participant itsp05.

Figure 24 shows the change in medium syntax skill per chapter for participant itsp07 in both conditions. Participant itsp07 seemed to perform well in both conditions, except for the first chapter of the EMBRACE condition. He or she repeatedly switched the subject and object in one particular step, which decreased the skill value. This step was for the medium complexity sentence “He picked it up and carried it home to give to his father.” The sentence required moving the “Lucas” object to the “bottle” object, but the participant moved the objects in the reverse order. In the second chapter, he or she would make the same mistake once but then immediately correct the error afterwards.

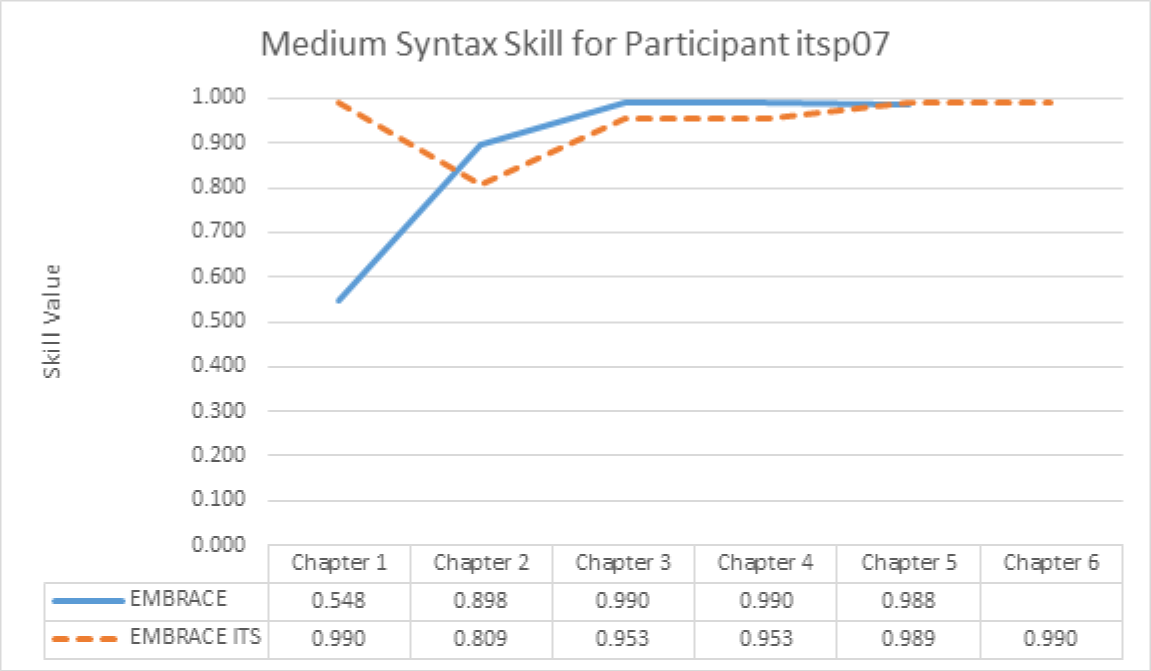
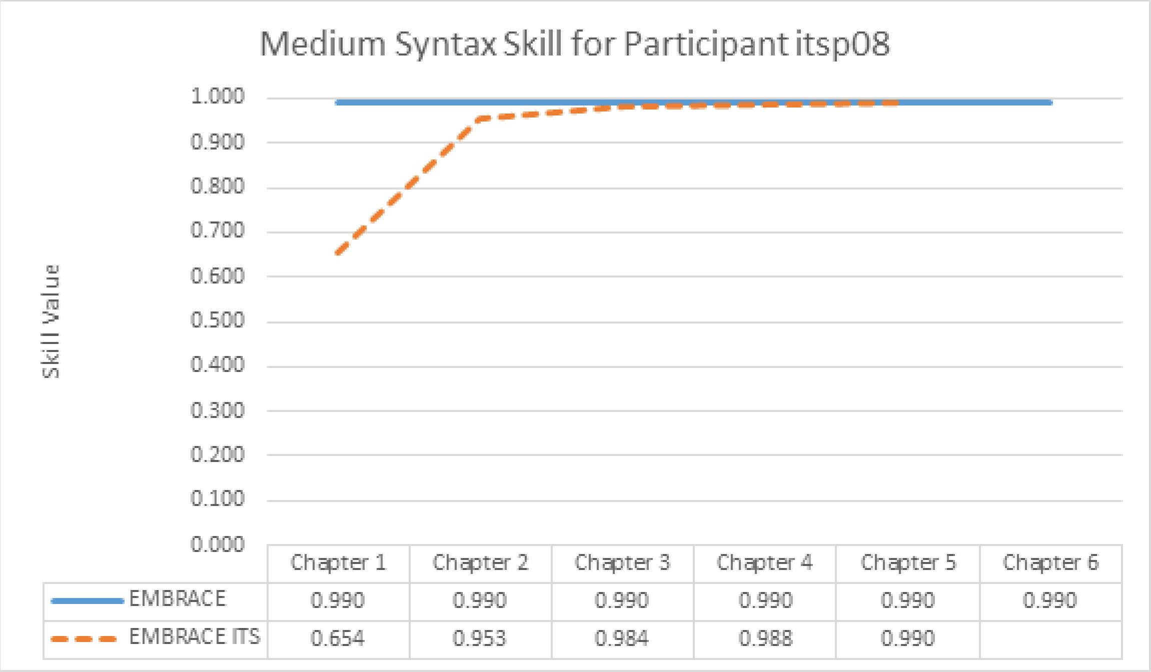


Figure 24. Final medium syntax skill per chapter for participant itsp07.

