University of North Dakota
UND Scholarly Commons

Theses and Dissertations | Theses, Dissertations, and Senior Projects

January 2014

The Effect Of Self-Explanation And Strategy Training On L2 Reading Comprehension Using An Intelligent Tutoring System

Hongxia Fu

Follow this and additional works at: https://commons.und.edu/theses
How does access to this work benefit you? Let us know!

Recommended Citation

Fu, Hongxia, "The Effect Of Self-Explanation And Strategy Training On L2 Reading Comprehension Using An Intelligent Tutoring System" (2014). *Theses and Dissertations*. 1653.
https://commons.und.edu/theses/1653

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact und.commons@library.und.edu.

THE EFFECT OF SELF-EXPLANATION AND STRATEGY TRAINING ON L2 READING COMPREHENSION USING AN INTELLIGENT TUTORING SYSTEM

by

Hongxia Fu
Bachelor of Arts, China University of Mining and Technology, 1989
Master of Science, Dakota State University, 2009

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota
December
2014

This dissertation, submitted by Hongxia Fu in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

_____
Dr. Richard Van Eck

_____
Dr. Anne Walker

_____
Dr. Shelby Barrentine

_____
Dr. Mark Grabe

_____
Dr. Tanner Jackson

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

_____
Wayne Swisher
Dean of the School of Graduate Studies

_____
Date

PERMISSION

Title The Effect of Self-Explanation and Strategy Training on L2 Reading Comprehension Using an Intelligent Tutoring System

Department Teaching and Learning

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Hongxia Fu
October 13, 2014

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ACKNOWLEDGMENTS

I thank all the educators who have guided and supported my doctoral study journey. Without their patience, support and expertise, never would I have enjoyed such a beautiful learning journey. I also thank my family members and friends who always stand behind me.

I sincerely thank my adviser Dr. Richard Van Eck who guided my whole journey in ND with all his invaluable professional and friendly support, encouragement, suggestions, and guidance; many thanks to my committee members who have always been prepared to help me in every way through my dissertation writing process. They are my mentors and friends. I owe much gratitude to these loving characters: Dr. Richard Van Eck, University of North Dakota (UND); Dr. Shelby Barrentine, UND; Dr. Anne Walker, UND; Dr. Mark Grabe, UND; and Dr. Tanner Jackson, Educational Testing Service.

I sincerely thank Dr. Danielle McNamara and her Science of Learning and Educational Technology Lab (SoLet) members at Arizona State University (ASU). Only with their kind support and expertise, could I have completed my current dissertation. I owe much gratitude to these wonderful professionals: Dr. Danielle McNamara, Dr. Jianmin Dai, Dr. Matt Jacovina, and PhD student, Erica Snow.

I sincerely thank my friends and colleagues whose support and help are appreciated during my dissertation journey. Much gratitude to these special people: Dr. Wendy Barnard, ASU; Dr. Kent Saber, ASU; Mr. John Kozel, ASU; Dr. Woei Hung, UND; Dr. Don Piper, professor emeritus from UND; M.A. Steven Finney, UND; M.A. Ivona Todorovic, Red River High School, ND; PhD Candidate Debbie Jackson, UND; Dr. Adrienne Salentiny, UND; Dr. Bing Liu, Taiyuan University of Technology (TYUT); M.A. Mei Hao, TYUT; Dr. Aiping Guo, TYUT; M.A. Rongxiang Li, retired faculty member from TYUT.

I sincerely thank all the students who participated in my study and provided rich feedback for this study. They are the treasure of this study.

I sincerely thank my husband Yonglong Liu, my son Yuqi liu, and my nephew Xiwei Wang. They have made my heart warm, love-filled, and hopeful through my doctoral learning journey.

To my mom Qiumang and dad Tonglai

ABSTRACT

While research suggests that secondary language (L2) learners at postsecondary institutions face academic reading challenges, and that reading strategy training can improve primary language (L1) learners' reading comprehension, it remains a challenge to find scalable ways to deliver such training to L2 learners. Intelligent tutoring systems (ITS) have been shown to be nearly as effective as human tutors while reaching potentially unlimited numbers of learners in a variety of subjects, including reading comprehension. However, few studies have explored the effectiveness of such systems for improving L2 learners' reading comprehension. Self-Explanation Reading Training (SERT) is an instructional model that combines self-explanation and five reading strategies (monitoring, paraphrasing, prediction, elaboration, and bridging), and has been shown to be effective. SERT has also been built into a game-based intelligent tutoring system environment called Interactive Strategy Trainer for Active Reading and Thinking ---Motivationally Enhanced (iSTART-ME). Studies have demonstrated the effects of iSTART–ME in improving L1 students' reading comprehension and learning motivation, but little evidence exists for its efficacy for L2 learners.

This research tested the reading strategy training effect through iSTART–ME on 34 incoming international L2 students admitted to a large public American higher institution in the Southwest. In addition to pretests, presurveys, posttests, and

postsurveys, these students received three hours training within two consecutive days right before their fall semester school courses formally started. The results showed that their self-explanation quality scores, short-answer reading comprehension test scores, andlearning motivation scores were significantly improved with a medium effect size. The results also suggested that students with lower self-explanation and comprehension ability benefited the most, although all students benefitted from the training. After the training, the interviewees reported that their learning with iSTART-ME was interesting and successful, expressed a desire to learn more strategies with iSTART-ME in the future, and expected to apply the strategies they learned to other subjects.

This study implied that iSTART-ME, with low cost in reaching large numbers of students, effectively taught the incoming international college students reading strategies, and improved their L2 reading comprehension abilities and learning motivations.

CHAPTER I

INTRODUCTION

The National Center for Education Statistics reports that the number of English language learners (ELL) in U.S. public schools in 2011–2012 was 4.4 million, or 9.1 percent of all students. According to The Institute of International Education, the number of international students studying in U.S. higher education institutions has increased 7.2% from 2012, reaching a total of 819,644 international students in the fall semester of 2013-2014 school year, with an estimated financial impact of $24 billion annually to the US economy. These students have a unique set of characteristics that impact their academic and professional success. Researchers have found that international students at higher institutions believed that their English-related skills were problematic, including reading comprehension (Lee, 1997; Lewthwaite, 1996; Senyshyn et al., 2000). Holmes (2004) found that some hard-working international students did not get good grades. The international students in New Zealand often had to read the textbook more slowly than their New Zealand classmates, and even read the book multiple times (Holmes). Academic English reading ability is not only crucial to a non-native English speakers' (NNES) academic success (Cummins, 1979a), but is also a determinant for their acculturation (Ying, 2001, Andrade, 2006). Boosting non-native English speakers' reading proficiency in a short time would be a significant accomplishment.

This introductory chapter describes the rationale for the current study of reading strategy training using current educational technology, specifically a game-based intelligent tutoring system, for non-native English speakers.

Background of the Study

The Fast-Growing Number of English Language Learners

According to the National Clearinghouse for English Language Acquisition (2011), from 1997 to 2009, the number of English Language Learners (ELLs) enrolled in American public schools increased by 51 percent, to 5.3 million, whereas the general student population grew only by 7.2 percent. The Institution of International Education (2014) reported that from 2011, international undergraduate enrollment began to exceed graduate enrollment. Most of the international students are non-native English speakers (NNES). Besides experiencing cultural shock, these incoming NNES students face language challenges too, even though the majority of them have passed one of the English proficiency tests required by different institutions. English language reading skill is one of the most significant challenges and has been found a determinant of acculturation for Chinese students in English-speaking countries (Ying, 2001; Kuo & Roysircar, 2004). This fast growing ELL, NNES, secondary language (L2)[1], or/and multilingual learner population poses unique challenges for educators, who must provide instruction to help their students improve their English-language skills in addition to mastering subject content knowledge and skills.

[1] The definitions of these terms vary according to different contexts, or these terms have different connotations for different people. The definitions of these terms and their specific meanings in this dissertation are presented at the end of this chapter. In this dissertation, L2 learners are a general term referring to all people whose native language is not English.

English as a Lingua Franca

Improving ELLs' reading proficiency is not just an important topic in English-speaking countries, and the problem extends well beyond formal education environments. The professional organization of TESOL (Teachers of English to Speakers of Other Languages) described in their 2009 position statement that "As a result of complex economic, cultural, and technological forces English has become a "lingua franca," or common language, in many regions throughout the world….As a result, the vast majority of those using English worldwide are themselves nonnative speakers." Moreover, it is estimated that about 2 billion people are learning English in the world (Walker, 2009). Considering that written English is an important means of international communication, English reading proficiency is crucial for all English language learners in the world to engage in serious international communications in the global market. Furthermore, written English is the major means of disseminating important academic research, so English reading proficiency is not only a prerequisite for ELLs' academic success in English-speaking countries, but crucial to the advancement of research in all disciplines worldwide.

Academic Reading Challenges Facing L2 Learners

According to *Education Week* (2011), only three percent of ELLs met the standard for 8th grade reading on the NAEP (National Assessment of Educational Progress) in 2009. That was 31 percent lower than non-ELLs. The gap between ELLs and Non-ELLs was bigger in reading than in math. Shneyderman (2012, p. 1) found that "For the majority of students who enter as ELLs in 9th grade or higher, the time in high school is not sufficient to reach reading proficiency in English." Moreover, because of the fast

growing number of international L2 learners on post-secondary campuses, the challenge of teaching L2 learners at higher educational institutions is no longer limited to the relatively small group of educators teaching in graduate schools, but now affects the much larger number of faculty members who teach undergraduates. The problem is often invisible; while many L2 learners may seem orally proficient in everyday interpersonal communication, they may lack the necessary proficiency in academic English for higher education and occupational purposes.

The academic reading proficiency gap between L2 students and L1students may directly increase the workload of L2 students at postsecondary institutions, including international students. It is a challenge that these students have to acquire English language proficiency, but also to keep pace with their classmates who are native English speakers. Research reveals that language issues tend to be a primary focus of academic adjustment for international students (Andrade, 2006; Yeh & Inose, 2003) and that academic difficulties are likely to affect their psychological adjustment (Lin & Yi, 1997). With the constraints of classroom time, it may be necessary to provide efficient instruction or support for L2 learners outside of classroom. While an L1 student might finish reading and comprehending a given set of instructional materials in one hour, an L2 student might need to spend several hours on it, which still might not guarantee adequate comprehension. At worst, the increased cognitive load of addressing language issues in addition to the effort required by the subject itself (intrinsic cognitive load) leaves fewer resources for actually learning and encoding the new knowledge (germane cognitive load), which leads to increased workload or frustration on an assignment (Paas, Renkl, & Sweller, 2003). Developing effective, practical, and scientific ways to improve

L2 learners' reading proficiency is thus a significant academic challenge for formal education, the advancement of research, and the efficiency of business worldwide. Addressing the problem will help approximately two billion English language learners worldwide, including international L2 students in higher educational institutions, to succeed in their academic pursuits or occupations.

College Instructional Material Challenges

Challenges for L1 students. Strong American Schools (2008) found 29% of undergraduates enrolled at 4-year public institutions required reading comprehension remediation; Terry (2007) found that 24% of Texas high school graduates were not able to read college level materials; Wilkins, Hartman, Howland, and Sharma (2010) concluded that "At the 75 percent comprehension level, 51 percent are able to read and comprehend 95 percent of the textbooks used in entry-level English courses; 80 percent are able to read and comprehend 50 percent of the textbooks; and 9 percent are able to read no more than 5 percent of the textbooks" (p.10). This finding was also reflected in other researchers' work. The National Endowment for the Arts (2007) found that students at college and university level did not read as well as undergraduates and graduates in the USA from prior years.:

- Among college graduates, reading proficiency has declined at a 20%–23% rate from 1992 to 2005.
- The average score declined for the bottom 90% of readers.
- Among high school seniors, the average score has declined for virtually all levels of reading.
- 35% of high school seniors now read proficiently (p.13).

