Student Modeling for English Language Learners in a Moved By Reading Intervention

by

Nicolette Furtado

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

Approved June 2016 by the
Graduate Supervisory Committee:

Erin Walker, Chair
Ihan Hsiao
M. Adelaida Restrepo

ARIZONA STATE UNIVERSITY

August 2016
ABSTRACT

EMBRACE (Enhanced Moved By Reading to Accelerate Comprehension in English) is an IPad application that uses the Moved By Reading strategy to help improve the reading comprehension skills of bilingual (Spanish speaking) English Language Learners (ELLs). In EMBRACE, students read the text of a story and then move images corresponding to the text that they read. According to the embodied cognition theory, this grounds reading comprehension in physical experiences and thus is more engaging.

In this thesis, I used the log data from 20 students in grades 2-5 to design a skill model for a student using EMBRACE. A skill model is the set of knowledge components that a student needs to master in order to comprehend the text in EMBRACE. A good skill model will improve understanding of the mistakes students make and thus aid in the design of useful feedback for the student. In this context, the skill model consists of vocabulary and syntax associated with the steps that students performed. I mapped each step in EMBRACE to one or more skills (vocabulary and syntax) from the model. After every step, the skill level is updated in the model. Thus, if a student answered the previous step incorrectly, the corresponding skills are decremented and if the student answered the previous question correctly, the corresponding skills are incremented, through the Bayesian Knowledge Tracing algorithm.

I then correlated the students’ predicted scores (computed from their skill levels) to their posttest scores. I evaluated the students’ predicted scores (computed from their skill levels) by comparing them to their posttest scores. The two sets of scores were not
highly correlated, but the results gave insights into potential improvements that could be made to the system with respect to user interaction, posttest scores and modeling algorithm.
DEDICATION

Dedicated to my mother, Nirmala Furtado, who does too much for me.
ACKNOWLEDGMENTS

None of this would be possible without the guidance and support of my advisor, Dr. Erin Walker. Thank you for giving me the opportunity to work with you, for being so approachable and for always answering all my questions, even the stupid ones. I am extremely grateful for her honest and prompt feedback, constant support, and encouragement. Her passion and dedication to her work is truly inspiring. Being mentored by her has been a wonderful experience that I will always be grateful to have had. I am also grateful to my thesis committee members, Dr. M. Adelaida Restrepo and Dr. Ihan Hsiao, for agreeing to be on my committee and for their valuable time and feedback.

I am indebted to my parents, especially my Mom. Their prayers, guidance and belief in me are the reason I could do this. Thank you for your constant support, for listening to me, even through some very long international phone calls and for making sure I was working.

I would also like to thank the developers working on the EMBRACE system, Audrey, James, Jonatan and Jithin explaining the details of EMBRACE to me and answering any question I have had about EMBRACE through the course of my thesis. Thanks also to the group of Masters and PhD students at the 2Sigma Lab, Audrey, Victor, Anil, Scott and Shang, for their useful feedback and suggestions on my work.

I am extremely grateful to have had the opportunity to work at the FACT lab. I am indebted to Dr. Kurt VanLehn and Dr. Salman Cheema for this opportunity and it has been a
privilege to have worked there. My time at FACT taught me a lot of things and I had the opportunity to work on a lot of interesting features.

I am lucky to have found a great bunch of friends and support system here at ASU. Nikhil and Tanushree, thank you for your friendship. All my friends at FACT, especially Jasdeep, Ankita, Pramodh, Adithya, Ayush, Nitin, Kenil thank you for being awesome co-workers and friends. I have had an enriching experience at FACT because of you guys. I am also thankful to Sachin Grover for giving me advice and feedback whenever I needed it. My college friends, Niyati, Nisha and Divya, thank you for being there for me, even with all the distance between us. Lisa, I cannot thank you enough for helping me through a tough time.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................................. ix

LIST OF FIGURES .............................................................................................................................. x

CHAPTER

1 INTRODUCTION ................................................................. .......................................................... 1

2 OVERVIEW OF THE THESIS ................................................................. ............................... 5

3 BACKGROUND .......................................................................................................................... 6

   Challenges in Second Language Learning ...................................................... 6

   The Role of Vocabulary ............................................................................. 6

   Intelligent Tutoring Systems ................................................................... 8

   Intelligent Tutoring Systems for Languages ......................................... 11

   REAder Specific Practice(REAP) ............................................................ 11

       Interactive Computer Identification and Correction of Language

   Errors(ICICLE) ................................................................................................. 13

   Interactive Strategy Training for Active Reading and Thinking (iSTART) .............................................. 14

   Reading Tutor .......................................................................................... 15

   The Military Language Tutor(MILT) ........................................................ 17

   Mastery Learning .................................................................................... 18

   Bayesian Knowledge Tracing ................................................................... 19

4 METHODOLOGY ................................................................. .................................................. 23

   Moved by Reading ....................................................................................... 23
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5  THE EMBRACE APPLICATION .................................................................. 25</td>
<td></td>
</tr>
<tr>
<td>6  CORPUS ................................................................................................. 28</td>
<td></td>
</tr>
<tr>
<td>Pre-processing of the Data ......................... 30</td>
<td></td>
</tr>
<tr>
<td>7  KNOWLEDGE TRACING ITERATIONS ..................................................... 31</td>
<td></td>
</tr>
<tr>
<td>Initial Models .............................................. 30</td>
<td></td>
</tr>
<tr>
<td>Considering Every Attempt of the Student ........... 31</td>
<td></td>
</tr>
<tr>
<td>Considering the Initial Attempt Multiple Times ...... 32</td>
<td></td>
</tr>
<tr>
<td>Basic Model ..................................................... 34</td>
<td></td>
</tr>
<tr>
<td>Considering the Initial Attempt Multiple times along with Help Requests and Syntax as a Skill ..................................................... 34</td>
<td></td>
</tr>
<tr>
<td>Enhanced Model ............................................... 35</td>
<td></td>
</tr>
<tr>
<td>Considering the Previous Model along with a More Detailed Syntax Decomposition ................................................................. 35</td>
<td></td>
</tr>
<tr>
<td>8  RESULTS AND EVALUATION ................................................................. 37</td>
<td></td>
</tr>
<tr>
<td>Help Requests for Each Word ......................... 39</td>
<td></td>
</tr>
<tr>
<td>Attempts at Each Skill ....................................... 40</td>
<td></td>
</tr>
<tr>
<td>Skills per Chapter ....................................... 41</td>
<td></td>
</tr>
<tr>
<td>Student Attempts per Chapter ......................... 42</td>
<td></td>
</tr>
<tr>
<td>Mean Skill Values ............................................. 43</td>
<td></td>
</tr>
<tr>
<td>Correlation Results .......................................... 44</td>
<td></td>
</tr>
<tr>
<td>9  DISCUSSION AND CONTRIBUTION ....................................................... 50</td>
<td></td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>53</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table Page
1. Skills per Chapter per Condition for Correlation .......................................................... 37
2. Skills per Condition for Correlation .............................................................................. 38
3. Action and Non Action Verb Posttest Questions per Chapter ...................................... 39
4. Help Requests Across Students for Each Word ............................................................. 40
5. Number of Attempts for Each Word ............................................................................. 41
6. Number of Skills per Chapter ........................................................................................ 42
7. Student attempts per chapter ........................................................................................ 43
8. Mean Skill Values for the Basic Model ........................................................................ 43
9. Mean Skill Values for the Enhanced Model .................................................................. 44
10. Correlation Results for the Basic Model in the Action Condition .............................. 45
11. Correlation Results for the Enhanced Model in the Action Condition ...................... 45
12. Correlation Results for the Basic Model in the Non Action Condition ....................... 46
13. Correlation Results for the Enhanced Model in the Non Action Condition ................ 47
14. Correlation Results for the Basic Model in the Complete Condition ......................... 47
15. Correlation Results for the Enhanced Model in the Complete Condition ................. 48
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Conditional Probability after Correct Response</td>
<td>21</td>
</tr>
<tr>
<td>2.</td>
<td>Conditional Probability after Incorrect Response</td>
<td>22</td>
</tr>
<tr>
<td>3.</td>
<td>Probability That the Student Learned the Skill after Correct/incorrect Response</td>
<td>22</td>
</tr>
<tr>
<td>4.</td>
<td>A Screenshot of the Embrace Application after a Move Action Has Been Performed</td>
<td>27</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