Figure 25 shows the change in medium syntax skill per chapter for participant itsp08 in both conditions. Participant itsp08 appeared to perform well in both conditions. He or she received syntax error feedback in chapters 3 and 4 after mixing up the subject and object in the step. In chapter 4, the participant made the error on a complex level sentence “His papa told him to follow him as he walked down the hill while carrying the baby.” The participant moved the “Lucas” object before the “papa” object, so the system displayed the simpler version of the sentence: “‘Follow me,’ said Papa as he walked down the hill while carrying the baby.” After receiving the feedback, the participant was able to immediately correct the error.

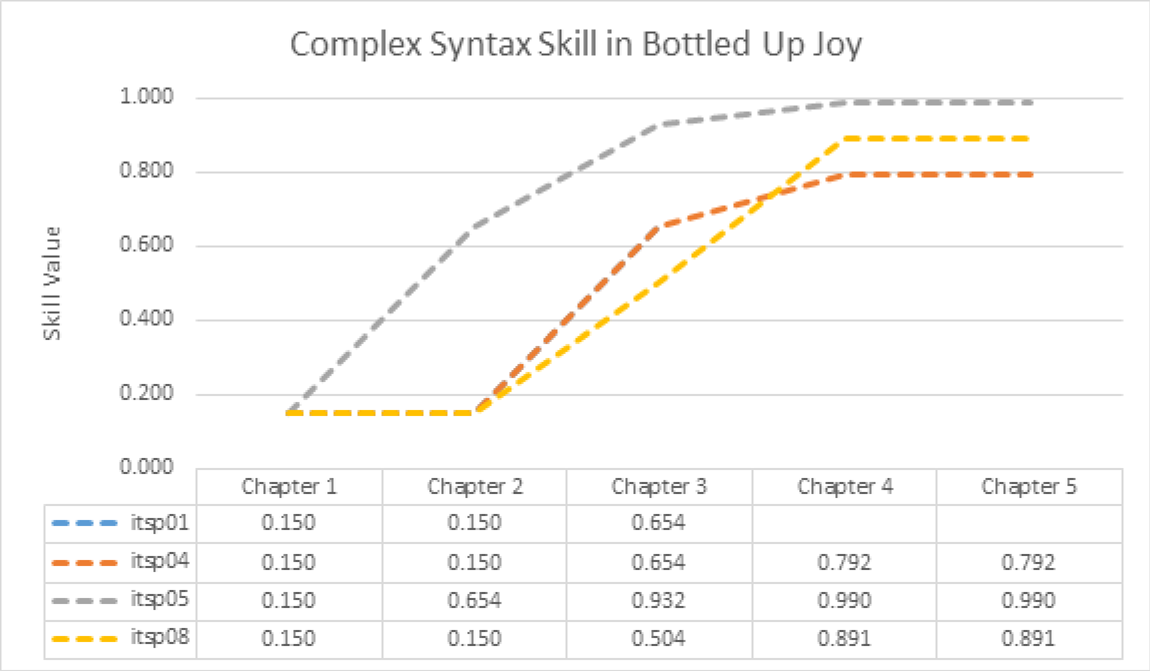


	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
EMBRACE	0.990	0.990	0.990	0.990	0.990	0.990
EMBRACE ITS	0.654	0.953	0.984	0.988	0.990	

Figure 25. Final medium syntax skill per chapter for participant itsp08.

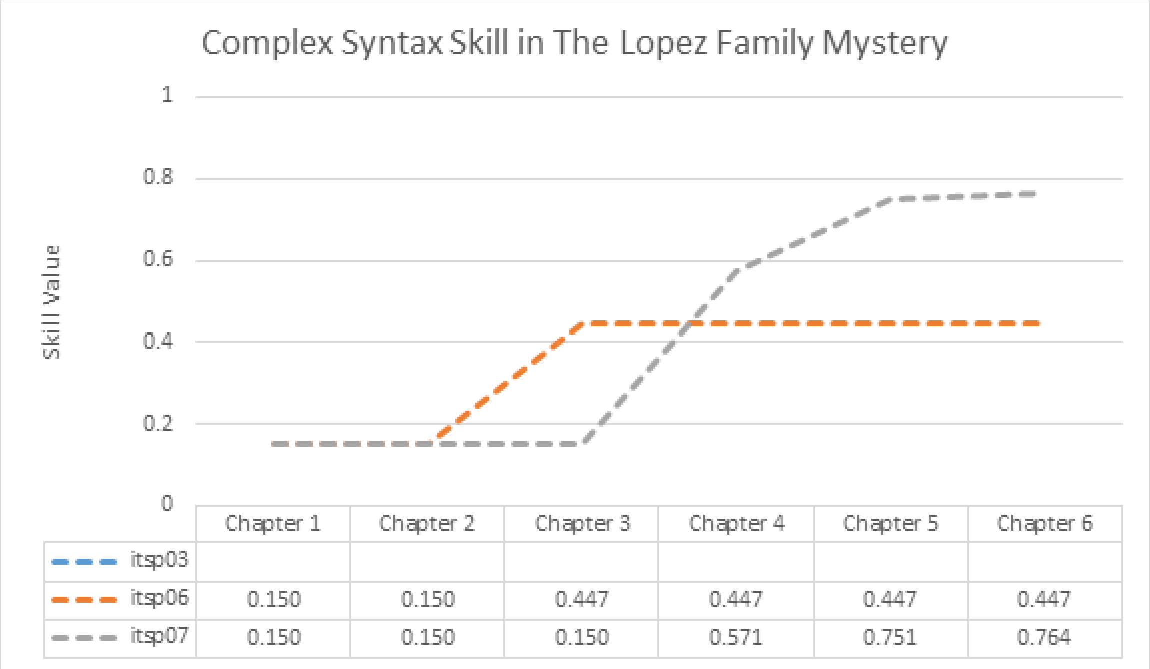
Overall, the system appears to correctly identify syntax errors when participants switched the order of the subject and object in a step, but it often did not provide feedback because the skill value was too high. When the system did provide feedback, it was helpful for one participant but not helpful for another participant.