Although not all students are able and independent readers, Alvermann and Nealy (2004) found that content area teachers tended to assume that by the time students enter middle and high school, they have become active and independent readers and learners. It is not unexpected that professors at post-secondary institutions regard their students as strategic readers and learners, yet students frequently show that they struggle with understanding expository texts in the content areas (Irvin, Buehl, & Klemp, 2007). Reading expository texts calls for different schematic knowledge than does reading narratives (Smagorinsky, 2009).

At the level of higher education, reading becomes a primary tool for learning. When traditional classroom lectures are replaced by classroom discussions, it is assumed to be the students' responsibility to acquire the information contained in the assigned textbooks or other instructional materials for active classroom discussions. Reading comprehension ability thus plays a crucial role in students' academic success. Under such a situation, effective scaffold of systematic and focused reading strategies training becomes demanding, especially for understanding expository texts. Reading strategies have been recognized as essential to overcome reading problems and improve comprehension (Mcnamara, 2009; McNamara, Ozuru, Best, & O'Reilly, 2007). The College Board Standards included reading strategies in their English Language Arts College Board Standards for College Success™ (2006), which is a sign of the importance of reading strategy training.

Challenges for L2 students. Comprehending college textbooks also remains a big challenge for L2 students attempting to solve domain specific problems. Take Norwegians, who are known for their English fluency, as an example. After examining

578 Norwegian university students, Hellekjær (2009, p. 198) found that "about 30% of the respondents had serious difficulties reading English, while an additional 44% found it more difficult than reading in their first language." In fact, the majority of L2 learners find their depth of reading comprehension in L2 is generally not as good as that in their native language. Maarof and Yaacob (2011) found that proficient readers in L1 in secondary Malaysian schools were not necessarily good readers in L2 (English), even though English is an official language in Malaysia. Moreover, Zwann and Brown (1996) found that those L1 college students who registered for French classes for two years generated more explanatory inferences in English (native language) than in French after reading English and French stories, reflecting more skilled or deeper comprehension in English than in French. Linguistic barriers in French having been excluded from this study, these students seemly did not transfer their reading strategies in English that helped their explanatory inferences to their French reading. So explicit strategy training in L2 may be helpful for L2 learners to transfer their L1 strategies to their L2 reading. Probably these L1 reading strategies that L2 learners have acquired can or/and need to be activated by explicating teaching and practicing these strategies in L2 so as for the L2 learners to successfully use them in L2.

Given the reading challenges that L1 students face in post-secondary institutions, it is inevitably also a challenge for L2 students to comprehend the assigned reading materials and participate in active classroom discussions related to the assigned reading materials. It was reported that majority Japanese students at college have difficulties in comprehending instructional materials when they entered the United States (Yuko, 2008). Lack of English proficiency is one of the key factors contributing to Chinese international

students' academic stress (Yan & Berliner, 2009), and English remained their biggest obstacle in higher education (Yuan, 2011). Without deep comprehension of assigned reading materials, active classroom discussions become ineffective, at best. Unfortunately, sometimes L2 students not only cannot understand instructional materials, but also cannot understand the assignment instructions. In current post-secondary institutions, if students have problems with their writing, they can go to writing center for help. However, for students who have difficult in reading, they are not likely to find an official reading service center on campus. Therefore, convenient and low cost reading comprehension training is critical and in short supply to help L2 students become academically and professionally successful.

Reading Strategies

Research in both L1 and L2 demonstrate that college textbook comprehension is a big challenge for a great number of college students, both L1 and L2, although reading comprehension ability sets the basis for learners to succeed in academic areas. To improve reading comprehension, extensive research has shown the effectiveness of reading strategies instruction on improving reading comprehension (e.g. Pressley, 1998, 2000, 2001, 2002, 2006; Trabasso & Bouchard, 2002; Wilkinson & HyneSon, 2011). According to Afflerbach, Pearson, & Paris (2008), "Reading strategies are deliberate, goal-directed attempts to control and modify the reader's efforts to decode text, understand words, and construct meanings of text" (p. 368). According to Wilkinson and HyneSon (2011), research on strategy instruction has included laboratory and classroom-based studies, which leads to the conclusion that "there is now no doubt that instruction in small repertoires of comprehension strategies, when implemented well, connecting,

and etc. produces robust effects on measures of comprehension, including standardized tests (e. g., Andeson, 1992; Brown et al., 1996; Collins, 1991)" (p. 364). The number of strategies is huge, for example, visualization, note-taking, read-aloud, re-reading, previewing, setting purpose, summarizing, graphic organizing,

Given these reading challenges, providing convenient and affordable reading comprehension strategy training to college students, especially L2 students, should be instrumental for these students to successfully learn their subject content materials.

Statement of the Problem

Strategy training is a low-cost way to improve reading comprehension in developing readers. Willingham (2006) compared the effectiveness of strategy instruction on reading comprehension to "a bag of tricks that can indirectly improve comprehension. These tricks are easy to learn and require little practice" (p.45). If readers can retain the strategies they have learned and apply them in their reading, they can improve their comprehension. This does not deny the role of decoding, vocabulary and background knowledge in reading comprehension, which is still critical. However, decoding, vocabulary and background knowledge does not ensure reading comprehension, reading strategy training can significantly improve the reading comprehension skills based on enough decoding, vocabulary and background knowledge.

Because strategy training boosts L1 reader's comprehension, it may hold some promise to boost secondary or foreign language (L2) reader's as well. Although many studies (e.g. Cummins, 1979, 1980, 1991; Horita, 1996, 2000; Yamashita, 2002) have found that L2 reading comprehension is correlated to L2 linguistic ability, L1 reading

ability, and metacognition, the relationship between strategy training and reading comprehension lacks consensus due to the limited number of studies on strategy training with L2 learners (Grabe, 2009).

Grabe (2009) believed that simply reading alone could not develop effective reading strategies which should be explicitly instructed. Regarding the rather limited empirical studies on L2 strategy research, this study will focus on the effectiveness of reading strategies training on L2 learners. Much research supported that sufficient L2 linguistic knowledge was the premise for text comprehension and applying reading strategies (Alderson, 1984; Clarke, 1979; Cummins, 1979, Laufer, 1997), so intermediate proficiency L2 learners will be the targeted research population. Given the challenge for L2 learners to understand the college textbooks and academic journal papers (Snow, 2002), this study will focus on the comprehension of academic exploratory reading materials.

L2 Reading Strategies

Compared with the extensive reading comprehension strategy training research in L1, studies on reading strategy training with L2 learners are limited. In a meta-analysis of empirical research on L2 reading strategy instruction, Taylor, Stevens, and Asher (2006) found only ten published studies and 12 dissertations that met their review criteria. Because each study used a different design and framework in L2 reading strategy study (Yamashita, 2002), the matter was further complicated. More studies are warranted in order to understand the relationship between strategy training and L2 reading comprehension.

If strategy instruction can boost L2 learners' reading comprehension, it will be important to understand more about what kind of strategy instruction will be most effective for L2 learning. Sheorey & Mokhtari (2001) found that college-level L2 students with high reading ability, like their L1 peers with high reading ability, reported higher usage of cognitive and metacognitive reading strategies. Walter (2004) has found that intermediate L2 learners did poorly on the overall comprehension of texts, even though they understood individual sentences in L2 text. So it is relating ideas between sentences or with the reader's prior knowledge that may cause the comprehension failure for intermediate L2 learners. Taylor (1999) defines metacognition as "an appreciation of what one already knows, together with a correct apprehension of the learning task and what knowledge and skills it requires, combined with the ability to make correct inferences about how to apply one's strategic knowledge to a particular situation, and to do so efficiently and reliably" (p.37). Metacognitive strategies include intentional strategic approaches to learning, such as the reader adjusting his/her reading strategies by monitoring his/her comprehension. It is reasonable to assume that metacognitive reading strategy training would help L2 learners to read most efficiently. Because of this, and because of the research on L1 strategy training, it is reasonable to assume that metacognitive reading strategy training will help L2 learners to read most efficiently. Due to the scarce reading comprehension resources existing on campus, classroom instruction and human tutoring are usually the only available options, and they are in short supply. How, then, to deliver reading comprehension strategy training to the growing number of L2 learners?

Intelligent Tutoring Systems

Coaching and tutoring have proved to be among the most important learning or teaching methods for a range of learning domains, including reading comprehension. However it is a challenge to hire many coaches or tutors in a short time to teach many international students reading strategies, and the cost is high. The potential exists to solve this problem using current information technology. Grasser (2007) argues that "Computers are able to train many reading comprehension strategies and are expected to take a more prominent role in the future" (p.20). This can overcome the limitations of social human resources in providing individualized training to more L2 learners worldwide. Regarding the advantages of computers over human, Grasser elaborates as follows:

> Computers do not have the same limitations on fatigue, memory, and grain size [amount of computation] that human instructors face. They can potentially diagnose hundreds of reading problems, maintain a student profile on hundreds of variables, tune strategies with unlimited degree of complexity, and flexibly tailor a particular strategy to the student's learner profile. (p.20-21)

Such approaches can also save time and costs, and moreover, extend the availability of effective training to students through online access. Compared with human instructors, conventional computers may be criticized for lacking the human face-to-face personal interactions, feelings, and adaptability. For example, conventional computer-assisted language learning (CALL) involves description of strategies, examples of strategies, and practices in text or video, which is didactic and lacks individualized feedback and guidance. However, the computers of today are becoming more sophisticated and are able to provide interactive and adaptive timely feedback. Intelligent

tutoring systems (ITSs) with animated conversational agents can speak in natural language with a kind of "personal" communication through careful word choice. For example, the Interactive Strategy Trainer for Active Reading and Thinking (iSTART) is an ITS which tutors reading strategies using animated agents, provides feedback to each user's performance in natural language, guides each user to learn the strategies, self-explanation skills, and improve their reading comprehension of expository texts like textbooks.

Self-explanation is a self-generated explanation of the meaning of information to oneself, while reading a text. McNamara and Magliano (2009) defined self-explanation as "the process of explaining text or material to oneself either orally or in writing" (p. 61). Self-explanation can be initiated or occur naturalistically to explain what is reading. Roy and Chi (2005) have explained the function of self-explanation as "a domain general constructive activity that engages students in active learning and insures that learners attended to the material in a meaningful way while effectively monitoring their evolving understanding." Self-explanation overlaps with think-aloud and retelling, but they are different. Thinking-aloud is usually used as a measurement of thinking, which is not specifically directed as self-explanation is collected, but only collect the verbalized thoughts and processes of which the participant was conscious (Trabasso & Magliano, 1996). Retelling is a strategy to monitor how much the reader remembers of the text by recalling in the reader's words the information or stories contained in the text. Koskinen, Gambrell, Kapinus, and Heathington (1988) stated that "Retelling requires the reader to organize text information in order to provide a personal rendition of it. Engaging in retelling focuses the reader's attention on restructuring text holistically" (p. 892).

iSTART has demonstrated its effectiveness in improving L1 students' self-explanation quality and reading comprehension by providing reading strategy scaffolding (O'Reilly, Sinclair, & McNamara, 2004; McNamara, O'Reilly, Best, & Ozuru, 2006). It has not been studied, however, as much with L2 learners.