English Language Learner (ELL) students in the U.S. hail from over 400 different cultures. It is estimated that, at present, one in nine students is an ELL. However, most ELLs, around 80 percent, come from a Spanish speaking background [1]. Nationally, it has also been found that there is a significant disparity between the reading comprehension proficiencies of language-minority and English-only children [2]. This is especially concerning since most ELLs are in an English-only medium of instruction school and their academic achievements are low [1].

ELLs who experience slow vocabulary development have lower text comprehension abilities [3]. This can severely affect their academic progress in English, which is not only the dominant medium of instruction in schools, but also the medium through which they are tested [2]. This is unfortunate since their poor performance in school impacts their future academic and career opportunities. Society on the whole, also loses out on the potential achievements of these children while bearing the costs of such a substantial underachievement. However, improved reading skills in English Language Learners improve their overall English skills, which will have an overall positive effect on their learning [2].

Improving the reading skills of these children is a necessary step that will help them academically. Teaching reading strategies to struggling readers in order to improve reading comprehension has been found to be an important intervention to improve
reading and reading comprehension. Reading strategies are actions that students actively use to better their language skills [2]. Although time consuming at first, these strategies become natural to the user after sufficient practice. The goal of employing such strategies is to make the use of these strategies a natural part of reading in the future. After all, a successful reader actively employs a number of interdependent strategies to adequately comprehend a text [4].

Enhanced Moved by Reading (EMBRACE), is an Ipad application with a game-like environment to improve reading comprehension among ELLs. It uses animated stories that the users can interact with in order to teach reading comprehension to students with Spanish as their first language.

Researchers have debated whether a second language must be taught directly, through interactions with other speakers or through looking up words and their meanings in a dictionary. It has, however been established that direct instruction is more effective than the dictionary approach.[1] Also, integrating language learning in the day along with regular course work has not been found to be as effective as a separate instructional time devoted solely to second language learning [1]. In accordance with this, EMBRACE trials are conducted as separate sessions from regular classes. This separate block of instruction time has been correlated with higher scores on standardized tests [1].

Reading, being a difficult skill to learn and acquire for some people, could be harder for children reading in a second language. Moreover, reading comprehension requires the
reader to read between the lines [4]. ESL readers may understand each word in a text separately, but may not have developed adequate language skills to understand the underlying meaning of the text they read. It has been established that effectively teaching a second language requires a combination of

- Explicitly teaching students the features of the second language i.e. syntax, vocabulary, grammar, pronunciation.
- Opportunities to use their knowledge in the second language in order to further their learning [1].

It is known that for adults, a large vocabulary of around 8000-9000 word families for reading and 5000-7000 word families for conversation are required to function in English. This is a huge overhead that English language learners need to cover [5]. A useful method for vocabulary learning is to boost the learners’ engagement with the words, which is an approach the EMBRACE application already uses.

Most vocabulary learning applications have target vocabulary that the student is expected to learn through the use of the application. It is believed that a student needs to know the meaning of 98-99% of words in a text in order to adequately comprehend the text. It has also been found that a sufficient level of comprehension can be gained from knowing a lower percentage of words in a passage. However, a knowledge of less than 80% of words in a text is almost always related to low text comprehension skills [5].
Through this thesis, I have attempted to analyze the role of vocabulary in reading comprehension skills of ESLs while using EMBRACE. Although, vocabulary skills are a very important part of acquiring second language skills, there may be other language components that appear in a reading comprehension text e.g. syntax, that affect language acquisition in spite of vocabulary skills. I have attempted to track the path of the student as they use the application and relate their skill levels to their posttest scores at the chapter level. I used correlation to observe the predictability of their skill levels (as per their first response) and posttest scores.
CHAPTER 2

OVERVIEW OF THE THESIS

The EMBRACE application uses the Moved By Reading technique to improve the English skills of bilingual students with Spanish as their first language. This thesis uses the student log data from previous studies using EMBRACE to categorize and analyze student errors as they relate to vocabulary, syntax and how they relate to their posttest results. The student log data contains details about each student’s usage of the application. I experimented with basic to more complex variants of skill models for students using EMBRACE. I associated every step in EMBRACE with skills present in the skill model. I then used Bayesian Knowledge Tracing to predict the skill levels of students in all the models. I compared their skill levels at the end of each chapter with their performance in the posttests for each chapter.
CHAPTER 3

BACKGROUND

Challenges In Second Language Learning

There has been a great amount of research on effective text comprehension among native English speakers [6]. It has been found that just as for native English speakers, oral language proficiency is positively correlated with written language proficiency. Knowledge of vocabulary is an important skill, as explained in the next section [6]. Decoding is another important skill that is important in language learning irrespective of language. A language learner must be able to recognize a word fast enough to decode its meaning [6].

The role of vocabulary

There is a very pronounced difference in the vocabulary knowledge in English of native English speakers and bilingual people. This difference gives bilingual speakers less vocabulary to work with and thus they find it difficult to communicate in and comprehend English [6]. It also affects higher levels of understanding language such as grammatical processing, construction of schemata and text models [3].

Vocabulary has always been considered to be an important part of second language learning. Knowing a word implies many less obvious aspects. It implies that the learner is aware of the literal or most common meaning of the word, the contexts in which it can be used, how it can be used in various types of sentences, as well as its synonyms and antonyms, if they exist. The knowledge of these features of a word suggests that the
learner has acquired, not only a certain breadth of word knowledge, but also a depth of word knowledge which has led him/her to have such complete knowledge of the word in question. Current research suggests that ELLs have a lower breadth and depth of word knowledge as compared to native English speakers [3].

A challenge for instructors of a second language course is the selection of effective vocabulary. It is important that the vocabulary chosen for students is not only age appropriate but also unique and unlikely to be learned by students in other settings. This is to ensure that the students gain the most out of the second language instruction provided to them.

It is also important the words chosen appear often enough for the students to become familiar with the words and learn them fully. The probability of learning an unknown word that a second language reader comes across in a text is low, even more so for younger learners and for learners who are new to the language. Since ELLs lack the higher decoding abilities of native English speakers, it is unrealistic to expect them to accurately decode the meanings of words they chance upon. A strategy through which they are directly taught the meanings of unfamiliar words as well as are exposed to these words frequently enough, is a pragmatic approach to vocabulary learning for ELLs [7].

An emerging approach for imparting second language skill is through the use of Intelligent Tutoring Systems, which are explained in the next chapter. Intelligent tutoring
systems may use various techniques to diagnose the cause for specific student errors and help remediate them.

**Intelligent Tutoring Systems**

Intelligent tutoring systems provide interactive learning environments for students to learn and explore a particular domain. The diversity of intelligent tutoring systems that have been developed use different methods to provide feedback and thus support learning. Some systems allow students to experiment with a domain but do not track their errors. Other systems offer more feedback but do not provide enough direction to learners.