Figures 26 and 27 show the complex syntax skill per chapter for participants who read Bottled Up Joy or The Lopez Family Mystery, respectively, using the EMBRACE ITS. The skill value shown is the final value by the end of the chapter.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
itsp01	0.150	0.150	0.654		
itsp04	0.150	0.150	0.654	0.792	0.792
itsp05	0.150	0.654	0.932	0.990	0.990
itsp08	0.150	0.150	0.504	0.891	0.891

Figure 26. Final complex syntax skill per chapter in Bottled Up Joy.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
itsp03						
itsp06	0.150	0.150	0.447	0.447	0.447	0.447
itsp07	0.150	0.150	0.150	0.571	0.751	0.764

Figure 27. Final complex syntax skill per chapter in The Lopez Family Mystery.

Most participants remained at 0.150 through chapter 2, which reflects that most participants did not reach the complex syntax level until chapter 3. Afterwards, all participants across both stories increased their complex syntax skill. However, the values are higher in Bottled Up Joy than The Lopez Family Mystery. Participant itsp05 reached the complex syntax level in chapter 2, mastered it in chapter 3, and maxed out the skill in chapters 4 and 5. Participant itsp06 reached the complex syntax level in chapter 3, but his or her skill level remained relatively low. It stayed the same for the rest of the story, as he or she returned to the medium syntax level in chapter 4. The log data indicates that this participant had not made any errors on complex sentences in chapter 3, but because there were so few complex sentences in the chapter compared to medium sentences, there were not enough opportunities to increase the complex syntax skill and stay at that level.

Usability Skill and Feedback

How does the usability skill compare between the EMBRACE condition and the EMBRACE ITS condition?

Table 5 shows the average usability skill for each participant in both conditions. The average was calculated from the final usability skill value at the end of each chapter.

Participant	**EMBRACE**	**EMBRACE ITS**	**AVG**	**STDEV**
itsp01	0.649	0.871	**0.760**	**0.157**
itsp03	0.827		**0.827**	
itsp04	0.850	0.777	**0.813**	**0.052**
itsp05	0.718	0.990	**0.854**	**0.192**
itsp06		0.858	**0.858**	
itsp07	0.784	0.852	**0.818**	**0.049**
itsp08	0.990	0.800	**0.895**	**0.134**
AVG	**0.803**	**0.858**	**0.830**	**0.039**
STDEV	**0.118**	**0.074**	**0.043**	**0.064**

Table 6. Average usability skill.

There does not appear to be a major difference between the two conditions. Whereas some participants increased their usability skill reading the EMBRACE ITS, other participants actually decreased. Overall, the usability skill is high for both conditions, despite the EMBRACE ITS providing a fair amount of usability error feedback. Because usability error feedback automatically performs the step for the user, there is no opportunity to correct usability errors immediately after receiving feedback.

Errors Per User Step

How many errors per user step did participants make in the EMBRACE condition compared to the EMBRACE ITS condition?

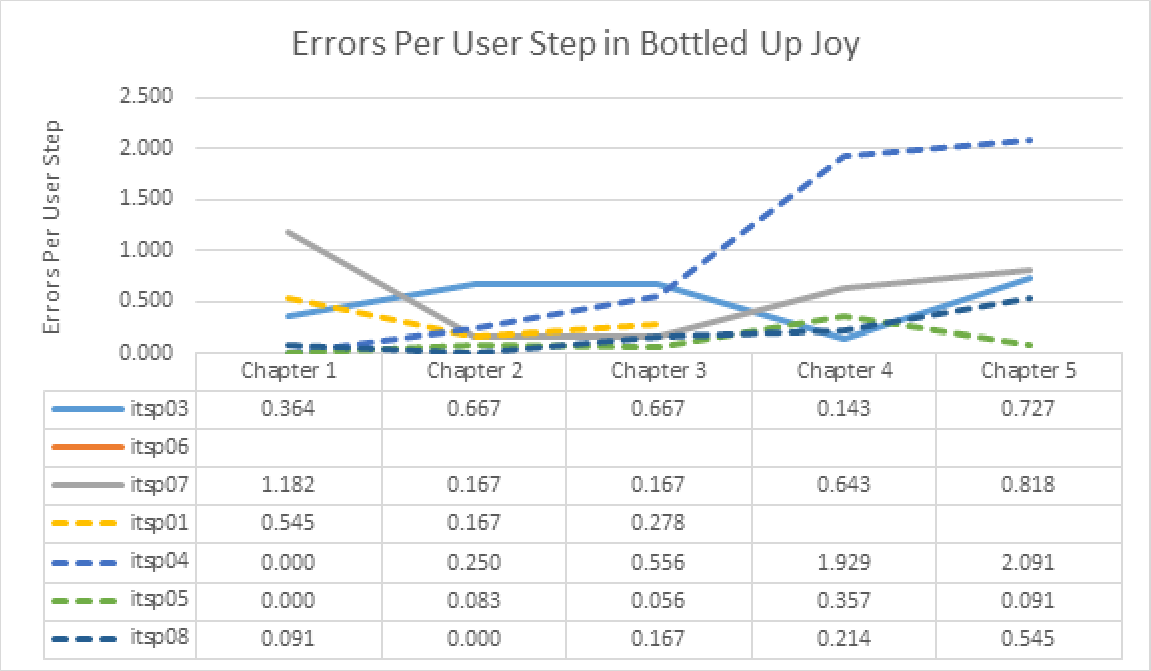
Table 6 shows the average errors per user step for each participant in both conditions. A user step is any step that is performed by the user--mostly “group” and “check” steps. Some steps are performed automatically by the system, so these are not included in the counts. The total number of errors in each condition was divided by the total number of user steps.