Game-Based ITS Environment

One potential weakness of ITSs is that users can become disengaged and bored over time, although the ITSs are novel to them at the beginning (Baker, D'Mello, Rodrigo, & Graesser, 2010; Bell & McNamara, 2007). On the other hand, digital games provide a new medium to motivate and engage today's learners who grow up with digital technology (Prensky, 2005; Van Eck, 2006). Digital game-based learning (DGBL) has become an accepted, effective instructional method. To improve the engagement of iSTART, the iSTART research group built up a game-based environment --- iSTART–ME (Motivationally Enhanced) for extended practice, on top of iSTART. Jackson and McNamara (2013) found that high school students preferred to work with iSTART–ME rather than iSTART, and these two training systems produced equivalent performance at the posttest and delayed retention test.

However, it remains unknown whether the current version of iSTART-ME would successfully motivate and tutor L2 users to learn these reading strategies and self-explanation approach to improve their English linguistic ability. It is also unclear whether international students newly admitted to higher educational institutions could benefit from the iSTART-ME program, considering that they have already been literate in their L1, because the participants in this study are international undergraduates and graduates

admitted to a large southwestern university, who have finished their k-12 or k-12 and undergraduate education in their L1.

Research Purpose and Questions

The purpose of this study is to explore whether newly arrived international students in American higher institutions could successfully learn reading comprehension strategies through iSTART-ME, and whether the strategy learning could improve their English reading comprehension as it has with L1 learners. A second question focuses on the participants' perceptions of their training experience, for example, whether the participants like the iSTART-ME system, whether they like to continue studying with the system after the training intervention. The rationale for this second question is that what appeals to L1 learners may not appeal to L2 learners. Of course, a subsidiary question is whether the iSTART-ME system is user friendly enough for L2 users to easily learn to use and navigate in the system, taking account of the L2 users' language and technology background. The third question is what individual characteristics are related to or can predict the training effects, such as linguistic ability, learning motivation level, and prior reading strategies. These three questions become the main research questions of this study.

This study used a mixed methods research approach to study the following research questions:

1. What are the effects of iSTART–ME on L2 reading comprehension and self-explanation?

1.1. Do participants improve their reading comprehension after the training?

1.2. Do participants improve their self-explanation quality after the training?

1.3. Do participants write longer self-explanations after the training?

2. What are participants' perceptions of their learning experience with iSTART-ME?

2.1. Has participants' learning motivation changed through the training?

2.2. Do participants hold a positive attitude towards iSTART-ME?

3. What factors account for changes in reading comprehension and self-explanation?

3.1. Is there correlation between self-explanation word count change and self-explanation score change?

3.2. Is motivation change correlated to participant's self-explanation score change?

3.3. Are vocabulary scores correlated to participant's self-explanation score change?

3.4. Are prior metacognitive strategies related to self-explanation score change?

3.5. Are prior metacognitive strategies related to reading comprehension score change?

3.6. What individual characteristics predict self-explanation score change?

3.7. What individual characteristics predict comprehension score change?

Significance of the Study

Regarding the rather limited studies on L2 strategy research and the advantages of ITSs and DGBL, this study explores the effectiveness of iSTART-ME in teaching post-secondary international students reading strategies as well as the effect of applying these strategies in L2 reading comprehension. This study is important for several reasons. First, it provides a new environment, a combination of ITSs and games, to test the relationship between strategy training and reading comprehension improvement for L2 learners. Second, it explores the relationship between a number of factors --- vocabulary, pre-strategy application, pre-comprehension level, motivation, self-explanation, and strategy learning --- and reading comprehension for L2 learners. Third, it adds learners' learning experiences as a new perspective from which to study the relationship between strategy training using iSTART-ME and L2 users' reading comprehension improvement. Fourth, because this training involves both reading and writing (participants are asked to write down their self-explanation), L2 learners may derive specific benefits from the written component. Because the function of self-explanation in L2 comprehension is hardly researched, this present study is an exploration study.

Given doubts expressed in the literature regarding the automatic transfer of L1 reading skills to L2 reading (e.g. Carrell, 1991; Clarke, 1980; Taillefer, 1996), explicitly teaching self-explanation skills and the other five strategies taught in iSTART-ME may improve readers' overall comprehension. So far, few studies have researched the possibility of improving L2 learners' reading comprehension through strategy and self-explanation training at all, let alone with ITSs and digital games. This research may provide information for understanding not only the role of self-explanation as a strategy

approach in L2 reading, but also the potential role of game-based ITS environment in L2 teaching and learning.

Definitions of Terms

- **Academic English**

Academic English is English used in the learning of academic subjects in a formal schooling context. It is also called Cognitive Academic Language Proficiency (CALP) in Cummins' work (Cummins, 1981). It took 5-7 years to develop the proficiency to cope with the academic demands in school content areas, including specific academic terms, technical language, and speech registers related to each field of study.

- **Animated Conversational Agent**

An animated conversational agent is a talking head or figure built in an intelligent tutoring system (ITS) to help students actively construct knowledge through conversations in a conversational interface (Graesser, VanLehn, Rose, Jordan, & Harter, 2001). Learners can type in their responses and the agent can respond with something relevant and useful.

- **Bridging Strategy**

"Bridging "is the process of linking ideas and understanding the relations between separate sentences in the text. Deep comprehension requires more than merely interpreting individual sentences; the reader must also be able to integrate individual sentence meanings into a coherent text level representation (Kintsch, 1988; 1998). Making inferences is critical to text comprehension because texts normally do not (or

cannot) state all of the relevant information (e.g., McNamara et al., 1996).” (McNamara, 2009. p.36).

- **Comprehension Monitoring Strategy**

“Comprehension monitoring is the process of being aware of understanding. In effect, the process of comprehension monitoring falls out of using effective reading strategies because to use a strategy the readers must be at least somewhat aware of their level of understanding.” (McNamara, 2009. p.35)

- **Digital Game-Based Learning (DGBL)**

Digital Game-Based Learning (DGBL) is an alternative instructional method that integrates educational content in video games with the goal of combining serious learning with interactive entertainment.

- **Educational Games (Serious Games)**

An educational game, or serious game, is a game designed to help people to learn a subject, a concept, an event, a culture, or a skill, including board, card, and digital games.

- **Elaboration Strategy**

“Elaboration is the process of making inferences that link what is in the text or sentence to related knowledge. For example, when reading the following sentence about heart disease, ‘Coronary artery disease occurs when the arteries become hardened and narrowed,’ the reader might make the link to prior knowledge that arteries supply blood to the heart muscle. The reader might also use general knowledge or logic to infer that narrowed arteries would reduce blood flow to the heart muscle, result in a lack of oxygen supply, and potentially lead to a heart attack” (McNamara, 2009. p. 36).

- **English Language Learner** (**ELL)**

"English language learners (ELL) are also known as ESL (English as a second language) students or bilingual students, or LEP (Limited English proficiency) students. According to the federal government, an LEP/ELL is an individual who is enrolled or preparing to enroll in an elementary or secondary school, whose native language is other than English, which has had a significant impact on the individual's level of English language proficiency, and whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual the ability to meet the State's proficient level of achievement on State assessments."

(Source: Public law 107-110, title ix, part a, sec. 9101, (25)

- **First Language (L1)**

In this study, the first language (L1) is not differentiated from native language, mother tongue, or arterial language. The key denotation is the language(s) a person has learned from birth or within the critical period, including the vocabulary, linguistic variations (i.e. grammar, semantics), or other types of communication (e.g. body language). It allows the person to practice a particular 'slang' to communicate within the environment he lives in. The L1 creates a specific mental structure --- code that will affect a person's future languages acquisition. A person's first language is not necessarily the language he/she uses most or is most comfortable with.

- **Intelligent Tutoring Systems (ITS)**

Intelligent Tutoring Systems (ITS) are educational systems combining artificial intelligence, cognitive science, and education to provide instruction tailored to the needs of individual learners.

- **International Students**

According to UNESCO, an international student in the USA is an individual who is enrolled for credit at an accredited higher education institution in the U.S. on a temporary visa, and who is not an immigrant (permanent resident with an I-51 or Green Card), an undocumented immigrant, or a refugee. (Source: UNESCO)

- **iSTART**

Interactive Strategy Trainer for Active Reading and Thinking (iSTART) is an automated trainer program with animated agents. It focuses on teaching trainees various reading strategies through self-explanation to comprehend expository texts. It is designed for secondary school or college students.

- **iSTART-ME**

iSTART-ME (Interactive Strategy Trainer for Active Reading and Thinking-Motivationally Enhanced) is built on iSTART, but digital games have been added for engaging students in extended strategy application practice.

- **L2 Learners**

In this study, the term "L2 learners" does not specifically refer to a group of people whose English has not reached certain standards, as the term "ELL" used in k-12 schools; on the contrary, "L2 learners" refers to any nonnative English speakers (NNES), regardless their English proficiency.

- **Learning Motivation**

According to a social-cognitive point of view, motivation is directly linked to a person's learning activities (Duncan & McKeachie, 2005). Although many models of motivation may be relevant to student learning, this study concentrates on three general

types of motivational beliefs, measured by Motivated Strategies for Learning Questionnaire (MSLQ). These three motivational beliefs are: (a) self-efficacy beliefs (judgments of one's capabilities to do the academic task), (b) task value beliefs (beliefs about the importance of, interest in, and value of the task), and (c) goal orientations (whether the focus is on mastery and learning of the task).

(Source: Pintrich, 1999).

- **Literacy**

"The ability to identify, understand, interpret, create, communicate, and compute, using printed and written materials associated with varying contexts. Literacy involved a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society." (p.13)

(Source: United Nations Educational, Scientific, and Cultural Organization (UNESCO), 2004, the plurality of literacy and its implications for policies and programmers. Paris: UNESCO).

- **Native English Speakers (NES)**

Native English Speaker (NES) is a person whose first language he learns to speak is English. According to Davis (2004), "A child may be a native speaker of more than one language as long as the acquisition process starts early and necessarily prepuberty. After puberty (Felix, 1987), it becomes difficult--not impossible, but very difficult (Birdsong, 1992)--to become a native speaker."

(Source: Alan Davies, "The Native Speaker in Applied Linguistics." The Handbook of Applied Linguistics. Blackwell, 2004)

- **Non-Native English Speakers (NNES)**

An acronym of native English speakers. In this study, this term is not different from L2 learners.

- **Prediction Strategy**

"The prediction strategy involves thinking about what might be coming next in the text" (McNamara, 2009. p.36).

- **Paraphrasing Strategy**

"Paraphrasing is the process of restating the text in different words, or in the reader's own words. It doesn't go beyond the information in the text, so it's not an explanation of the text. In the reading strategy literature, paraphrasing is often not recognized as an effective strategy. However, it is an important part of the explanation because many readers often paraphrase the sentence to begin an explanation" (McNamara, 2009. p.35).

- **Reading Comprehension**

Reading comprehension is the "process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (Snow and RAND, 2002, p.11).

- **Reading Strategy**

"Reading strategies are deliberate, goal-directed attempts to control and modify the reader's efforts to decode text, understand words, and construct meanings of text" (Afflerbach, Pearson, & Paris, 2008, p. 368). There are many different strategy names, given from different perspectives. Examples of strategies are previewing, monitoring,

summarizing, and questioning. Reading strategies also belong to the broader category of learning techniques.