The past few years of research in intelligent tutoring systems has seen a lot of analysis and improvement in the student model section of intelligent tutoring systems.

Traditionally, the intelligent tutoring system architecture consists of the following components [8]:

*Problem solving environment.* The problem solving environment contains, at the very least, an interface for the student to interact with the system. The interface may be as minimal as a text box for the student to enter answers or question for the system to evaluate. This is particularly important in instruction domains where the real world environment may pose a threat to the student. It is also the medium through which the system can provide feedback to the student.

Designing an effective problem solving environment must consider the following:
• The problem solving environment must be as close to the real world as possible.
• It must also be constructive for the learning process of the student. This guideline gives rise to the need for feedback while learning in order to improve learning [8].

*Domain Knowledge.* The domain model contains the knowledge that the student is expected to gain from the system. It is the foundation of the intelligent tutoring system. According to changes in the domain model across systems, the interpretation of student actions may change [8].

An important consideration in the design of the domain module is the representation of the knowledge that needs to be taught to the student. It is also the source through which student actions can be evaluated [9]. Thus, the domain model must contain the logic required to interpret and evaluate student actions. An ideal domain model will solve a problem in the same way as a human.

*Pedagogical Module.* The pedagogical module is in charge of presenting items of learning to the user [8]. It is responsible for what will come next in the student’s journey to mastery of the subject. Thus, the implementation of the pedagogical module may be as personalized as the designers of the intelligent tutoring system want it to be. The following are the five types of actions that the pedagogical module may perform while tutoring the student:

• The system may teach the student by walking them through an example.
• The system may present to the student the series of steps that they need to carry out in order to solve the problem correctly.
• The system helps out while the student is solving the problem.
The system keeps a track of the topics and the student’s performance in each of them.

The system allows the user to explore the problem solving environment.

In addition, the system may not do anything at all while the student is solving a problem [8].

**Student Model.** The student model is a record of the student’s state as he/she uses the application. It usually consists of two components:

- The record of the activity of the student throughout the application and their performance over all the problems solved.
- The list of misconceptions or gaps in knowledge that the student may have acquired while using the application [8].

An important consideration here is the granularity of the knowledge being tested at each step. In order to get an accurate estimate of the student’s knowledge at any point of time, activities must be accurately mapped to a topic that must be learned by the student.

Student modeling is a very important part of Intelligent Tutoring Systems since it can be used to understand how learners learn, to make an educated estimate of the knowledge they have gained, along with any misconceptions they may have acquired on the way. Determining the student's knowledge state can be done through knowledge tracing, which is explained in the next section. This knowledge can be used to predict the future behavior of the student in order to better assist them as they use the application. Another use of the student model is to test the parameters of the application, in order to tweak them if necessary or to test how effective they are in predicting student behavior [10].
Intelligent Tutoring Systems For Languages

In recent times, a number of intelligent tutoring systems have been implemented for different types of students with varied strategies to improve reading comprehension. Some of the intelligent tutoring systems with English as their language are described below:

**REAder specific Practice (REAP):** The REAP system, as the name suggests, was developed with the intention of providing reader-specific texts in order to improve reading comprehension skills [16]. REAP was used to improve reading comprehension in students enrolled in University of Pittsburgh’s English Language Institute (ELI) Reading Level 4 course, an upper level course for English as a Second Language (ESL). It procured documents from the Web for reading by the students. The documents it selected had to satisfy certain criteria regarding their appropriateness for the students [17]. These criteria included the following:

- Complexity of the text: This was calculated using a mixture of language models.
- Grammatical structure: This included sentence level as well as word level features.
- Document structure: This included document level features such as the title, metadata and keywords.
- Topic category: This included the genre of the passages selected.
- Named entity tags: BBN’s Identifinder was used for tagging.
- Text coherence
Every document in REAP is annotated with these features [18]. It was found that less than one percent of the articles containing the required words were suitable for reading by the students. However, these features enabled a certain amount of individualization in the reading practice provided to students.

Before the students started reading any passage on REAP, they were presented with a pre-test for the application to learn what the student must learn. After this, the application selects a passage for the student to read. While the students read the passage, the target words were highlighted as and when they appeared in the text. If the students did not understand the meaning of the target words, they could learn their meanings from the electronic version of the Cambridge Advanced Learner’s Dictionary that was integrated into REAP. After each reading exercise, the students also had to complete exercises to assess their reading comprehension. The results of these exercises were also used to select further reading passages for the student [17]. The exercises were dynamically generated based on target vocabulary found in the reading.

The student model in REAP was quite simple, including the reader’s interests, reading level, and vocabulary acquisition. The model is updated dynamically as the students use the application [16].

The students who used REAP during the classroom studies conducted found it to be user friendly and useful in improving reading skills. But, they also wished for better personalization for better engagement [18]. Thus, although well received, the system
faced certain problems with the low text quality and the contexts of the articles that the system sometimes selected. Also, certain articles were found to contain slang from other countries and the students reading the articles were unfamiliar with their meanings. Most students were American and hence slang from Australia or Britain was unfamiliar to them [17].

**Interactive Computer Identification and Correction of Language Errors**

(ICICLE): ICICLE was developed to support English literacy skills, mainly writing, among deaf students who are more comfortable with American Sign Language (ASL) than with English. It has been observed that the reading comprehension skills of hearing impaired students is significantly lower than those of their peers. The developers of this system also observed that the errors these students made in writing were different from those made by students who could hear.

It is argued that although ASL and English are assumed to be identical, there are distinct differences in sentence construction in both. As a result, they treated ASL as a different language altogether and their approach to the system draws from second language learning theories [19].

ICICLE, in its current form, is a Windows-based application. It contains a text parser that goes through user input, checking it against some pre-defined common user errors. The system’s interpretation of a user’s text in case of many or conflicting issues, depends on the learner’s knowledge at that point of time [20].
It presented students with a text in English in which the students had to find grammatical errors. While doing this, it engaged the student in tutorial dialogue to help the student come up with corrections to the text. The system tracked student progress after every task and generated feedback which it presented to the user. Since the developers of this system also tried to adapt this system to the skill level and learning features of the users. Thus, this feedback was tailored to the users [19]. The feedback could be in the form of a tutorial session on the concepts that user needs to improve on [20].

**Interactive Strategy Training for Active Reading and Thinking (iSTART):** iSTART is a web based intelligent tutoring system that teaches a reading strategy called SERT, to adolescent students. SERT is based on the idea that students who can self-explain text are likely to have understood the text to a larger extent. SERT is made up of comprehension monitoring, paraphrasing, making bridging inferences, elaboration, prediction and bridging techniques. Comprehension monitoring allows children to identify their inability to understand a piece of text. Paraphrasing teaches the students to connect the material in the text to ideas they already know or to other ideas in the text. Making bridging inferences requires the user to associate the recently covered text to previously covered text in the application. The purpose of this step is to encourage a global understanding of the text as a whole in the student. Students then employ prediction to judge what the future text may involve. Ultimately, elaboration lets students connect their own knowledge with the current sentence. iSTART incorporates SERT by administering it via animated pedagogical agents [21].
In the Introduction Module, the student is introduced to three animated characters that cooperate with each other in order to teach reading comprehension to the student. The characters speak using a text to speech synthesizer and possess a variety of gestures. These characters simulate the working of SERT to the student. After the presentation of each strategy in SERT, the students need to take a quiz that assesses their understanding of SERT.

In the second phase called the Demonstration Module, the student learns to self-explain a science text. Again, two animated characters help the student in this process. The student is expected to learn from examples provided by the animated characters in this phase.