Participant	EMBRACE	EMBRACE ITS	AVG	STDEV
itsp01	0.790	0.330	**0.560**	**0.325**
itsp03	0.513	■	**0.513**	■
itsp04	0.603	0.965	**0.784**	**0.256**
itsp05	0.420	0.117	**0.269**	**0.214**
itsp06	■	0.483	**0.483**	■
itsp07	0.595	0.287	**0.441**	**0.218**
itsp08	0.332	0.203	**0.268**	**0.091**
AVG	**0.542**	**0.398**	**0.470**	**0.102**
STDEV	**0.160**	**0.304**	**0.178**	**0.085**

Table 7. Average errors per user step.

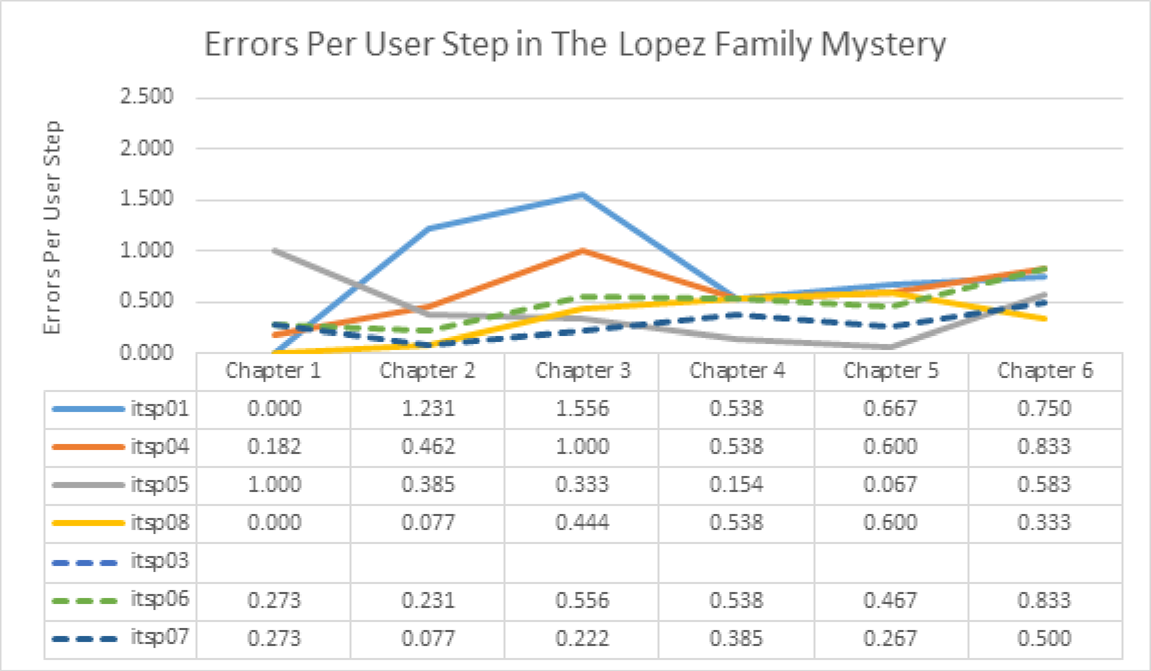
On average, ITS users have fewer errors per step than EMBRACE only users. Participant itsp04, however, increased in errors per step in the EMBRACE ITS condition. This makes sense as he or she received the most amount of error feedback.

Figures 28 and 29 provide more detail about the change in errors per user step throughout each chapter.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5
—— itsp03	0.364	0.667	0.667	0.143	0.727
—— itsp06					
—— itsp07	1.182	0.167	0.167	0.643	0.818
- - - itsp01	0.545	0.167	0.278		
- - - itsp04	0.000	0.250	0.556	1.929	2.091
- - - itsp05	0.000	0.083	0.056	0.357	0.091
- - - itsp08	0.091	0.000	0.167	0.214	0.545

Figure 28. Errors per user step in Bottled Up Joy. Solid lines represent EMBRACE participants, while dashed lines represent EMBRACE ITS participants.