- **Reading Skills**

"Reading skills are automatic actions that result in decoding and comprehension with speed, efficiency, and fluency, and usually occur without awareness of the components or control involved" (Afflerback, Pearson, & Paris, 2008, p.368).

- **Secondary Language (L2)**

A secondary language is any language that a person learns or acquires in addition to his/her first language, no matter whether the language is generally used or not in the area he/she learns the language. In this study, just as in the second language acquisition field, L2 is not specifically differentiated from foreign language.

- **Self-Explanation**

Self-explanation is a self-generated explanation of the meaning of information to oneself while reading a text. McNamara and Magliano (2009) defined self-explanation as "the process of explaining text or material to oneself either orally or in writing" (p. 61). In this study, participants either spontaneously self-explain or are prompted to do so (McNamara, 2004). According to Roy and Chi (2005),

> Self-explanation is a domain general constructive activity that engages students in active learning and insures that learners attend to the material in a meaningful way while effectively monitoring their evolving understanding. Several key cognitive mechanisms are involved in this process including generating inferences to fill in missing information, integrating information within

the study materials, integrating new information with prior knowledge, and monitoring and repairing faulty knowledge. (p. 272)

Self-explanation overlaps with think-aloud and retelling, but they are different. Thinking-aloud is usually used as a measurement of thinking, which is not specifically directed as self-explanation is collected, but only collect the verbalized thoughts and processes of which the participant was conscious (Trabasso & Magliano, 1996). Retelling is a strategy to monitor how much the reader remembers of the text by recalling in the reader's words the information or stories contained in the text.

- **Self-Explanation Word Count (SEWC)**

Self-Explanation Word Count **(S**EWC) is the total number of words in a self-explanation, which represents how long a written SELF-EXPLANATION is in this study.

CHAPTER II

LITERATURE REVIEW

In order to provide context for the current study, which examines the effects of strategy training through SELF-EXPLANATION using ITS, this chapter begins by reviewing available research on L1 and L2 reading comprehension and reading strategies. Secondly, specific factors related to L2 learners' reading comprehension are reviewed. Finally, technology in reading strategy training is introduced. The literature review provides evidence of the necessity and promise of studying the effects of reading strategy and SELF-EXPLANATION training for L2 learners. The advantages and disadvantages of ITS are reviewed, especially the publications related to iSTART-ME, which shows a potential as an effective tool for training L2 learners in reading strategies and SELF-EXPLANATION, although limited research is available for the application of this tool to L2 learners.

Reading Comprehension

Definition

What is Reading? Before answering what reading comprehension is, we need to understand what reading is. The Reading First Program of No Child Left Behind (2001) defines reading as follows:

> The term reading means a complex system of deriving meaning from print that requires all of the following: (A) The skills and knowledge

> to understand how phonemes, or speech sounds, are connected to print. (B) The ability to decode unfamiliar words. (C) The ability to read fluently. (D) Sufficient background information and vocabulary to foster reading comprehension. (E) The development of appropriate active strategies to construct meaning from print. (F) The development and maintenance of a motivation to read. (p. 103)

The above *reading* definition indicates that NCLB identifies "deriving meaning"—comprehension—as the purpose of reading, and explains the variables in reading instruction: phonemic awareness, decoding, fluency, background knowledge, vocabulary, active strategies, and motivation. This definition interprets reading from a reader's perspective, and implies an instructional shift from an emphasis on decoding and recitation in the 19th century to meaning-oriented instructional models of the 20th century (Stahl, 1999). The goal of reading instruction is comprehension. After the decoding and fluency phases have been passed, developing content vocabulary, strategies, and motivation becomes important in reading instruction, which has prompted this present study to include strategy and motivation variables in this comprehension study.

What is reading comprehension? If meaning-making, comprehension, is the goal of active reading, then what is comprehension, per se? Since the 1980s, researchers' definitions of reading comprehension have emphasized elaborative inferences and the reader's active role in reading, beyond literal information recognition and recall (e.g. Kintsch &Van Dijk, 1978; Just & Carpenter, 1980; McClelland & Rumelhart, 1981; Farrar, 1986, Stahl, 1999). This constructivist view of meaning construction is also found in the report of the RAND Reading Study Group (RRSG), funded by the Department of

Education's Office of Educational Research and Improvement (OERI) in 2002. In a 7-page chapter, the Rand Group (2002) defines reading comprehension mainly in terms of the three elements involved in the reading process: *reader*, *text*, *reading activity*. These three interactive elements are situated in a larger *social-cultural context*. See Figure 1.

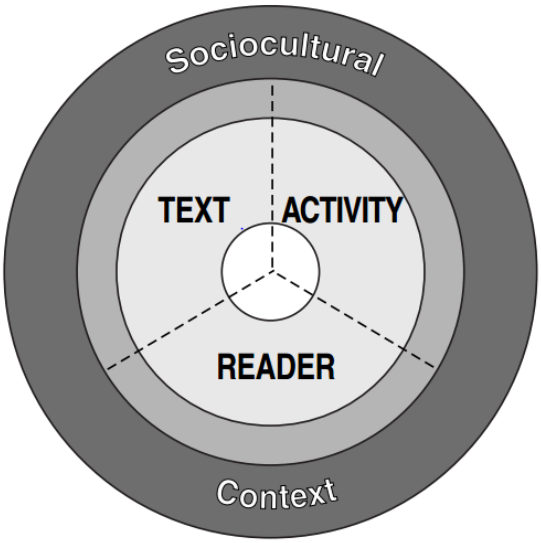


Figure 1: Elements of Reading Comprehension

(Source: An Heuristic for Thinking about Reading Comprehension, from Snow and RAND: 2002, p. 12, permitted by RAND.)

They broadly defined reading comprehension as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (p. 11), but elaborately discussed the factors that have impacts on the *reader*, *text*, and *reading activity*. Comprehension is the result of the relationship between the features of the text, the reader's knowledge, abilities, and activities have a large effect on comprehension. In terms of the *reader*, many factors related to the reader affect the reading process, such as the reader's knowledge, experience, and cognitive, motivational, linguistic, and non-linguistic capacities. In terms of the *text*, features like content,

vocabulary, sentence structure, and format have a large effect on comprehension, as comprehension is the result of the relationship between the features of the text and the reader's knowledge, abilities, and activities. *Reading activity* involves purposes, processes, and consequences. A reader's purposes prior to reading may change as he/she reads; during reading, the reader engages in a complex array of cognitive activities to process the text through and beyond "decoding, higher-level linguistic and semantic processing, and monitoring" (p.15). The direct consequences of a reading activity include knowledge, application, and engagement. Reading activities may also have indirect consequences, such as learning new vocabulary, and acquiring incidental knowledge. The *sociocultural context* often refers to the settings in which reading activities occur, like the context of instruction, classrooms, schools, and any other neighborhood context, involving many factors like economic resources and class membership.

This definition of comprehension includes the factors that are related to the reading process as well as the factors that affect the final text comprehension, or the purpose of reading. This definition also applies to the reading comprehension in L2, although more variables, like L1 and L2 linguistic ability, are important factors. Among the three key elements of reading process, *reader*, *text*, and *reading activity*, this present study examined the *reader's* motivation and reading ability change after intervening the *reading activity* with self-explanation and strategy training, while controlling the vocabulary, sentence structures, and the format of *text.* The purpose of this design was to compare the consequences of the reading strategy training intervention.

Interactive reading models. The above definition of comprehension is the result of different reading theories, including but not limited to schema theory, reading and

writing model, and attitude influence model. Schema theory emphasizes that the reader constructs the meaning by activating his/her current organized knowledge, schemata (Anderson, 1984); the reading/writing model stresses that meaning is negotiated between the reader and writer through the text (Pearson & Tierney, 1984); the model of attitude influence addresses the active role of motivation and attitude in reading (Mathewson, 1994). As is clear in the above definitions, the interactive reading models recognize the interactive meaning-construction process which involves the text and the reader's experience and reading skills throughout the whole process (McCormick, 1988, Samuels & Kamil, 1988). Willingham (2004) stated that effective readers relate sentences in two levels: a text base and a situation model. "A text base is a web of connected ideas created from what you've read"; "situation model relies on both the text and the reader's background knowledge" (Willingham, p. 40). Building a text base is necessary for reading comprehension, but not sufficient for real, deep, and rich comprehension which requires a situation model. Furthermore, reading comprehension is not only a process of building a connection between what the reader knows and what the reader does not know (Searfoss & Readence, 1994), but is also affected by a cognitive component (information processing and evaluation), an affective component (attitude, motivation, and self-esteem), and a conative component (personality, volition, and temperament). Mathewson (1994) and Stahl (1999) have emphasized the importance of addressing readers' affective issues and increasing their reading motivation. The present study used affective and motivation measurements to investigate the affective effects of strategy training in a game-based environment on the participants, in addition to measuring their learning achievement effect, comprehension level, after the training or intervention.

Reading Strategies

One proven way to facilitate effective reading processes is training students to use reading strategies, as specified in the definition of *reading* by NCLB (2001). Reading strategies, such as monitoring, paraphrasing, prediction, connection, and summarizing, help readers monitor their reading process, connect the text with their prior knowledge, bridge the associations among sentences, make inferences, etc. Extensive research has confirmed the effectiveness of explicit reading strategy training in L1 reading (e.g. Pressley, 1995, 2000, 2002a, 2002b, 2006; Trabasso & Bouchard, 2002; McNamara, 2004; Wilkinson & Hyneson, 2011), but the relationship between strategy training and L2 reading comprehension lacks consensus due to the limited number of studies (Grabe, 2009).

Classification of Strategies

Reading comprehension ability lays the foundation for learners to succeed in academic areas. Extensive research has shown the effectiveness of reading strategy instruction for improving reading comprehension (Snow & RAND, 2002; Trabasso & Bouchard, 2002; McNamara, 2004; Wilkinson & Hyneson, 2011). Researchers have identified a broad array of reading strategies, including visualization, highlighting, note-taking, read-aloud, re-reading, previewing, setting purpose, summarizing, graphic organizing, connecting, etc. They have classified these strategies in various ways as a consequence of their various views on reading processes and strategies (Koda, 2004). For example, Chamot and O'Malley (1994) classified strategies to three function-based strategy clusters: cognitive (e.g. paraphrasing), metacognitive (e.g. monitoring), and social and affective strategies (e.g. seeking outside help); Paris, Wasik, and Turner (1991)

simply classified strategies to before-, during-, and after-reading strategies (e.g. prediction, monitoring, and evaluation respectively). Other researchers have categorized reading strategies into different groups. For example, Anderson (1991) has introduced five categories: supervising, supporting, paraphrasing, establishing text coherence, and test taking. After reviewing 205 experimental studies on reading strategy instruction, Trabasso and Bouchard (2002) identified 12 strategy instructional categories: comprehension monitoring, graphic organizers, listening actively, mental imagery, mnemonic instruction, prior knowledge, question answering, question generation, and story structure. Under each category, there are many strategies. Mokhtari and Reichard's (2002) Metacognitive Awareness of Reading Strategies Inventory (MARSI) covers metacognitive awareness of reading strategies, global reading strategies, problem-solving strategies, and support reading strategies. Among these strategies, cognitive and metacognitive strategies appear frequently in L2 strategy research literature and appear to be the most effective and unifying strategies. A synthesis of the related literature is given below.