The third phase, called the practice phase, the student attempts to self-explain a text. Another animated character, helps the student do this via feedback. Once the student completes this, the character asks the student to identify which strategy from SERT was used. In this phase, the interactions of the user with the agent are monitored [22].

**Reading Tutor:** Oral reading is an important skill that is developed through a mix of classroom instruction and individual practice. Listening to children reading aloud is important, in order to assess their reading skills. It allows the listener to detect word comprehension errors so that they can be minimized. Correcting word reading errors has been found to improve word recognition accuracy and comprehension [23].
Project LISTEN’s Reading Tutor is a computer tutor that assesses the reading abilities of students. They used an automatic speech recognition technology to listen to children reading aloud. Humans, while judging oral reading proficiency, can judge a student on a number of factors by just listening to him/her read. The challenge for this tutor was to incorporate these scoring features into a computer based tutor [24]. The purpose was not just to detect mistakes but also act as motivation for the student as the tutor also identifies what the student knows [23].

The Reading Tutor evaluates the student’s reading ability, one sentence at a time. An animated agent listens to the student as he/she reads. When the system detects the end of the sentence, it checks the recognized output against the original sentence, in order to score the student’s performance. If no important words were missed, it displays the next sentence to the student to read. Otherwise, the animated agent provides help if needed, in the following ways: by handing out the word in question, speaking the text that appeared before the word in question or reading the sentence if more than one word was missed [23].

One of the features that was used for scoring reading ability was inter-word latency. Inter-word latency is the time taken between reading successive words and is defined only for words that are read correctly and if the previous word was read by the student. Latency also involves how long the student took to pronounce the whole word. The goal for calculating latency is to estimate the time taken by a student to pronounce a word [24]. Another feature that was assessed was fluency.
The analysis on the log data first took into account only the help requests by the students to observe whether students could be assessed only by their help requests. This model was compared to the combined model of speech recognition data and help requests. They found significant improvement between 6% and 10% variance. They also found that using speech recognition data does in fact, account for at least 20% more variance [24].

The results of the assessments from the Reading Tutor were used to generate automated continuous assessments of the students’ reading skills. These results were used to better fit the level of the stories chosen to the reading levels of the students and the help given to them [25].

**The Military Language Tutor (MILT):** The Military Language tutor is a natural language authoring and tutoring system developed by the U.S Army Research Institute (ARI), for military personnel to develop and sustain foreign language skills in a post-cold-war world. It has been observed that foreign language skills are usually short lived. Soldiers have been found to lose their foreign language abilities after their language schooling is over. MILT focused on teaching and sustaining language skills in Spanish and Arabic. The goal of the MILT system was to produce ‘overlearning’. Overlearning involves learning a skill beyond the point of mastery. This is useful when the skills learnt will have to be recalled or used in stressful situations [26].

MILT’s two dimensional microworld utilized speech recognition as well as user input through the keyboard. It also had an animated agent that interacted with the student as they used the application. The student’s speech was recognized through the Dragon
speech recognition software [26]. Its natural language capabilities were advanced enough to detect common student misconceptions and mistakes [27].

Students were given exercises that used language production to manipulate objects in a graphics microworld [26]. Experiments with students using MILT found that the students enjoyed their learning experience. MILT aimed to improve vocabulary, grammar, pronunciation and fluency in Spanish and Arabic. There were learning gains for students of both languages, especially Arabic [27].

The above systems have used a number of different techniques to teach English or reading comprehension in English. They use different methods to assess and store student state. In the next section, Bayesian Knowledge Tracing, which is a method to trace the knowledge acquired by the student through the course of the application is explained.

The above systems have used a number of different techniques to teach English or reading comprehension in English.

**Mastery learning:**

Mastery learning is a technique that aims for students to master what they have been taught. This philosophy has given rise to two strategies in instruction. The Learning for Mastery (LFM) strategy emerged from the field of education, while the other, Personalized System of Instruction (PSI) evolved from the field of psychology [11]. Both approaches, however, share some common characteristics:
They propose that students can learn for mastery in order to acquire some basic intellectual competencies.

These competencies will, in turn, let them advance in their learning.

Also, this process will give them positive feelings towards learning, in general [11].

Learning for Mastery (LFM): The Learning for Mastery approach was designed for use in a classroom setting where instructional time is relatively flexible. According to this method, if each student spent enough time required for him/her to reach some pre-defined level, then he/she would probably reach that level [11].

Personalized System of Instruction (PSI): The Personalized System of Learning approach was designed to be an individual-based and student-paced approach to mastery learning. It requires that the time available for instruction to students be relatively unrestricted [11].

Bayesian Knowledge Tracing:

Bayesian Knowledge Tracing is a method to track the acquisition of skills in mastery learning. The knowledge state of the student, as determined by knowledge tracing determines how close the student is to mastery of the skill. Knowledge tracing tracks the acquisition of skills across practice opportunities presented to the student over a period of time [12]. Each practice opportunity presents only two outcomes, right or wrong, and this outcome is used by the algorithm to update the knowledge state of the student. A student
can only learn a skill, unlearning or forgetting a skill is not considered. However, if a rule is said to have been learnt by the student, the student may still slip and make a mistake. Conversely, if a student has not learnt a skill, they may still guess the answer to a step. These possibilities are accounted for through the guess and slip probabilities in the algorithm [13].

Thus, the traditional knowledge tracing algorithm takes into account the following parameters to calculate the probability of the student having learnt the skill after each step:

1. Transition P (T): The probability that a student has learnt a skill after a certain step.
2. Guess P (G): The probability that a student got a step correct in spite of not having learnt the skill.
3. Slip P(S): The probability that a student got a step wrong in spite of having learnt the skill.
4. The probability that a student got a step wrong because they have not learnt the skill.
5. The probability that a student got a step right because they have learnt the skill.

The transition, guess, and slip parameters take values between 0 and 1 in the basic knowledge tracing model. Most intelligent tutoring systems commonly fix the values of the guess parameter between 0 and 0.3 and the slip parameter between 0 and 0.1 [14].
Thus, Bayesian Knowledge Tracing, is essentially a model that determines, when, over time, the learning of a particular skill occurred. Every step presents to the student an opportunity to succeed (and learn) or fail at a task [12].

Bayesian knowledge tracing has two stages: the learning stage and the inference stage. The learning stage involves learning all the four parameters of the algorithm. The inference stage is the probability of the student having learned the skill at a point in time [15].

Thus, during the inference stage, the probability of a student’s knowledge at time \( n \), when the student gave a correct response is as follows:

\[
p(L|\text{correct}_{n-1}) = \frac{p(L_{n-1}) \cdot p(\neg S)}{p(L_{n-1}) \cdot p(\neg S) + p(\neg L_{n-1}) \cdot p(G)}
\]

Fig: Conditional probability after correct response

\[
p(L|\text{incorrect}_{n-1}) = \frac{p(L_{n-1}) \cdot p(S)}{p(L_{n-1}) \cdot p(S) + p(\neg L_{n-1}) \cdot p(\neg G)}
\]

Fig 1: Conditional probability after incorrect response

Conversely, the probability of a student’s knowledge at time \( n \), when the student gave an incorrect response is as follows:
Fig 2: Conditional probability after incorrect response

After this learning opportunity, the probability that the student learned the skill is calculated as follows:

\[ p(L_n | \text{Incorrect}_n) = \frac{p(L_n) \cdot p(S)}{(p(\neg L_n) \cdot p(S)) + p(\neg L_n) \cdot p(\neg G))} \]

Fig 3: Probability that the student learned the skill after correct/incorrect response

\[ p(L_n) = p(L) + p(L) \cdot p(T') \]
CHAPTER 4

METHODOLOGY

Moved By Reading

From a psychological perspective, learning is more effective when grounded in physical experiences. It has been observed that some children may not enjoy reading in a language even if they may employ the same language to speak or watch movies in. Psychologists from the embodied cognition school of thought believe that this may be due to a lack of physical experiences while reading as opposed to speaking (e.g. waving hands). Thus, reading may become a boring activity, and this is exacerbated in case of reading in a second language which may be a difficult activity in itself [28].