	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
—— itsp01	0.000	1.231	1.556	0.538	0.667	0.750
—— itsp04	0.182	0.462	1.000	0.538	0.600	0.833
—— itsp05	1.000	0.385	0.333	0.154	0.067	0.583
—— itsp08	0.000	0.077	0.444	0.538	0.600	0.333
- - - itsp03						
- - - itsp06	0.273	0.231	0.556	0.538	0.467	0.833
- - - itsp07	0.273	0.077	0.222	0.385	0.267	0.500

Figure 29. Errors per user step in The Lopez Family Mystery. Solid lines represent EMBRACE participants, while dashed lines represent EMBRACE ITS participants.

For EMBRACE ITS users, the errors per user step appear to increase throughout the chapters, which coincides with the syntax complexity increasing in difficulty. Participant itsp06 switched down from complex syntax to medium syntax in chapter 3, and the errors per user step appear to decrease in chapters 4 and 5 but jump up in chapter 6. EMBRACE only users, on the other hand, experience a fluctuation of increases and decreases.

Errors Per Assessment Question

How many errors per assessment question did participants make in the EMBRACE condition compared to the EMBRACE ITS condition?

Table 7 shows the average errors per assessment question for each participant in both conditions. The total number of errors on assessment questions in the condition was divided by the number of assessment questions.

Participant	EMBRACE	EMBRACE ITS	AVG	STDEV
itsp01	0.241	0.179	**0.210**	**0.044**
itsp03	0.075	■	**0.075**	■
itsp04	0.256	0.389	**0.323**	**0.094**
itsp05	0.166	0.125	**0.145**	**0.029**
itsp06	■	0.196	**0.196**	■
itsp07	0.136	0.028	**0.082**	**0.076**
itsp08	0.024	0.207	**0.115**	**0.130**
AVG	**0.150**	**0.187**	**0.168**	**0.027**
STDEV	**0.091**	**0.119**	**0.087**	**0.040**

Table 8. Average errors per assessment question.

Due to participants itsp04 and itsp08, the average increased for the EMBRACE ITS group. Participant itsp04 may have performed worse because he or she made more errors per user step in the EMBRACE ITS condition than in the EMBRACE condition which demonstrates a lack of understanding of the text. Although participant itsp08 had a slightly lower average errors per user step using the EMBRACE ITS, he or she seemed to make more errors towards the end of the story and increase the number of errors per assessment question as a result. Other users decreased in errors per assessment question. Overall, the two averages appear similar across both conditions.

CHAPTER 8

EXIT INTERVIEW RESULTS

Exit interview questions and answers are provided below. When participants were asked about each type of adaptation and error feedback, they were shown a demo of the feature on the iPad.

Reading Habits and Using EMBRACE to Read

How often do you read? Where? With who?

Most participants said that they read a lot or everyday; one participant said they did not read that much. Most participants like to read at school; one participant also added that they like to read at home. Participants like to read with friends, siblings, or by themselves.

Do you like to read? Why or why not?

They all said they like to read; cited reasons were because reading is fun, is good for the brain, helps them learn new words, and improves reading skills.

Did you like reading with the iPad? What did you like about using the iPad to read these stories? What did you not like?

All the participants enjoyed reading using the iPad because of the ability to move the pictures. None of them said there was anything that they did not like about using the iPad.

Adapting Complexity of Vocabulary and Syntax

What did you think about the vocabulary list at the beginning of each chapter? Did you think the words were helpful? Explain.

All of the participants found the vocabulary lists helpful because they taught new words. One participant mentioned that the vocabulary lists were a clue for what words would be underlined in the chapter and what the chapter was going to be about.

What did you think about the number of words in the list?

All of the participants thought the number of words was “just right.” One participant liked the lists. Another found them easy to read. And another participant did not seem to think too much about them and just went through the process of tapping on each word.

How did you feel about the difficulty of the chapters? Did the difficulty change as you used the app?

One participant found that the chapters got easier, while another participant found the chapters hard and got harder over time. Another participant felt that the chapters were easy and stayed the same level through the story. The last participant thought that the chapters were a mix of both easy and hard, where some chapters got easier and others got harder.

Providing Feedback for Vocabulary, Syntax, and Usability Errors

If you moved the wrong pictures, sometimes the iPad would highlight other pictures on the screen. What did you think this meant? What did you do when the iPad highlighted these other pictures? Was this helpful or not? Why?

The participants found the vocabulary error feedback helpful because it told them where to move the pictures, which is exactly what they would do when they saw the feedback. One participant said they did not notice this type of feedback though.

If you moved the wrong pictures, sometimes the iPad would pop up a message box with one or two sentences in it. What did you think about the sentences in the message box compared to the sentence you were reading? What did you do when the iPad showed this message box? Was this helpful or not? Why?

Some participants found the syntax error feedback helpful. They read the sentence and moved the correct pictures afterwards. One participant did not see it often and did not read the message; the participant would simply press "OK." Another participant did not think the feedback was helpful and seemed to disregard the message completely.

If you kept getting an error, sometimes the iPad would pop up a message box saying "The iPad will show you how to complete this step." Then it will move the pictures for you. Was this helpful or not? Why?

All of the participants found the usability error feedback helpful for understanding where pictures are supposed to be moved. One participant noticed this feedback a lot. Another participant thought that the feedback showed the step being completed the same way he or she had already attempted.

Improvements to EMBRACE

What would you change about the iPad app? How can the iPad improve its ability to help you and other kids learn to read?