Cognitive strategies are "internal processes by which learners select and modify their ways of attending, learning, remembering, and thinking" (Gagne, Brigg, & Wagner, 1988, p. 67), like summarizing and note-taking. Metacognitive strategies monitor or regulate cognitive strategies to control the reader's thinking or learning, like making predictions before reading and monitoring comprehension during reading, and to assist the reader in constructing meaning from the text. Researchers have reported that students make significant improvement in learning when metacognitive strategies are taught and

practiced in reading (Cross & Paris, 1988; Schraw & Moshman, 1995). Mokhtari and Reichard (2002) summarize that

> […]indeed, researchers agree that awareness and monitoring of one's comprehension processes are critically important aspects of skilled reading. Such awareness and monitoring processes are often referred to in the literature as metacognition, which can be thought of as the knowledge of the readers' cognition about reading and the self-control mechanisms they exercise when monitoring and regulating text comprehension. (p.249).

However, there is considerable debate about the exact scope and meaning of metacognition, and the relationships and nature of different types of metacognitive processes (Alexander, Schallert, &Hare, 1991; Schraw & Moshman, 1995), although it is agreed that metacognitive activities include planning, monitoring, and evaluation (Jacobs & Paris, 1987; Kluwe, 1987). Specifically, there is general agreement about the concept of metacognitive and cognitive strategies, but the classification of specific strategies with modest differences varies. For example, prediction is classified as metacognitive strategy by Sheorey and Mokhtari (2001), whereas inferring (prediction is one kind of inference) is classified as a cognitive strategy by O'Malley and Chamot (1990). The perspective from which each strategy is viewed determines the classification. Take the inferring strategy as an example. If the reader *chooses* to *infer knowledge from the text*, the decision or *choosing* is a *metacognitive* process. The inferring process also involves *evaluating* the information collected from the text, and understanding what is not stated literally in the text. Therefore, inferring can be classified as metacognitive strategies. However, the inferring itself involves *cognitive* activities, including understanding,

recalling, and synthesizing the information contained in the text and his/her own prior knowledge; therefore, inferring, understanding what is not stated literally in the text, can also be classified as a cognitive strategy. Metacognitive theories are "systematized cognitive frameworks" (Schraw & Moshman, 1995, p.351). In the present study, self-explanation strategy, bridging, elaboration, prediction, and monitoring can be regarded as metacognitive strategies, considering *planning, monitoring,* and *evaluating* activities are involved, but they can also be classified differently.

Self-Explanation

Definition. Self-explanation is a process for explaining the meaning of a text while reading (McNamara, 2004) that has been recommended as an effective and promising technique to improve students' learning (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Whereas extensive studies in domain fields have established the effectiveness of self-explanation in comprehending instructional materials, it is seldom included in reading strategy studies in literacy and language teaching fields. Since *self-explanation* is scarcely studied as a reading strategy in the literatures related to literacy and L2, this chapter reviews it separately from other reading strategies. Mcnamara and Magliano (2009) have discussed the nature and process of self-explanation as below:

> The reading strategies that readers engage in while self- explaining or thinking aloud are heavily guided by metacognition. The act of self-explanation by its very nature requires the reader to be aware of the comprehension process. The process of self-explanation brings to the surface and externalizes the comprehension process for the reader. As such, it is a process that both requires and induces metacognition.

> Metacognition as it relates to reading involves the reader being aware of the reading process and knowing what to do when his/her level of comprehension is not sufficient, or does not satisfy his / her goals and desires.

According to Mcnamara and Magliano, and Chi, Leeuw, Chiu, and Lavancher (1994), self-explanation is a metacognitive learning technique or reading strategy which requires readers to employ other reading stategies to process the information, monitor their understanding, and attain the reading goal.

Self-explanation and think-aloud. Self-explanation is a learning strategy, while think-aloud is a research method, or a measure of thinking. Chi, Leeuw, Chiu, and Lavancher (1994) have argued that self-explanation is different from think-aloud(Chi, Leeuw, Chiu, & Lavancher, 1994)(Chi, Leeuw, Chiu, & Lavancher, 1994)(Chi, Leeuw, Chiu, & Lavancher, 1994)(Chi, Leeuw, Chiu, & Lavancher, 1994). Chi, Leeuw, Chiu, and Lavancher have explained the difference as follows: Traditional protocol collection is a process for participants to articulate the information passing through their short-term or active memory as they are solving a problem, like displaying problem-solving; self-explanation is a process of reflection, so participants are encouraged to reflect and infer what they are reading, thus constructing a mental structure. According to Vygotsky (2012), speech serves thinking, so prompted self-explanation theoretically improves comprehension and problem-solving. The purpose of self-explanation and think-aloud is different, but the format looks similar. Self-explanation quality improvement is one of the goals of this study.

Self-explanation and strategy training. Studies have shown that training in self-explanation and reading strategies promotes knowledge acquisition and problem-solving skills. For example, Bielaczyc, Pirolli, and Brown (1995) found that university students who received explicit strategy and self-explanation training showed significantly greater gains in the application of self-explanation and strategies as well as problem-solving performance in programming tasks than the control group, which received similar interventions without the explicit training. Researchers have also found that self-explanation can facilitate high order learning. In a self-explanation-eliciting study (the instruction given to participants was like a short one-time training) involving 24 eighth graders (L1), Chi, Leeuw, & Chiu, M, Lavancher (1994) found that the gains were greater for the prompted self-explanation group (32%) than for the unprompted group (22%) on the short-answer questions, especially on the more complex questions including generating new knowledge and understanding implications (22.6% versus 12.5%, $t(22)=2.64$, $p<.01$), which is important in facilitating higher order learning in the content area.

Moreover, researchers have also confirmed the effectiveness of self-explanation on reading comprehension. After studying the effects of providing reading strategy training with self-explanation to an experimental group of 21 university undergraduates with low prior domain knowledge, McNamara (2004) found that, compared with the 21 undergraduates assigned to the reading-aloud control group, the students who received the strategy training showed improved ability to comprehend the most difficult text, along with an improved quality of self-explanation. Mcnamara, O'Reilly, Best, and Ozuru (2006) also found that compared with the control group who received only a

presentation of self-explanation descriptions and examples via iSTART, explicitly teaching self-explanation and other reading strategies in the same environment for an average time of 1 hour and 44 minutes significantly improved adolescents' (L1) reading comprehension and self-explanation quality. So explicitly teaching self-explanation and strategies maximizes the reading comprehension effect.

Furthermore, Ozuru, Briner, Best, & McNamara (2010) found the contribution of self-explanation to the comprehension of low cohesive texts (lacking certain cues) is larger than its contribution to the comprehension of high cohesive texts, which indicates that self-explanation induces readers to actively bridge the cohesive gap between sentences and to draw inferences from texts. This finding indicates its potential for L2 learners, because a high cohesive text for an L1 reader may be a less cohesive text in a L2 reader's eyes, if he/she comes across linguistic problems such as unknown vocabulary. Thus, based on Ozuru, Briner, Best, & McNamara's finding that self-explanation is more effective with low cohesive texts, self-explanation seems to be a promising technique for L2 learners. Based on the extensive studies that have established the effectiveness of self-explanation and strategy training from various perspectives, there exists the possibility that self-explanation and strategy training may benefit L2 learners.

SERT Instructional Model

What is SERT? SERT is an instructional model that combines reading strategies with self-explanation, called Self-Explanation Reading Training (SERT). According to McNamara (2004),

> The reading strategies covered in SERT were included because their use is particularly characteristic of successful, skilled reading. The

strategies included are: monitoring comprehension, paraphrasing, predicting what the text will say, making bridging inferences to link separate ideas in the text, and elaborating by using prior knowledge and logic to understand the text. (p.2)

SERT was designed to address the reader's comprehension *activity*, one of the three elements (*reader, text, reading activity*) in the definition of *comprehension* by Snow and RAND Group (2002). Its purpose is to improve students' comprehension and learning by providing reading strategy instruction with self-explanation. In the table below, McNamara (2004, 2009) illustrates the strategy usage in the participants' self-explanation of the sentence "Mitosis guarantees that all the genetic information in the nuclear DNA of the parent cell will go to each daughter cell" in the text of *Cell Mitosis*.

Table 1: Examples of Strategies Used by Participants for Sentence 3 of Cell Mitosis

Strategy	Self-Explanation Examples (Two examples were provided for each strategy)
Comprehension monitoring	1. "I don't remember what DNA stands for." 2. "So I guess daughter cells are a part of a larger cell or came from a larger cell—I don't know."
Paraphrase	1. "So each daughter cell will receive a duplicate copy of the same strand of DNA from the parent cell." 2. "Ok through this process of mitosis all the genetic information belongs in the DNA of the parent cell and that is transferred over to the daughter cell."
Bridging	1. "So, yeah, so all the genetic information is in the chromosomes and each cell gets a complete set, so that's mitosis—when each cell has just as much DNA as the first mother cell—main cell—parent cell." 2. "So mitosis—the first stage of cell division where each set of chromosomes that goes to each daughter cell will contain DNA."

Table 1 cont.

Strategy	Self-Explanation Examples (Two examples were provided for each strategy)
Elaboration	1. "Ok so there's the daughter cell and then there's a parent cell—mitosis it has to do with genetic information so when I'm thinking of cell division I'm thinking of maybe how a baby is made and how it's developing." 2. "Ok what they're saying is that mitosis will make sure that an equal amounts of genetic information will go to each of the cells—equal amount will go to each daughter cell that way. They will develop basically the same—multiply the same."
Prediction	1. "Ok this is the separation of the cell—the DNA—the next one should be the RNA." 2. "So that's the first stage, now they'll give the second one."

Notes: This table is from Table 1 in McNamara (2004, p. 3; 2009, p.35).

There are many instructional models implemented in language arts or literacy classrooms, which have demonstrated significant effects in boosting students' reading ability (Trabasso & Bouchard, 2002). For example, in the K-W-L model, "teachers guide students to think about what they already know about a topic (K), what they want to learn (W), and what they learned as a result of their reading (L)" (Duffy, 2002, p.29). Through this model, readers make connections between their prior knowledge and what they are reading. Another example is the reciprocal teaching model, which teaches students to emulate a teacher's question-asking and combine the strategies of prediction, clarification, self-questioning, and summarizing in reading.

However, while research suggests that self-explanation is an effective reading strategy (e.g. Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Chi, Leeuw, Chiu, & Lavancher, 1994; Mcnamara, O'Reilly, Best, & Ozuru, 2006), studies in literacy or L2 field rarely explore the effect of the instructional model, SERT, combining self-

explanation with other reading strategies as a package in teaching. Actually there is little investigation into the effectiveness of self-explanation in literacy or language study at all.

Strategies included in SERT. The purpose of SERT is to improve the reader's self-explanation quality and comprehension ability by providing strategy instruction in combination with self-explanation. Research has shown that skilled readers have higher self-explanation quality than less skilled readers (Chi, Leeuw, Chiu, & Lavancher, 1994; McNamara, 2004). In this model, SERT combines reading strategies with self-explanation in order to train students to construct meaning by monitoring their own comprehension, bridging ideas in the text, connecting the text with their own prior knowledge, and making inferences. Self-explanation functions as an overarching reading approach, employing the five reading strategies to explain the meanings of sentences, paragraphs, and a passage: *monitoring*, *paraphrasing*, *prediction*, *bridging*, *elaboration*. self-explanation also serves as an instrument to measure the application of the five reading strategies.