Moved by Reading is an activity which requires students to manipulate the objects that appear in sentences they read [28]. The EMBRACE application aims to provide a connection between the written word and their meanings through the Moved by Reading strategy. The student needs to ‘tap’ on the correct object to manipulate and move the object on the screen to its correct position as per the sentence [29]. This associates reading comprehension with certain physical actions. The meaning of the sentence is thus ‘embodied’ in their actions.

Research on memory has proved that students who mime phrases they read retain more information than students who try to memorize without miming [30]. Moving objects on the tablet allows for active processing of words and their meanings within the context.
Active processing of words has also been found to lead to better recall and understanding of words and their meanings [3].
CHAPTER 5
THE EMBRACE APPLICATION

The EMBRACE application is an IPad app with movable animations. Every story in the app is organized into chapters. Each chapter is spread across a number of pages. At the beginning of every story, students are presented with a vocabulary list for the story. The words in the vocabulary list appear at least once throughout the course of the story. The student is expected to have learned the meanings of the words in the vocabulary list by the end of the story. Knowledge of vocabulary is an important aspect in learning a second language. Individual words are found to be more understandable when they are included in sentences that are understandable to the reader. Also, unfamiliar words are more likely to be understood if they are embedded within an understandable story [31].
Every page, as shown in the figure above, contains a box on the left side of the screen containing some of the sentences in that chapter. The sentences that have manipulations associated with them are displayed in blue, while the sentences without any manipulations are displayed in black. Every occurrence of any of the words in the vocabulary list is underlined. The student can tap on the underlined word to listen to the pronunciation and meaning of the word as many times as he/she wants. It has been found that the listening comprehension skills of second language learners positively influence their reading comprehension skills [6].

The sentences with manipulations associated with them require the student to move an object to another object based on the text in the sentence. In some cases, after a student moves an object to another object, the application asks the student to choose between two scenarios. The ‘move’ action is associated with a verb, referred to as an ‘action verb’ in the context of the EMBRACE application. The ‘move’ action helps students associate action verbs with the objects that interact with each other and thus understand the meaning of the sentence. The figure below is a screenshot of the EMBRACE application after the student moves the farmer to the cat. The first scenario shows the farmer near the cat, while the second scenario shows the cat standing on the farmer’s head. The student needs to choose between one of the two scenarios in order to move to the next step.
Fig 4: A screenshot of the EMBRACE application after a move action has been performed.

Further, it has been found that media in its various forms such as illustrations, audio and video enhance the learning experience for children and adults alike. It helps the child ‘picture walk’ through the story as they read the text on each screen [32].
CHAPTER 6
CORPUS

The corpus contains data from a study that tested 70 children studying in grades 2 to 5. These children were enrolled in extracurricular activities among two schools, three community centers, and one library. All the children were Spanish speaking with the ability to decode in English. As a pretest, the students were provided with a Qualitative Reading Inventory (QRI) consisting of 40 words. An experimenter first read aloud the words in the list. Then the child was given the list to read aloud on their own. Children who read at least half of the words correctly were participants in the study. The children worked on the Best Farm story for the first four days of the study.

The Best Farm is a story in the EMBRACE application with seven chapters and 838 words. The Flesch-Kincaid level for the readability of the story is 2.6. The story follows the central character, farmer Manuel as he, along with other farm animals like the cow, horse, and pig compete to win the Best Farm award.

The dataset I used consists of the logs of 20 students. I included only those students who had complete data for all chapters of ‘The Best Farm’ and did not repeat any of the steps in any chapter. The logs include details of every action of the participant while using the application. These details are organized in the following categories:

- Actor: The actor may be the user or the computer.
- Action ID: The action ID of the action performed
- Selection: The object that the user selects in order to perform move actions.
- **Action:** This column contains the type of action that was performed, including Move to Hotspot, Move to Object, Tap, Display Menu, Load Next Sentence, Load Next Page, and Play Word among others.

- **Input type:** This column provides information about the objects that are moved and the start and end positions of those objects.

- **Verification:** Based on the objects that are moved for a particular step in a particular sentence, the action is assigned a correct or incorrect verification.

- **Error type:** This contains one of the types of actions that was responsible for the incorrect action for a particular step in a particular sentence.

- **Status:** The presence of the entry in the log data indicates that the use completed this step. Hence, the value here is always completed.

- **School:** The name of the school the participant attends.

- **Condition:** The condition may be one of the following,
  - Bilingual Control (BC)
  - Bilingual EMBRACE (BE)
  - English-only Control (EC)
  - English-only EMBRACE (EE)

- **Day:** The day the action was performed.

- **Participant ID:** The ID assigned to the student.

- **Experimenter:** The experimenter in charge of conducting the experiment.

- **Story:** The story that the student is currently working on.

The posttest questions on each chapter consist of six to seven comprehension questions based on the text in that chapter. The questions that have their answers in a sentence
containing a manipulation by the student are action verb questions. The questions that have their answers in a sentence with no manipulation are non-action verb questions. Some of the questions were inference based while others had an answer that was directly present in the text. The children were asked a follow-up question, after every question that presented them with two options, one of which was the answer. If the child answered both the first and the follow-up questions, they were given two points, for answering just the follow-up question they were given one point and zero points if they were unable to answer both questions.

**Pre-Processing of the Data**

I used the actor, action, verification, input data and user step fields of the log data for my analysis. The actor could be the user or the computer, depending on who executes the user request. The actions that I considered were ‘Move to Object’, ‘Move to Hotspot’ and ‘Play Word’. The ‘Move to Hotspot’ and ‘Move to Object’ actions are move actions performed by the student, therefore they are classified as ‘User’ actions while the ‘Play Word’ action represents the help request by the student. Since the help request is essentially an audio of the word and its meaning, which is played by EMBRACE, the ‘Play Word’ action is classified as a ‘Computer’ action.

For every student in each of the knowledge tracing iterations, I filtered out the relevant columns and their fields in a CSV file which was the input for the knowledge tracing program. I put the ‘Input Data’ field in a text file, to be read and processed separately since it also contained details like the coordinates of the objects moved.
CHAPTER 7

KNOWLEDGE TRACING ITERATIONS

During the course of developing a student model for EMBRACE, I developed four models, two of which were initial models and the other two were basic and enhanced models.

Initial Models:

- Considering every attempt of the student
- Considering the initial attempt multiple times

Basic Model:

- Considering the initial attempt multiple times along with help requests

Enhanced Model:

- Considering the previous model along with a more detailed syntax decomposition

All the attempts used knowledge tracing to model the knowledge of the student as they used the application.

Initial Models

In the basic version of the model, I used only information about the attempts that students made for the various manipulations in ‘The Best Farm’ story. The goal of the initial models was to come up with a working model of Bayesian Knowledge Tracing tailored to EMBRACE. The two versions of this model are described below.

Considering every attempt of the student
This initial model was a naive implementation of knowledge tracing for the student. The model took into account every attempt that the student made for every question. After every user step, I incremented or decremented the skill level of the student according to the Bayesian Knowledge Tracing algorithm. The guess, transition and slip probabilities were set to 0.1. The skills in this model were the target words in the vocabulary list. Here, the tap action on each word was considered to be a step, conceptually in the model.