None of the participants had any answer for changes to improve the iPad. Most of the participants also did not have an answer for how the iPad can help kids read better. They seemed to misinterpret one of the questions as "How does the iPad help you learn to read?" One participant commented that the iPad is helpful in the way it sometimes reads a chapter aloud and other times asks the user to read it out loud by himself or herself. Another participant commented that the iPad helps when he or she is stuck on a word.

CHAPTER 9

DISCUSSION

The results from the study revealed both strengths and weaknesses of the EMBRACE ITS in regards to its ability to adjust relevant skills based on user actions, adapt complexity of vocabulary and syntax, and provide appropriate error feedback. Recommendations for future development on ITSs for reading comprehension in this context can be made according to these strengths and weaknesses.

Skill updates appeared the most accurate when users made particular kinds of errors. For example, if the user moved a completely irrelevant object that was far from the correct object, then the system was able to correctly identify the vocabulary error. If the user switched the subject and object of a step, then the system was able to correctly identify the syntax error. Users had a tendency to make this particular syntax mistake, so the system was successful in adjusting the relevant skill type in most syntax-related cases; however, the system failed to correctly identify possible syntax errors in other cases, such as when syntax skills were too high to provide feedback for syntax or when users moved a relevant object to an unnecessary, irrelevant location. Vocabulary and usability skills were often penalized instead. The system also demonstrated a bias towards increasing the medium syntax skill, as medium-level sentences were the most common in chapters, even when the complexity level had increased. Sometimes there were not enough complex sentences in a chapter to give the user enough opportunity to develop his or her complex syntax skill. Additionally, it was often observed that users who made enough of the right kinds of errors could actually end up increasing their vocabulary skill values more than users who did not make any errors at all. This behavior may be both a success and a failure of the system. On one hand, users who complete a step correctly the first time likely know the skill well, so it seems unfair that users who repeatedly complete only part of the step correctly can unintentionally make themselves seem as though they know the skill better. On the other hand, users should be credited for the skills they do know even if they do not know all of them. Adjusting the amount that skills values are dampened after the first attempt at a step is a possible solution. Ideally, students should only receive credit for the same skill on a step once.

When the EMBRACE ITS adapted the complexity of vocabulary and syntax, giving extra vocabulary did not seem as valuable as giving more difficult sentences. However, the system did appear to add a reasonable number of words to the vocabulary lists. Almost no users received the maximum number of words possible in any chapter. This is valuable, as too many words can overwhelm or bore users, but according to the answers from the exit interviews, most of the users found the lists helpful and that the amount of words was satisfactory. Nonetheless, adapting the vocabulary lists unfortunately made little impact on the vocabulary skills of EMBRACE ITS readers other than giving the skills a slight boost. The system often added easy words only because it was early in the story and the user had not been able to practice the words much. It also rarely added non-manipulatable words, which users may need the most practice with, because these words usually only appeared in one or two chapters and users hardly requested them. As for adaptations to syntax, many users seemed to perform well on medium sentences, so the EMBRACE ITS was successful in challenging them with more complex sentences. However, some of these medium sentences should probably have been encoded as simple or complex sentences instead. A few readers struggled with the sentences, yet the system still increased their complexity levels. Other readers might have been doing fine with a higher complexity, but as mentioned previously, since most sentences are encoded as medium-level, they had fewer opportunities to practice complex sentences, so the system ended up decreasing their complexity level instead.

The EMBRACE ITS error feedback features experienced its share of strengths and weaknesses as well. Although the system was likely to provide error feedback after only one mistake early in the story when skill values were still low, it did not seem to provide so much error feedback that it became disruptive to the reading. Vocabulary error feedback appeared to be the easiest to understand. Many users found this feature helpful, and in most cases, they were able to immediately correct their mistakes after receiving it. However, syntax error feedback was the most challenging. It did not seem to help users as they continued to make syntax errors even after receiving a simpler version of the current sentence. This might mean that the simpler version was not actually simpler at all. Also, based on the answers from the exit interview, many participants seemed to confuse the syntax error feedback for the usability error feedback. After seeing the message in the usability error feedback for the first time, they simply ignored the message in the syntax error feedback likely because the two types look so similar. Lastly, it was often observed that when users developed higher skills later in the story, they almost never received any error feedback. In some ways, this behavior may be considered successful for an ITS. It is reasonable to reduce the amount of help as student knowledge increases, but it may not be reasonable to remove that help completely. Often times high skill readers were only able to receive error feedback because they drove their skills low enough by repeatedly making an error that should have been considered correct. The system failed to provide usability error feedback quickly enough in these situations.

System bugs, such as failing to credit users for correct answers and failing to map the correct vocabulary word to encoded object IDs, also affected the overall functioning of the EMBRACE ITS. These should be fixed in the next iteration of the system.

Based on the strengths and weaknesses demonstrated by the results of the EMBRACE ITS study, my strongest recommendation for future development is to adapt the thresholds for providing error feedback as skills change. Users should be able to make at least one mistake before receiving help. This gives them the opportunity to try to fix the mistake on his or her own. Additionally, users with very high skills should still be able to receive error feedback after making multiple mistakes without driving their skills too low. This essentially means that the probability of making a usability error should actually increase as vocabulary and syntax skills increase because if the user has high knowledge of vocabulary and syntax, then repeated mistakes on the same step should be an indication of a system error rather than a user error. Another possible solution would be to implement usability errors as a flag to fix in the system rather than as a skill by itself. Another recommendation would be to consider the importance of how error feedback is presented. If two types of feedback look too similar, then the user may not notice the difference and error feedback will become less useful. Instruction on how each type of error feedback works would allow users to better understand what they can expect from using the ITS.