The rationale for including self-explanation in the SERT model can be found in the above discussion on the positive effects of self-explanation in reading and learning under the section of self-explanation and Strategy Training. The selection of the specific five strategies included in SERT is due to the findings about the characteristics of successful and skilled readers (McNamara, 2004). Researchers have found, for example, that skilled readers closely *monitor* their understanding so as to be aware of what they do and do not understand through the reading process (e.g. Brown, 1982; Pressley, 2002a and 2002b). Based on this finding, the *monitoring* strategy is included in SERT. *Paraphrasing* is included in the model because it is a strategy for the reader to rephrase

the meaning of a sentence or a clause in the model, which is often the start of a self-explanation process (McNamara, 2009). Another difference between skilled and less skilled readers is that skilled readers are more likely to generate inferences (e.g. Hansen, & Pearson, 1983; Chi, Leeuw, & Chiu, M, Lavancher, 1994). Three strategies that help students generate various inferences are *prediction*, *bridging*, and *elaboration*. *Prediction* involves students drawing inferences about what might happen next in the text; *bridging* involves students connecting the ideas scattered in separate sentences and paragraphs of the text so as to make inferences; *elaboration* involves students connecting text content with their own prior knowledge to construct new content (inference) beyond the text.

With its emphasis on both monitoring and generating inferences, the SERT instructional model involves both cognitive and metacognitive strategies, if one accepts the general agreement that planning, monitoring, and evaluating are metacognitive activities (Jacobs & Paris, 1987; Kluwe, 1987). When students intentionally explain a sentence's meaning, they may employ various strategies, such as *paraphrasing* the meaning of the sentence, *monitoring* their own understanding of it, *elaborating* on its meaning, *bridging* the idea expressed in this sentence with the ideas contained in other sentences, or *predicting* the information or ideas to appear in the following sentences. These activities involve both cognitive and metacognitive strategies.

Effects of SERT. SERT has been conducted using both human instructors and animated conversational agents in iSTART and ISTART-ME. Extensive studies with different designs and participants have established its effectiveness in improving secondary and post-secondary school students' reading comprehension and self-explanation quality (e.g. McNamara, 2004, 2006; McNamara, Levinstein, & Boonthum,

2004; Best, Rowe, Ozuru, & McNamara 2005; McNamara & Magliano, 2009; Jackson, Dempsey, & McNamara, 2012; Jackson, & McNamara, 2013; Jackson, Varner, Boonthum-Denecke, & McNamara, 2013). Based on the findings related to the features of self-explanation and the five strategies SERT employs, an investigation of the effect of the SERT instructional model on L2 students' reading comprehension activity is promising.

SERT and L2. Willingham (2006) has argued that individual sentences do not pose a comprehension problem for a proficient L1 decoder, provided he/she knows the vocabulary and has sufficient knowledge of the subject matter. Relating sentences to one another is essential for reading comprehension. However, besides vocabulary and background knowledge, L2 learners need to have syntactic knowledge, and sometimes cultural background as well, to understand a single sentence. Understanding single sentences is as essential as relating sentences to each other. This poses an extra challenge for L2 readers because they need to overcome the linguistic and cultural problems to understand the meaning of single sentences, and then to further relate sentences to one another as L1 readers do is another challenge. Walter (2004) did find that intermediate L2 learners did poorly on the overall comprehension of texts, even though they understood individual sentences in L2 text. Due to these considerations, it is expected that the strategies included in SERT may help L2's reading comprehension. The participants in the present study were skilled readers in L1, but the literature casts doubts on the automatic transfer of L1 reading skills to L2 reading (e.g. Carrell, 1991; Clarke, 1980; Taillefer, 1996). Therefore, explicitly teaching the comprehension strategies in the SERT model may quickly boost these students' comprehension ability.

L2 Reading Comprehension

"A Reading Problem or a Language Problem?"

Alderson (1984) titled one of his book chapters with *Reading in a Foreign Language: A Reading Problem or a Language Problem?* "Reading Problem" refers to the lack of high-level cognitive activities, such as comparing, classifying, synthesizing, evaluating, and inferring; "Language Problem" refers to the lack of necessary linguistic knowledge and skills specific to L2, like phonological, lexical, syntactic, and discoursal knowledge. After reviewing the relevant research, Alderson and other researchers (e.g. Carrell 1991; Bossers 1991, 1992; Brisbois 1995; Taillefer 1996; Lee and Schallert 1997; Yamashita 1999) concluded that L2 reading involves both language and reading problems. Because L2 reading involves reading problems, the reading strategies widely used in L1 have been studied for their effectiveness in L2. In the below section of L2 reading strategies, these studies are reviewed. Additionally, the role of learners' L1 is an unavoidable variable in studying L2 reading comprehension, which is also discussed below. Because L2 reading is both a language problem and a reading problem, researchers have studied the relationships between L2 language proficiency, L1 ability, strategy training, and reading comprehension ability. However, the research is far from extensive. This present study investigated these variables in the course of the strategy training research.

Variables Related to L2 Comprehension

Linguistic factors. Many studies have found that L2 reading comprehension is correlated to L2 linguistic knowledge, especially vocabulary and grammar knowledge. Studies show that L2 readers draw heavily on their linguistic knowledge in reading

various L2 texts (Bossers, 1992; Cummins, 1980; Cziko, 1980; Horita, 1996, 2000; Nassaji, 2003; Tailefer, 1996). An example of such linguistic knowledge is syntactic awareness, which refers to the ability to understand the grammatical structures within a sentence (Tunmer & Hoover, 1992) as well as the ability to "reflect on the syntactic structure of language and regard it objectively and separately from the meaning conveyed by language" (Blackmore, Pratt, & Dewsbury, 1995, p. 405). The relationship between L2 syntactic awareness and reading comprehension has been well presented in empirical studies (e.g. Gelderen et al., 2003; Kirajima, 1997; Verhoeven, 1990). The reader's linguistic knowledge set the foundations for reading comprehension, but reading strategies can somewhat compensate for the linguistic knowledge insufficiency when the reader's linguistic knowledge reaches a certain level (e.g., threshold level).

L1 Transfer. As noted above, studies show that L2 readers draw heavily on their L2 linguistic ability in reading various L2 texts (Bossers, 1992; Cummins, 1980; Cziko, 1980; Horita, 1996, 2000; Nassaji, 2003; Tailefer, 1996). At the same time, their L2 learning is influenced by their L1 knowledge and skill. Goodman (1971) assumed that the reader's knowledge and skills in L1 can be transferred to L2 reading, and the metacognitive knowledge plays a large role in reading comprehension. However, researchers (e.g. Carrell, 1991; Clarke, 1980; Taillefer, 1996) have doubted the automatic transfer of L1 reading skills to L2 reading. Related to these studies are the linguistic threshold hypothesis and the linguistic interdependence hypothesis.

The linguistic threshold hypothesis proposes that the transfer of L1 reading ability happens when L2 proficiency, such as knowledge of vocabulary and grammar, reaches a certain threshold level (Cummins, 1979; Clarke, 1978, 1980). The main idea is that L2

learners are not able to read effectively until they reach some level of proficiency in the target language, at which their L1 skills can be transferred to L2 reading (Alderson, 1984). The threshold level may vary from task to task and from reader to reader, and it also may be the deciding factor in success or failure in L2 reading.

The linguistic interdependence hypothesis, in contrast, proposes that L1 reading knowledge and skills can be transferred to L2 reading (Goodman, 1973; Coady, 1979; Cummins, 1979, 1991). Very few studies have researched reading strategy transfer from L1 to L2 . Yamashita (2002) compared four groups (3 participants/group) of readers with different reading ability background in L1 (Japanese) and L2 (English) by way of thinking aloud their reading test taking processes. Each reading test was composed of multiple choice questions and gap-filling items. The author found that "readers tend to transfer their L1 reading strategies to their L2 reading", and "language independent strategies are more likely to be transferred from L1 to L2 than language dependent strategies." He also found that participants at various English proficiency levels reported less global strategies and more local strategies in L2 than in L1. Davis and Bistodeau (1993) also found that low L2 level readers used more bottom-up strategies in L2 and more top-down strategies in L1, although the advanced L2 level readers did not show a significance difference between L1 and L2. So metacognitive strategies training in L2 like monitoring, prediction or elaboration involved in SERT may be effective for the college level L2 learners.

Van Gelderen, et al. (2004) summarized the studies related to L1 transfer as "most findings have shown that L2 reading is modestly correlated with L1 reading but that the correlation between L2 reading and L2 knowledge is higher, especially for readers with

less advanced levels of L2 proficiency" (P. 20). Van Gelderen, et al. (2007) compared Dutch secondary school students' L1 (Dutch) and L2 (English) reading and found that in L1 reading comprehension, only metacognitive strategies made a significant contribution to the multiple regression analysis model, whereas in L2, both metacognitive strategies and vocabulary knowledge contributed significantly to L2 reading comprehension. When they added L1 reading comprehension ability to the L2 reading comprehension regression model, however, they found that the contribution of L1 reading comprehension and L2 vocabulary was significant and that metacognitive strategies were not significant any more. It may be interpreted that metacognitive strategies are correlated with L1 comprehension, so when the L1 comprehension variable is identified as a variance contributor, metacognitive strategies become indirectly correlated to L2 reading.

Although the topic of L1 transfer still needs further exploration, most researchers agree that L1 ability is transferred to L2 when the L2 learner's language proficiency reaches a certain level. Yamashita (2002) found that L1 reading ability and L2 proficiency complement each other to achieve the highest possible level of L2 reading comprehension. In the present study, the participants had already reached the language proficiency level required by the university's admission requirements, and were at the intermediate to high level based on their TOEFL scores. They can thus be considered to have reached the threshold level required for L1 transfer. Under such conditions, and taking the L1 transfer into account, this study explored the effectiveness of strategy training.

L2 Reading Strategies

The Relationship between L2 Strategies and Comprehension

Since L2 is still a reading problem, it is reasonable to conclude that intervening in high level cognitive activity like synthesizing and inferring should help L2 learners to improve their reading comprehension. The previous literature has shown the positive relationship between reading strategy usage and L2 reading comprehension (Hacquebord, 1989; Anderson, 1991; Ehrman & Oxford, 1990, 1995; Stevenson, Schoonen, & De Glopper, 2003; Yamashita, 2002; Fung, Wilkinson, & Moore, 2003; Zhang, 2010). Zhang's (2010) study on Chinese college students found that there was a strong positive correlation between metacognitive strategies and L2 reading comprehension. Likewise, Sheorey & Mokhtari (2001) found that college-level L2 students with high reading ability, like their L1 peers with high reading ability, reported higher usage of cognitive and metacognitive reading strategies. Fitzgerald (1995) also reported that more proficient ESL readers used more various metacognitive and cognitive strategies. To further study the relationship between different factors and reading comprehension, Nergis (2012) studied college L2 students and found that reading strategies were a significant predictor of academic reading comprehension in a multiple regression model. These studies have established that training L2 students in cognitive and metacognitive strategies is liable to improve their reading comprehension ability and further improve their academic achievement.