Through this model, I could observe how the students learned the skills through the use of EMBRACE. The results from this model showed that the skill levels of all students were unnaturally high. The reason for this was that, students need to get every step right in order to move on to the next step. Therefore, even though a student struggled with certain words and sentences, the fact that they ultimately got them right contributed to the high skill levels. I noticed that the skill levels included the whole spectrum, from very low skill levels on some words in some sentences and very high skill levels. Also, it was evident that students struggled a lot more on certain steps in a few sentences than others.

**Considering the initial attempt multiple times**

According to [15], the first response is said to be the most informative. Other intelligent tutors such as the Cognitive Tutors and the ASSISTments platform that used Bayesian Knowledge Tracing also used the first response to evaluate the student but graded the student on the last response (usually the third one). However, there is some amount of debate to decide on the most informative response. Since the intelligent tutors mentioned
previously scored the first response and the students were aware of that fact, it is a bone of contention whether the first response is indeed the most informative one. Considering that the students knew that only the last response was graded, they were likely to employ guesses in their first try and only give serious thought to the last attempt [15]. However, as was observed in my previous model, the last response has a very high possibility of being correct. Also, the EMBRACE application also only moves ahead with correct answers by the student.

Thus, in order to get a more realistic snapshot of the user’s knowledge at each step, I used only the first attempt of the user for each attempt at every step of the sentence.

Another factor that I considered was the fact that a particular sentence tested more than one skill of the student. This was because the student had to move some object to another object. This feature in EMBRACE bears resemblance to the ASSISTments dataset. The ASSISTments dataset consists of a skill model where some problems are associated with multiple skills. The problems that were mapped to multiple skills were replicated once for each skill in the dataset [15]. It was found that this finer grained model was better for modeling and prediction [33]. In order to overcome the drawbacks of low skill levels by considering only the first attempt, I considered the first attempt of the student on every step of each sentence multiple times, once for every skill associated with the sentence in every step. Since every step in a sentence contains two skills, each step is repeated twice.
The results from this model also resulted in relatively high skill values for the students, which were not necessarily predictive of the student’s performance on the post tests.

**Basic Model**

The basic model uses the information from the initial models along with the help requests made by the student. I also added syntax skills in this version of the model.

**Considering the initial attempt multiple times along with help requests and syntax as a skill**

On going through the log data again, I also realized that students would sometimes move the objects in the sentence at the wrong times. For example, a student would move an object involved in step two and an object that is involved in step one, both in the first step. This could point to the idea that although the students know the vocabulary involved in the sentence, they have not understood the relationships between the words. I mapped these errors as syntax errors.

Another factor that I considered in this model was help requests. Help is provided on target words in EMBRACE by allowing the students to tap on a word and listen to the pronunciation and its meaning. These actions are called ‘Play Word’ in the log data. Since a help request implicitly implies that a student is unfamiliar with a particular word, it is treated as an incorrect attempt. Also, subsequent attempts on the same word in the
future, are treated slightly differently. A student after using the help feature on a particular word, is expected to get the manipulation right. In terms of knowledge tracing, this means that the student is more likely to guess the correct answer and thus less likely to slip on that attempt. Hence, I came up with two sets of guess and slip probabilities. The guess probability in case of a skill that a student had already received help on was slightly higher, while the slip probability in the same case was kept the same as before, since it was already high. Also, I slightly increased the slip probability for the skills in the regular case where no help was received.

**Enhanced Model**

Considering the previous model along with a more detailed syntax decomposition

Another pass on the log data, revealed that that there were a few more errors that students made, that could be attributed to syntax errors in general. I identified three common errors here: Errors made due to misunderstanding possession: Students often misunderstood the action to be performed in sentences that contained possession. For e.g. the Best Farm contains the sentence, ‘Manuel opened the goat’s pen’. Here, the student is expected to move the farmer (Manuel) to the goat’s pen. But a lot of students tended to move the farmer to the goat.
Errors made due to misunderstanding pronouns: The Best Farm also contains sentences that refer to the farmer as ‘he’. Here, I noticed students ignoring the presence of the farmer in the sentence and moving other objects in the sentence.

Usability errors: I noticed that students got confused with the different ‘pen’ objects in the story. The Best Farm contains pens for sheep, pigs, goats and cows. Thus, students would often move an animal to another animal’s pen. Although, EMBRACE classifies these actions as incorrect, the student does, in fact know the meaning of ‘pen’, they just weren’t moving the animal to the correct pen. These errors were classified as usability errors.

Apart from these, I attributed previous syntax errors to a general ‘syntax’ skill. The sets of transition, guess and slip probabilities were the same as those in the previous model.
CHAPTER 8
RESULTS AND EVALUATION

The goal of the knowledge tracing algorithm is to predict at every step, how close the student is to mastering the skills associated with that step. Since the students completed one chapter in a single session and answered the posttest questions for that chapter soon after, I used correlation to determine how accurately knowledge tracing could predict student performance in the posttests on every chapter. Chapters two to seven were used for the correlation, because chapter one is in Spanish for the bilingual version of EMBRACE. Also, I considered only the first posttest questions’ scores for correlation with the predicted skill levels. Therefore, students who answered only the follow-up question were considered to have not answered or answered incorrectly, therefore given zero points. The results show the numbers of the different types of questions and skills and the results of the correlation considering different conditions.

The total number of skills per chapter of the Best Farm associated with the action verb and complete conditions are as follows:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Action verb condition</th>
<th>Complete condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1: Skills per chapter per condition for correlation

Therefore, the total number of skills considered per condition for correlation is as follows:
Table 2: Skills per condition for correlation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of skills considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action verb</td>
<td>36</td>
</tr>
<tr>
<td>Complete</td>
<td>44</td>
</tr>
</tbody>
</table>

I performed correlation on two sets of results:

- Skill values against skills in action verb sentences (action): Here, I ran the correlation only for the skills involved in the posttest questions that were based on an action verb sentence in the story.

- Skill values against skills in non action verb sentences (non action): Here, I ran the correlation only for the skills involved in the posttest questions that were based on a non action verb sentence in the story.

- Skill values against skills in action as well as non action verb sentences (complete condition): Here, I ran the correlation for every posttest question and the skills associated with them.

For both conditions, the skill values required for a question were multiplied. This sum of this value for each question was then averaged out across the number of questions that were being considered. This is the probability of the student answering the questions correctly. This probability was then correlated to the actual average score across the questions to be considered for each student.

A posttest question is classified as an action verb question if the answer to that question lies in a sentence with a manipulation associated with it. Conversely, a posttest question is classified as a non action verb sentence if the answer to that question lies in a sentence with
no manipulation associated with it. The number of questions associated with an action verb sentence and a non-action verb sentence for each chapter of ‘The Best Farm’ is as follows:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Action Verb questions</th>
<th>Non action verb questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3: Action and non action verb posttest questions per chapter

Help requests for each word

The table below shows the number of help requests for a word and their associated standard deviations. Some of the words with the same number of help requests have the same standard deviations as well. The difference in the standard deviation is due to the fact that a few students had a lot of help requests for a single word. For e.g. In the case of ‘hay’, which has 6 help requests, a student made 5 help requests for ‘hay’ while another student requested help once. In the case of the words with 3 help requests, ‘farmer’, ‘judge’, and ‘gate’ each had single help requests by three different students and were thus equally distributed.
Table 4: Help requests across students for each word

<table>
<thead>
<tr>
<th>Number of Help requests</th>
<th>Words</th>
<th>Standard deviation (SD)</th>
</tr>
</thead>
</table>
| 1                       | Pen, purr, weeds, buckets, nest | SD for pen, purr, weeds, buckets, nest = 0.218218  
SD for barn = 0.436436 |
| 2                       | Jumped, trophy, hayloft, barn, healthy | SD for jumped = 0.307794  
SD for trophy, healthy = 0.300793  
SD for hayloft, barn = 0.436436 |
| 3                       | farmer, judge, gate           | 0.358569                |
| 4                       | Cart, pumpkins                | 0.679636                |
| 5                       | combed                        | 0.436436                |
| 6                       | Hay, farm                     | SD for hay = 1.101946  
SD for farm = 0.64365 |
| 8                       | owl                           | 1.745743                |
| 13                      | corral                        | 1.1169696               |

9 out of 20 students did not make any help requests.