CHAPTER 10

CONCLUSION

The ill-defined nature of reading comprehension makes it difficult to design an ITS that effectively represents all the necessary skills for students to learn. Figuring out how to use these skills to track the student's current reading level while adapting the complexity of the text to read presents another challenge, along with figuring out how and when error feedback should be provided. I attempted to address these challenges by applying the design principles of ITSs to EMBRACE, which allowed the system to become both a learning environment and an assessment environment in which students can develop their reading comprehension skills.

EMBRACE applies the first principle of designing ITSs by decomposing reading comprehension into a series of steps to complete. From these steps, I was able to extract the student's reading comprehension skills. While there are many possible ways to categorize these skills, I chose to focus on vocabulary and syntax in general, and added a new category, usability, to account for possible system design issues. Vocabulary skills were further decomposed into specific words. This was helpful for seeing which words users struggled with the most without users having to explicitly provide that information to the system via help requests. On the other hand, syntax skills were further decomposed into simple, medium, and complex levels. This was not as helpful because some sentences were encoded with the incorrect level, so syntax skill updates were a bit inaccurate. Finally, adding usability as a separate skill was not necessarily a good idea because system errors exist whether a user demonstrates high skill in using the system or not. Thus, the usability skill might be more beneficial if it was simply implemented as a flag to fix in the system. Overall, this classification of skills was not a bad approach for an initial reading comprehension ITS, as the skills can be later refined into more specific types, such as pronouns for vocabulary skills, depending on the needs of the system.

Adapting the complexity of text so that users of different reading abilities would not all be treated the same was another challenge. In accordance with the fifth principle of designing ITSs, minimizing working memory load, I attempted to introduce new skills by increasing complexity only if the user demonstrated mastery of the current skills. I chose to repeat words that users seemed to struggle with the most, but simply viewing the words again was not as helpful as actually practicing them. Additionally, I chose to start all users at the medium syntax level and increase or decrease the complexity of the next text to read based on their syntax skills. I found that this is helpful if (1) the sentences are encoded with the proper complexity level, as mentioned previously, and (2) users are given enough opportunities to practice one complexity level before switching to another. The initial reading level may not be as relevant if these two factors are considered.

In regards to error feedback, I applied the sixth principle of designing ITSs by providing immediate feedback to let the user know whether an action was right or wrong. Additionally, I chose to provide more specific feedback based on the skills to help the user figure out how to correct his or her mistake. An ITS should provide the first type at a minimum, but it is not as helpful without the second type. I was able to apply the eighth principle of designing ITSs, facilitating successive approximations to the target skill, by giving less feedback over time as skills improve. Ideally, users should always be provided immediate error feedback to let them know whether an action was right or wrong, and then they should be allowed to make at least one or two additional mistakes before being provided specific error feedback. This gives them an opportunity to correct their mistakes on their own. It is helpful to give different specific error feedback depending on the skill. However, the feedback appearance must be distinct enough that users do not confuse it for another type. Instruction on how to use the feedback would also help set user expectations on how to use the ITS.

Limitations

Several limitations can be noted about the EMBRACE ITS and study results. The most significant limitation is the small number of participants, which resulted in a wide range of reading abilities. There was also a lot of incomplete data as some participants did not finish both stories or participate in the exit interviews. Additionally, the system faced some issues with correctly identifying errors and updating skills, and it reset the skills of a few participants after the application crashed. Syntax error feedback may not be accurate because sentences are not necessarily encoded with the appropriate complexity level. On another note, the system may also be limited due to the fact that most of the manipulation steps that students were required to perform followed a specific language pattern in which one object is moved to another object or destination. Restricting instruction to primarily teach this language pattern makes it difficult to assess the system’s effect on comprehension for other language patterns. Lastly, the story texts may have cultural implications as well. The students that participated in the study were native English speakers, whereas the students for which the stories were written are Spanish-speaking English Language Learners, so the participants may not have been as familiar with some of the words and phrases that were used.

Future Work

Future development on the EMBRACE ITS involves expanding skill updates to taps on menu options in both physical manipulation and imagine manipulation activities, pressing the “Next” button when not all steps have been completed, and results from the assessment questions. The system should also consider timing information when updating skills and providing error feedback. Another development is to adjust error feedback thresholds according to skill values. A major focus should be on improving syntax error feedback in particular, as it appears the weakest. For example, instead of defaulting to usability error feedback when no simpler version of the current sentence exists, the system might try reading the sentence aloud to the student. Another option is to present syntax error feedback in a way that is more distinct from usability error feedback so that the user is less likely to ignore it. Finally, a third option is to highlight the correct objects to manipulate in the correct order. A larger population will be included in the next study using the improved system.

While English Language Learner children enjoy reading using an interactive storybook application like EMBRACE, they have the opportunity to benefit more from an application that applies the design principles of intelligent tutoring systems. An intelligent tutoring system provides a clearer understanding of the vocabulary and syntactic knowledge that children develop at each step in the reading process. Understanding what skills are known or not known and to what extent is valuable for both adapting complexity and providing feedback. The system can dynamically present text that best matches the current reading level of the student while allowing them to practice reading more complex materials as well. Additionally, the system can help the student correct his or her misconceptions in reading as soon as they arise. These features of intelligent tutoring personalize the learning experience in ways that empower English Language Learner children to become better readers while still having fun.