L2 Strategy Training Effect

In a meta-analysis to examine the effect of explicit reading strategy training on L2 reading comprehension, Taylor, Stevens, and Asher (2006) found only 21 published and

unpublished studies[2] that met their review criteria, specifically including a variable of reading comprehension and a test of explicit reading strategy training effect versus no explicit reading strategy training effect. The overall effect size was .54, indicating 68% of the students receiving strategy training exceeded the comprehension of the students who did not receive the strategy training. The moderating variables which had statistically significant effects were the overall length of the texts employed for the post-tests, readers' proficiency level and age. Test texts of 801 or more words had the biggest effect size; adult readers got the biggest effect size; the participants in their second or third year and beyond had a bigger effect size than the first-year participants. The effects of these moderating variables indicate that the college level L2 learners may gain a large effect size through strategy training.

Among these 21 studies, 12 featured some metacognitive strategy training with an effect size of .41. Given the limited number of samples, it is hard to reach a conclusion for both metacognitive and cognitive strategy training effects. For example, Carrell, Pharis, and Liberto (1989) found that metacognitive strategy training improved L2 learners' scores in an open-ended recall protocol, but not on multiple-choice questions, after four days training. The positive effect of strategy training in L2 was discovered in other studies. Fung, Wilkson, & Moore (2003) discovered that participants improved their comprehension monitoring ability in both L1 and L2 reading, after exposed to explicitly reciprocal teaching approach. They also found that the students could transfer

[2] 23 unique outcomes were used in this meta-analysis, because a researcher reported multiple experiments in the same publication. Taylor, et al (2006) calculated the included number of studies as 23.

"their newly acquired comprehension foster and monitoring strategies to a novel task" (p. 26). However, Barnett's (1988) study did not find any significant difference between the experimental and control groups in comprehension recall protocols after one semester of various kinds of reading strategy training.

Because of different research designs with different comprehension-measuring instruments, findings on the effect of strategy training on comprehension gains have been inconclusive (Yamashita, 2002). Grabe (2009) and Brantmeier (2002) have called for more studies in order to establish consistent results and reach generalizability about L2 reading strategy research. The present study answers this call and contributes to the literature of strategy training on L2 learners.

Self-Explanation in L2

As reviewed above, studies (e.g. Pirolli & Recker, 1994; Bielaczyc, Pirolli, & Brown, 1995; McNamara, 2004; Mayer & Johnson, 2010; Jackson & McNamara, 2013) have shown that self-explanation and reading strategy training significantly contribute to college students' comprehension of instructional materials and related problem-solving performance. However, few researchers have studied the effect of self-explanation and strategy training on college-level L2 learners' comprehension of instructional materials and related problem-solving performance. This study is not a meta-analysis study, but despite searches of "self-explanation," "think-aloud", "verbalization," "L2," "second language," "ESL," "ELL," "NNES," "language," "training," and "teaching" in Google Scholar and the electronic databases of humanities, science, social sciences, language and literature, and education, including ERIC, Proquest, EBSChost, JSTOR, Science Direct, and Sage Premier, very few studies were found that specifically focused on L2 students

and self-explanation, whereas many studies about self-explanation have been published in the science, math, and engineering fields (The majority of the participants were L1 students). That self-explanation is scarcely explored in L2 or literacy studies is probably related to the goal that self-explanation in content area is for improving problem-solving skills by self-explaining reading materials. The role of deep comprehension as the result of self-explanation and as the cause of improving problem-solving ability may not have attained the attention of literacy or language study researchers. The search did find one study by Wylie, Koedinger, and Mitamura (2009) about an experiment on the learning of English articles ("a," "an," and "the") within an intelligent tutoring system. The study found that students with self-explanation tutoring did not perform better than those with no self-explanation tutoring. The authors' tentative argument was that more practice opportunities were better in learning English articles than more reflective instructional practices. In order to learn the usage of these articles, it might not benefit so much by self-explaining why an article is used in a sentence or situation because it is hard to infer a definite usage rule of the articles. The current study is not about learning the English language, but about teaching L2 learners to use strategies and self-explanation to learn the same instructional materials as their L1 peers use in university classrooms. The search has not turned up any publications on this specific topic. These L2 learners in this study were proficient L1 learners, so explicitly teaching comprehension strategies and self-explanation may help their overall comprehension and improve their academic performance, considering that Walter (2004) found that intermediate L2 learners did poorly on the overall comprehension of texts, although they understood sentences in L2 text.

Technology in Reading Strategy Training

Traditional reading strategy training is conducted in a traditional classroom in which a large number of students may be taught by one instructor. At the other end of the spectrum is one-on-one tutoring. Although sometimes an unqualified human tutor in an academic setting may do more harm than good (Carlson, 1985), Glass, Cohen, Smith, and Filby (1982) have reported that the mean effect size for randomized studies of one-one-one adult tutoring was .62, and Cohen, Kulik, and Kulik (1982) have reported the mean effect size of learning gains was .40 in a meta-analysis. These effect size indices indicate a medium to large effect. However, one-on-one tutoring alone requires relatively more human resources and time than can be easily provided. McCardle, Chhabra, and Kapinus (2008) pointed a few issues demanding additional research to answer, including how technology can effectively be used to improve the reading comprehension. In recent years machine tutoring technology has made great progress, and intelligent tutoring systems have been playing an active role in the strategy training literature.

Intelligent Tutoring Systems

Graesser, Conley, and Olney (2012) have described intelligent tutoring systems (ITSs) as "computerized environments that incorporate computational models from the cognitive sciences, learning sciences, computational linguistics, artificial intelligence, mathematics, and other fields" (p.451). An intelligent tutoring system is commonly composed of these four models: the domain model, the student model, the tutoring model, and the user interface model (Nwana,1990; Freedman, 2000). Based on information collected about expert knowledge in the domain model and students' actions in the student model, the tutoring model decides tutoring strategies and actions, and the

interface model implements the actions made by the tutoring system (Nkambou, Mizoguchi, & Bourdeau, 2010). Through this mechanism, an ITS may track students' subject knowledge, emotions, and other attributes, so as to adaptively respond to each individual's performance (Graesser, Conley, & Olney), as is done by the Cognitive Tutor without an animated conversational agent and animated Tutors with an animated conversational agent. Studies have shown significant learning gains with these ITSs (e.g. Corbett, 2001; Ritter et al.2007; Nye, Graesser, & Hu, 2014). One ITS that uses a animated conversational agent is the web-based Interactive Strategy Trainer for Active Reading and Thinking (iSTART), which, along with its motivation-enhanced version, iSTART-ME, has trained thousands of secondary and postsecondary students in reading strategies, with significant learning gains. The present study specifically tests the effectiveness of iSTART-ME on L2 learners.

iSTART

iSTART uses animated conversational agents to tutor reading strategies, provides feedback on each user's performance in natural language, and guides each user to acquire strategies through self-explanation practices in order to improve the user's reading comprehension. iSTART has been continually developing new versions since the principal investigator, Dr. Danielle McNamara (Faculty at University of Memphis before 2011 and Faculty at Arizona State University from 2011 to present), received several grants from 2001 to present. As discussed above, researchers argue that self-explanation and strategy application can improve reading comprehension, because they help to organize and assimilate new knowledge to the learner's existing cognitive structure (Anderson, 1985).

iSTART has evolved continuously, and the most updated version is iSTART-ME. iSTART is composed of three modules: Introduction, Demonstration, and Practice. The Introduction Module is composed of videos introducing the concepts of self-explanation as well as the five reading strategies: monitoring, paraphrasing, prediction, elaboration, and bridging. These strategies are defined above under SERT.

The Demonstration Module provides more self-explanation examples to demonstrate how to "self-explain text and to use metacognitive reading strategies that improve self-explanation" (O'Reilly, Sinclair, & McNamara, 2004, p.174). Through the trials-and-errors of the student agent, Genie, the trainee is expected to learn how to improve a self-explanation by using the strategies.

In the Practice Module, Merlin, the instructor agent, reads each sentence appearing on the screen, and asks the trainee to type a self-explanation. Merlin gives feedback to the trainee for improvement. Once the self-explanation is satisfactory, Merlin asks the trainee to identify what strategy was used. At the end Merlin provides feedback.

iSTART has demonstrated its effectiveness in improving users' reading comprehension in their native language of English by providing reading strategy scaffolding through self-explanation (e.g. McNamara, O'Reilly, Best, & Ozuru, 2006, Jackson, Boonthum, & McNamara, 2010).

iSTART-ME

iSTART-Motivationally Enhanced (iSTART-ME) is an extension of the original iSTART program, and provides secondary and post-secondary students with reading strategy training and self-explanation training to better understand challenging science texts (Jackson & McNamara, 2011). Although all students benefited by studying with

iSTART (Jackson, Boonthum, & McNamara, 2010), the initial rather static iSTART practice became tedious over time for some students, as a normal ITS does (Brunelle, et al., 2010). iSTART-ME incorporates educational games for users to extensively practice the strategies and self-explanation, aiming to motivate the users to stick to the system and do enough practice. In other words, the motivational component of iSTART-ME, the games for extended practice instead of only coached practices as in iSTART, has increased the participants' motivation (Jackson & McNamara, 2013).

ISTART-ME also includes three modules: Introduction, Demonstration, and Practice. The Introduction Module is composed of a video introducing the concept of self-explanation, as well as individual videos on each of these five different reading strategies: monitoring, paraphrasing, prediction, elaboration, and bridging, followed by a summary video. At the end of each video are practice questions to ensure that the participant has understood the concepts and can identify the strategies in application. The Coached Practice is the initial practice of the Practice module. See Figure 2. ISTART-ME also includes extended practice composed of Coached Practice and video games designed for participants to practice the self-explanation approach employing the five reading strategies. The Introduction module, Demonstration module, and initial practice in this study took two hours, and the extended practice took one hour. However, the content can be adjusted based on different research purposes, and the total time can vary.

The Demonstration Module provides more self-explanation examples to demonstrate how to combine these reading strategies to improve self-explanation and comprehension of a science text. After leaving time for the participant to read through a paragraph, the virtual instructor first asks the participant to consider how to self-explain

one selected sentence. After a pause, the virtual instructor (animated conversational agent) gives his own self-explanation. He also asks the participant what strategies he or she used, but does not require the participant to answer. After a short pause, he elaborates the strategies he used as well as provides a definition for each strategy.