Attempts at each skill (word)

The table below shows the number of attempts a student has at each skill in ‘The Best Farm’ story. These are the number of manipulations associated with the word and are not to be confused with the number of times the word appears in the story. Here, since farmer Manuel is the main character in ‘The Best Farm’, it has the most number of manipulations associated with it. Other words such as ‘owl’ and ‘cow’ are introduced from do not appear in every chapter. For e.g., the ‘owl’ is introduced in chapter 5 and stays until the beginning
of chapter 6, while the ‘cow’ although introduced in chapter 1 itself, appears intermittently throughout the remaining chapters.

<table>
<thead>
<tr>
<th>No. of attempts in ‘The Best Farm’ story</th>
<th>Skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hayloft</td>
</tr>
<tr>
<td>2</td>
<td>Eggs, nest, cart, barn</td>
</tr>
<tr>
<td>3</td>
<td>Goat, corral</td>
</tr>
<tr>
<td>4</td>
<td>Hay, bucket, pen</td>
</tr>
<tr>
<td>5</td>
<td>Pig, horse, cat</td>
</tr>
<tr>
<td>6</td>
<td>Owl, cow</td>
</tr>
<tr>
<td>7</td>
<td>Chicken, judge</td>
</tr>
<tr>
<td>19</td>
<td>Farmer</td>
</tr>
</tbody>
</table>

Table 5: Number of attempts for each word

Skills per chapter

The below table shows the skills as they appear in every chapter. ‘Everyone Helps’, ‘Cleaning Up’ and ‘Getting Ready’ have the most skills. ‘Everyone Helps’ has the most number of skills as compared to the total number of sentences in the chapter while ‘The Wise Owl’ has the least number of skills compared to the number of sentences in the same chapter.
### Table 6: Number of skills per chapter

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Skills tested</th>
<th>Number of skills tested</th>
<th>Number of sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting Ready</td>
<td>Bucket, apple, pumpkins, barn, cat, farmer, goat, cart</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Cleaning Up</td>
<td>Horse, hay, pumpkins, barn, corral, pen, farmer, cow, hayloft, pig</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Who is the Best Animal?</td>
<td>Horse, hay, apple, chicken, cat, goat</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>The Wise Owl</td>
<td>Eggs, chicken, owl, farmer, pig</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Everyone Helps</td>
<td>Horse, bucket, apple, hay, chicken, pumpkins, owl, cat, pen, cow, nest, pig</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>The Best Farm Award</td>
<td>Horse, bucket, eggs, corral, farmer, cow, judge, pig</td>
<td>8</td>
<td>11</td>
</tr>
</tbody>
</table>

### Student attempts per chapter

The below table shows the mean and standard deviation of all student attempts in ‘The Best Farm’. The ‘Getting Ready’ and ‘Cleaning Up’ chapters have the most number of attempts as compared to the total number of attempts in these chapters.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Total number of Steps</th>
<th>Mean attempts</th>
<th>Std deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting Ready</td>
<td>14</td>
<td>23.428</td>
<td>6.953</td>
</tr>
<tr>
<td>Cleaning Up</td>
<td>11</td>
<td>19.809</td>
<td>4.884</td>
</tr>
<tr>
<td>Who is the Best Animal?</td>
<td>3</td>
<td>6.095</td>
<td>1.757</td>
</tr>
<tr>
<td>The Wise Owl</td>
<td>6</td>
<td>5.19</td>
<td>0.679</td>
</tr>
<tr>
<td>Everyone Helps</td>
<td>10</td>
<td>12</td>
<td>2.387</td>
</tr>
</tbody>
</table>
The Best Farm Award | 7 | 10.04 | 2.178

Table 7: Student attempts per chapter

Mean skill values

The below tables show the mean skill values for the basic and enhanced models.

Basic Model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Mean vocabulary skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.825</td>
</tr>
<tr>
<td>3</td>
<td>0.8713</td>
</tr>
<tr>
<td>4</td>
<td>0.84</td>
</tr>
<tr>
<td>5</td>
<td>0.958</td>
</tr>
<tr>
<td>6</td>
<td>0.966</td>
</tr>
<tr>
<td>7</td>
<td>0.958</td>
</tr>
</tbody>
</table>

Table 8: Mean skill values for the basic model

Enhanced model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Mean vocabulary skill</th>
<th>Mean syntax skill</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.825</td>
<td>0.934</td>
</tr>
<tr>
<td>3</td>
<td>0.896</td>
<td>0.966</td>
</tr>
<tr>
<td>4</td>
<td>0.84</td>
<td>0.977</td>
</tr>
</tbody>
</table>
I performed correlation between the posttest scores and the predicted scores that I calculated from the students’ skill levels. The predicted values from the two enhanced models were used for correlation. I calculated results for three conditions: action, non-action and complete. The action condition correlates scores of the action questions with the skills associated with them. The action questions are the questions whose answers lie in a sentence with a manipulation associated with it. The non action questions are the questions whose answers lie in a sentence with no manipulation associated with it. An interesting result is that the non-action condition appears to have better results than the action condition. Also, unfortunately there are no significant differences between the results in the basic and enhanced model.

The correlation results for the basic and enhanced for all three conditions are described below.

**Action condition:**

**Basic Model**
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Expected</th>
<th>Pearson correlation</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.815</td>
<td>0.952</td>
<td>-.135</td>
</tr>
<tr>
<td>3</td>
<td>0.828</td>
<td>0.8412</td>
<td>.119</td>
</tr>
<tr>
<td>4</td>
<td>0.6974</td>
<td>0.9841</td>
<td>-.122</td>
</tr>
<tr>
<td>5</td>
<td>0.9868</td>
<td>0.7301</td>
<td>-.354</td>
</tr>
<tr>
<td>6</td>
<td>0.9074</td>
<td>0.9642</td>
<td>.191</td>
</tr>
<tr>
<td>7</td>
<td>0.999</td>
<td>0.984</td>
<td>.484*</td>
</tr>
</tbody>
</table>

Table 10: Correlation results for the basic model in the action condition

* Correlation is significant at the 0.05 level (2-tailed)

Enhanced Model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Expected</th>
<th>Pearson correlation</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.2935</td>
<td>0.8715</td>
<td>-.178</td>
</tr>
<tr>
<td>3</td>
<td>0.8144</td>
<td>0.8412</td>
<td>.227</td>
</tr>
<tr>
<td>4</td>
<td>0.6974</td>
<td>0.9841</td>
<td>-.122</td>
</tr>
<tr>
<td>5</td>
<td>0.9361</td>
<td>0.7301</td>
<td>-.405</td>
</tr>
<tr>
<td>6</td>
<td>0.9074</td>
<td>0.9642</td>
<td>.191</td>
</tr>
<tr>
<td>7</td>
<td>0.999</td>
<td>0.984</td>
<td>.484*</td>
</tr>
</tbody>
</table>

Table 11: Correlation results for the enhanced model in the action condition
Non action condition:

Basic Model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Expected</th>
<th>Actual</th>
<th>Pearson correlation</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td></td>
<td>Pearson correlation</td>
<td>Sig (2 tailed)</td>
</tr>
<tr>
<td>2</td>
<td>0.669</td>
<td>0.738</td>
<td>-.010</td>
<td>.965</td>
</tr>
<tr>
<td>3</td>
<td>0.988</td>
<td>0.793</td>
<td>.387</td>
<td>.083</td>
</tr>
<tr>
<td>4</td>
<td>0.8531</td>
<td>0.7142</td>
<td>.265</td>
<td>.245</td>
</tr>
<tr>
<td>5</td>
<td>0.922</td>
<td>0.888</td>
<td>.549**</td>
<td>.010</td>
</tr>
<tr>
<td>6</td>
<td>0.9502</td>
<td>0.8571</td>
<td>.131</td>
<td>.573</td>
</tr>
<tr>
<td>7</td>
<td>0.947</td>
<td>0.9365</td>
<td>.172</td>
<td>.456</td>
</tr>
</tbody>
</table>

Table 12: Correlation results for the basic model in the non action condition

** Correlation is significant at the 0.05 level (2-tailed)

Enhanced Model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Expected</th>
<th>Actual</th>
<th>Pearson correlation</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td></td>
<td>Pearson correlation</td>
<td>Sig (2 tailed)</td>
</tr>
<tr>
<td>2</td>
<td>0.6676</td>
<td>0.738</td>
<td>-.012</td>
<td>.958</td>
</tr>
<tr>
<td>3</td>
<td>0.9705</td>
<td>0.79365</td>
<td>.315</td>
<td>.165</td>
</tr>
<tr>
<td>4</td>
<td>0.853</td>
<td>0.714</td>
<td>.266</td>
<td>.245</td>
</tr>
<tr>
<td>5</td>
<td>0.918</td>
<td>0.888</td>
<td>.383</td>
<td>.087</td>
</tr>
</tbody>
</table>
Table 13: Correlation results for the enhanced model in the non action condition

Complete condition:

Basic Model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Expected</th>
<th>Pearson correlation</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.731</td>
<td>0.8299</td>
<td>-.211</td>
</tr>
<tr>
<td>3</td>
<td>0.7379</td>
<td>0.8174</td>
<td>.170</td>
</tr>
<tr>
<td>4</td>
<td>0.775</td>
<td>0.849</td>
<td>.078</td>
</tr>
<tr>
<td>5</td>
<td>0.9298</td>
<td>0.809</td>
<td>-.027</td>
</tr>
<tr>
<td>6</td>
<td>0.9216</td>
<td>0.9285</td>
<td>.348</td>
</tr>
<tr>
<td>7</td>
<td>0.9731</td>
<td>0.9603</td>
<td>.067</td>
</tr>
</tbody>
</table>

Table 14: Correlation results for the basic model in the complete condition

Enhanced Model

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Expected</th>
<th>Pearson correlation</th>
<th>Sig (2 tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expected</td>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.717</td>
<td>0.8299</td>
<td>-.183</td>
</tr>
<tr>
<td>3</td>
<td>0.719</td>
<td>0.817</td>
<td>.135</td>
</tr>
</tbody>
</table>
Table 15: Correlation results for the enhanced model in the complete condition

<table>
<thead>
<tr>
<th></th>
<th>0.7753</th>
<th>0.849</th>
<th>.078</th>
<th>.738</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.9273</td>
<td>0.8095</td>
<td>-.092</td>
<td>.693</td>
</tr>
<tr>
<td>6</td>
<td>0.921</td>
<td>0.928</td>
<td>.355</td>
<td>.115</td>
</tr>
<tr>
<td>7</td>
<td>0.973</td>
<td>0.960</td>
<td>.067</td>
<td>.772</td>
</tr>
</tbody>
</table>
CHAPTER 11
DISCUSSIONS AND CONTRIBUTIONS

The results show various statistics from the data related to the features of ‘The Best Farm’ story, student actions and their calculated skill levels. The results show that the correlation in the non-action condition has better predictability while there is no significant difference between the basic and enhanced versions of the skill models.

An important consideration here is that EMBRACE has not been designed and implemented with the skills (that I considered) in mind. Therefore, there are unequal opportunities for students to apply these skills and the chapters themselves could be at different levels of difficulty. I also observed that students’ skill levels reached ceiling levels after the first four chapters. After they hit ceiling levels, skill levels do not show significant changes even if students get an answer wrong. In chapters 2 and 3, student scores are scattered and are not predictive of their posttest scores. The chapters after chapter 4 show more predictability.

Another observation is that the skills in the basic and enhanced model do not show a lot of variation, especially in the chapters after chapter 4. There are a couple of reasons for this. Firstly, the posttest questions after chapter 4 did not test the syntax skills that I included in the enhanced skill model. Secondly, the help requests made by students were not always very useful. A significant number of students did not make a help request at all. Students who made a help request sometimes requested help on words that were verbs (e.g. combed) or words that were not skills (e.g. ‘healthy). These help requests were not useful since they
were not skills as per the skill model. Also, the students did not make mistakes related to the syntax skills in the enhanced model often enough to significantly affect the results.

The irregularities in chapters 2 and 3 are more due to exploratory actions on the part of the students i.e. students sometimes moved completely unrelated objects. For e.g. for the sentence ‘He carried the milk bucket to the cat.’, a student moved the ‘bucket’ containing milk, to the ‘cow’. These errors are hard to predict and model. Conversely, chapters 5, 6 and 7 show better results in the correlations. However, these results prove that students did well in EMBRACE as well as on the posttests. There is no evidence showing that students who did not do well on EMBRACE, did not do well on the posttests. Therefore, the results from these chapters are not reliable. This is why the model is not very predictive in general.

Interestingly, the results for the non-action condition appear to be better than those for the action conditions. This is because the skills involved in the non-action posttest questions did not have ceiling values and could therefore be considered to be more realistic. I observed that the questions marked as non-action often did not have their answers directly present in the story text. They required a certain amount of inference from the student. An example of such a question is, ‘Why did Manuel want the farm to look good?’ the answer to which is ‘to win the Best Farm award’. The answer to this question is never explicitly present in the story. Thus, the students need to think and infer the answer to such questions from the story.
Even though the results are not as predictive as we expected them to be, they give us certain recommendations for the EMBRACE application. Future work on this application could include a method to check for knowledge of the verb. The skills modeled at present could be used to design appropriate feedback for the students while they use the application. A differentiation in the difficulty levels at the chapter or story level would also be a helpful feature. The skills for each difficulty level could be modeled differently. This will help determine exactly how well a student is doing in general and whether they need more practice with a particular level of difficulty. Also, in order to reduce the exploratory actions by students, a practice story where students can experiment with EMBRACE will be a good addition. A story for this purpose is being included in the newer version of EMBRACE. This will elicit genuine responses from the students from the beginning itself.

Through this thesis, I have attempted to construct a skill model to map the acquisition of skills as students use EMBRACE. Although the model in its current state is not very predictive of student performance, it could be used to advise further functionality in EMBRACE. Different types of feedback could be provided to a student who made a vocabulary error, a syntax error or a usability error. If a student corrects his/her answer based on the feedback provided, that could be better evidence for the working of the model. Admittedly, it does not do well when students make seemingly unexplained moves on the tablet.

Thus, my contribution through this thesis is the analysis of the various types of errors students make when they use EMBRACE. This model could be used for the initial
implementation of the EMBRACE ITS and can be subsequently refined to detect other
types of errors and student behaviors.
REFERENCES