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The journal of the learning sciences*, *4*(2), 167-207.

August, D. & Shanahan, T. (2006). Executive Summary. *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*, 1-9.

Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In *International Conference on Intelligent Tutoring Systems* (pp. 406-415). Springer Berlin Heidelberg.

Batalova, J. & McHugh, M. (2010). Top Languages Spoken by English Language Learners Nationally and by State. *Migration Policy Institute*, 1-5.

Beck, J. E., Chang, K. M., Mostow, J., & Corbett, A. (2008, June). Does help help? Introducing the Bayesian Evaluation and Assessment methodology. In *International Conference on Intelligent Tutoring Systems*(pp. 383-394). Springer Berlin Heidelberg.

Brown, J., & Eskenazi, M. (2004). Retrieval of authentic documents for reader-specific lexical practice. In *InSTIL/ICALL Symposium 2004*.

Duchateau, J., Kong, Y. O., Cleuren, L., Latacz, L., Roelens, J., Samir, A., ... & Verhelst, W. (2009). Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules. *Speech Communication*, *51*(10), 985-994.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424-436.

Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. *International Electronic Journal of Elementary Education*, 4, 5-18.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9, 558-565.

Glenberg, A., Willford, J., Gibson, B., Goldberg, A., & Zhu, X. (2012). Improving reading to improve math. *Scientific Studies of Reading*, *16*(4), 316-340.

Hagen, A., Pellom, B., & Cole, R. (2007). Highly accurate children’s speech recognition for interactive reading tutors using subword units. *Speech Communication*, *49*(12), 861-873.

Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi M. (2006). Classroom Success of an Intelligent Tutoring System for Lexical Practice and Reading Comprehension. *Proceedings of the Ninth International Conference on Spoken Language Processing*.

Heilman, M., & Eskenazi, M. (2006). Language learning: Challenges for intelligent tutoring systems. In *Proceedings of the workshop of intelligent tutoring systems for ill-defined tutoring systems. Eight international conference on intelligent tutoring systems* (pp. 20-28).

Heilman, M., Zhao, L., Pino, J., & Eskenazi, M. (2008, June). Retrieval of reading materials for vocabulary and reading practice. In *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 80-88). Association for Computational Linguistics.

Jacovina, E. J., & McNamara, D. S. (2016). Intelligent tutoring systems for literacy: Existing technologies and continuing challenges. In R. Atkinson (Ed.), *Intelligent tutoring systems: Structure, applications and challenges.* Hauppauge, NY: Nova Science Publishers Inc.

Lepper, M. R. & Chabay, R. W. (1985). Intrinsic Motivation and Instruction: Conflicting Views on the Role of Motivational Processes in Computer-Based Education. *Educational Psychologist*, 20, 217-230.

Levinstein, I. B., Boonthum, C., Pillarisetti, S. P., Bell, C., & McNamara, D. S. (2007). iSTART 2: Improvements for efficiency and effectiveness. *Behavior Research Methods*, 39, 224-232.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004) iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computer,* 36, 222-233.

McNamara, D.S., & the CSEP Lab. (2006). *iSTART: Benefits and Effects of Extended Practice.* Technical Report. University of Memphis.

McNamara, D.S., & the CSEP Lab. (2006). *iSTART: A classroom study. Technical Report.* University of Memphis.

Mokhtari, K. & Niederhauser, D. S. (2013). Vocabulary and Syntactic Knowledge Factors in 5th Grade Students' Reading Comprehension. *International Electronic Journal of Elementary Education*, 5, 156-170.

Mostow, J. (2012, June). Why and how our automated reading tutor listens. In

Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training (ISADEPT) (pp. 43-52).

Mostow, J., & Beck, J. (2006). When the rubber meets the road: Lessons from the in-school adventures of an automated Reading Tutor that listens. *Scale-Up in Education, 2*, 183-200.

Mostow, J., Burkhead, P., Corbett, A., Cuneo, A., Eitelman, S., Huang, C., ... & Tobin, B. (2003). Evaluation of an automated Reading Tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research*, *29*(1), 61-117.

Mostow, J., Nelson-Taylor, J., & Beck, J. E. (2013). Computer-Guided Oral Reading versus Independent Practice: Comparison of Sustained Silent Reading to an Automated Reading Tutor that Listens. *Journal of Educational Computing Research*, 49, 249-276.

Poulsen, R. (2004). Tutoring bilingual students with an automated reading tutor that listens: Results of a two-month pilot study. *Unpublished Masters Thesis, DePaul University, Chicago, IL*.

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-Speaking Children Reading in English: Toward a Model of Comprehension. *Journal of Educational Psychology*, 97, 246-256.

Reeder, K., Shapiro, J., Wakefield, J., & D'Silva, R. (2015). Speech Recognition Software Contributes to Reading Development for Young Learners of English. *International Journal of Computer-Assisted Language Learning and Teaching, 5*(3), 60-74.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46, 197-221.

VanLehn, K. (2006) The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16, 227-265.

Walker, E., Adams, A., Restrepo, M. A., Fialko, S., & Glenberg, A. M. (in press). When (and how) interacting with technology-enhanced storybooks helps disadvantaged readers. Manuscript submitted for publication.

Wessels, S. (2013). Integrating science and vocabulary instruction for English language learners. *Science as a Second Language*, 51, 50-53.