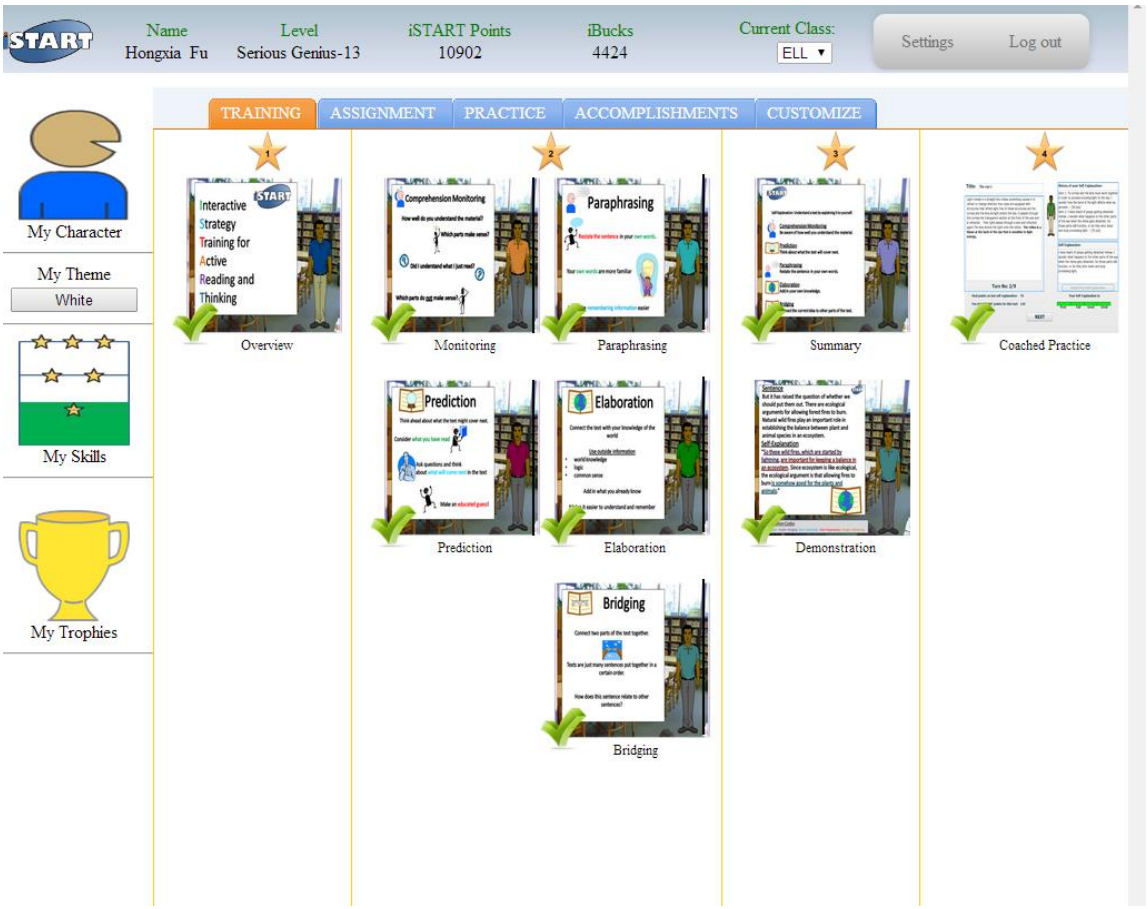


Figure 2: Screenshot of the Initial Interface of iSTART-ME

The Practice Module is a separate module involving interactions between the user and program with immediate feedback. The initial practice session is an instructed practice exercise that allows the participant to choose texts from the built-in library or to use a suggested text taken from a middle school science textbook. After submitting an self-explanation for each highlighted sentence, the user sees his or her performance rated

as poor, fair, good, or great. If the self-explanation is not rated highly enough, the system prompts the user to read previous sentences and get more information to improve his or her self-explanation and comprehension. The extended practice includes several games. The instructor or researcher can decide what games are available for his/her students. The extended game practice in iSTART-ME is designed to increase engagement, while practicing self-explanation and reading strategies.

iSTART-ME Training Effect

Jackson and McNamara (2011) found that their participants (10), college students, enjoyed their interactions with iSTART-ME and their boredom decreased across seven sessions. To further examine the effect of iSTART-ME, Jackson, Boonthum-Denecke, McNamara (2012) compared the performance of 125 high school students assigned to iSTART, iSTART-ME, or a control group, and found that the students assigned to iSTART or iSTART-ME performed better than the control group. Additionally, the less skilled students tended to gain more from iSTART-ME than the skilled students. To study the motivational component of the system, Jackson and McNamara (2013) examined the motivation and learning of 84 high school students studying with iSTART or iSTART-ME for eight 1-hour sessions and found that while they demonstrated equivalent target task performance, students with iSTART-ME exhibited significant higher levels of motivation and enjoyment. iSTART-ME has successfully demonstrated its effectiveness in improving L1 users' ability to explain difficult science texts and increase the engagement level of participants. However, the potential of iSTART-ME in tutoring L2 users is unknown.

iSTART-ME and L2

While iSTART-ME has consistently demonstrated its effectiveness in tutoring L1 users' reading strategies and self-explanation, can iSTART-ME be a good program for training L2 users? International L2 students admitted to American higher education institutions are supposed to have adequate English skills in listening, speaking, reading, and writing to do academic communications, although incoming international students are facing many academic challenges, including reading comprehension. Therefore, they should have the necessary basic reading ability for interacting with iSTART-ME. Additionally, Horiba (1996) found that L1 readers pay more attention to higher level processes such as connecting previous knowledge and generating inferences than their L2 counterparts, who devote more attention to lower level processes such as word identification. Because iSTART-ME targets higher level reading processes, it is assumed that iSTART-ME would work effectively in improving L2 readers' comprehension, for example, by learning the strategies to bridge sentences, elaborate knowledge, and make inferences, under the premise that they have had the necessary English skills to interact with iSTART-ME

Current Study

While international L2 students studying in American higher education institutions are proficient L1 learners, explicitly teaching self-explanation skills and other five strategies using iSTART-ME may help their overall comprehension, given the doubts expressed regarding the automatic transfer of L1 reading skills to L2 reading (e.g. Carrell, 1991; Clarke, 1980; Taillefer, 1996). However, it remains unknown whether these strategies and self-explanation training would indeed improve their reading comprehension, considering their L1 literacy background and intermediate English

linguistic ability. It is also uncertain whether iSTART-ME would successfully motivate L2 users to stick with the system long enough to finish the training, or even make them enjoy the learning process, and tend to apply the strategies in their future reading. This system may pose reading comprehension and writing challenges for participants, because they are requested to write down their self-explanations to each sentence selected by the system for them to explain after reading these sentences. Furthermore, if they do gain significantly from the system, how much they benefit and what factors contribute to the gains are both questions worth exploring.

In response to the rather limited studies on L2 strategy and self-explanation research and the advantages of ITS, this study will study the effectiveness of iSTART-ME in tutoring incoming international L2 students to American higher institutions to learn and apply reading strategies and self-explanation in science reading. Their reading comprehension improvement will be examined along with their self-explanation quality change, factors related to comprehension and self-explanation quality change, and their perceptions of the learning experience. So far, few studies have researched the possibility of tutoring L2 reading comprehension of instructional materials with ITSs, so this research provides information for understanding not only the role of reading strategy and self-explanation approach in L2 reading, but also the potential for using ITS to provide reading support to incoming international L2 university students as well as other L2 students in need.

CHAPTER III

METHODOLOGY

This interventional study employed a mixed research methods approach, including surveys, within-subject pretest-posttests, and interviews. This chapter is organized into the following major sections: 1) research rationale, 2) participants, 3) materials, 4) study procedures, 5) data analysis, 6) limitations, and 7) summary.

Mixed Research Methods Rationale

Different research approaches are related to different knowledge claims, worldviews, or paradigms. The quantitative approach mainly uses post-positivist claims for developing new knowledge; the qualitative approach primarily uses a constructivist or advocacy/participatory perspective to acquire knowledge; the mixed methods approach uses pragmatic knowledge to understand research problems (Creswell & Clark, 2011). The mixed methods approach is pluralistic, and emphasizes what works for the problems under study. The research questions (RQ) for this study required a mixed methods approach. Specifically, three research methodologies were adopted: the within-subject pretest-posttest method (RQ 1 and RQ 3), the survey method (RQ 2), and the interview method (RQ 2). The following section describes each method with its associated research questions, explanations for the method selection, and validity concerns.

Within-Subject Pretest and Posttest Design

Design rationale. As Creswell (2003) suggests, "The basic intent of an experiment is to test the impact of a treatment (or an intervention) on an outcome" (p.154). The largest part of this study was the within-subject pretest-posttest design, and was designed to examine the effect of iSTART-ME intervention on reading comprehension, self-explanations, and learning motivation. Specifically, this research method was chosen to answer the following major questions:

RQ 1. What are the effects of iSTART–ME on L2 reading comprehension and self-explanation?

1.1.Do participants improve their reading comprehension after the training?

1.2.Do participants improve their self-explanation quality after the training?

1.3.Do participants write longer self-explanations after the training?

RQ3. What factors account for changes in reading comprehension and self-explanation?

3.1. Is there correlation between self-explanation word count change and self-explanation score change?

3.2. Is motivation change correlated with participant's self-explanation quality change?

3.3. Is pre-vocabulary score correlated with participant's self-explanation quality change?

While an experimental design with randomly assigned participants to control and treatment groups could improve power and test causal relationships, there were several practical and methodological impediments to adopting an experimental design. First, the

question of content equivalency is hard, if not impossible, to solve. What would constitute "typical" non-treatment group versus a treatment group? For example, unguided practice with similar reading content on a computer would be comparing strategy to non-strategy training. Moreover, it is hard to control variables equivalent to those in iSTART-ME environment, such as to control the amount of reading materials equivalent to iSTART-ME or the equivalent self-study time. For example, if the materials were equivalent, the control group students would need less time to finish reading the material than the experimental group; if the time were equal, the control group would have more materials to read than the experimental group. Accounting for equivalence in the game portion of the strategy training was likewise problematic. Would it be best to provide different games from iSTART-ME to the control group or no games? How would these games be chosen for equivalence to the games in iSTART-ME? The games in iSTART-ME focused on strategy training, making it necessary to account for both games and strategies for equivalence. Based on their research purposes, the studies using experimental design related to iSTART-ME or self-explanation have various control groups. For example, the control group read the texts twice; the control group read aloud the texts; the control group was provided self-explanation instructions without prompts; and the control group used iSTART. All these kinds of control group design cannot help to answer the research questions that this study intends to answer.

Another potential experimental design would have been to use strategy training with a human tutor versus the iSTART-ME intervention. While this might be considered equivalent in terms of access to a tutor, this is not the research question or issue that this study intended to research. Many other studies have examined whether human and

artificial tutors are equivalent and require their own methodology; it would not be possible to test that in addition to the research questions under study here. Research has shown that ITS, while effective, are always less so than well-trained human tutors (Grasser, 2007). Yet, because well-trained human tutors are not readily accessible in sufficient numbers to meet the needs of ELLs worldwide, even less-effective tutoring can be beneficial because it can be available to anyone with Internet access. iSTART-ME can meet this need and has been shown to be effective in promoting reading comprehension through strategy training, yet it has not been tested with ELLs. The purpose of this study was to examine iSTART-ME's ability to deliver comparable results for ELL learners, not to establish whether iSTART-ME is or is not as good as human tutors.

Therefore, although the two different experimental designs referenced above are of value, I argue that they are subsequent to this initial study, which focuses on how the effects of strategy training via iSTART-ME does or does not transfer to incoming international L2 students admitted to American higher institutions, a special group of L2 learners.

As an initial study for iSTART-ME intervention on international college students, this study also intended to examine factors which might affect participants' learning with iSTART-ME and which might be unique to this population and thus provide guidance for improving iSTART-ME for future international students. Adopting this focus also solved a practical problem, which was how to get enough participants for this study, which was extremely labor-intensive and time-consuming for participants. The testing and intervention required five hours per participant. The within-subject pretest-posttest design directly increases power and allows a reduction in the number of subjects studied,

according to the relevant statistic formula which is not detailed here. This can be tested by using any online sample size calculation tool. Based on an online sample size calculation tool on Harvard University website, Statistical Considerations for Clinical Trials and Scientific Experiments, this design required a minimum sample size of 27, provided Type I error level α=.05, Type II error level β=.2, and effect size Cohen's d= .8. Of course, this effect size should be big enough to be discovered with a sample less than 30 participants, which were the least number of participants that this study would recruit, but as it was anticipated that the effect size could be smaller, it was desirable to increase the power of this study and to recruit as many participants as possible.

Validity threats. One of the threats to the validity of within-subject pretest-posttest design is the confounding problem of carryover caused by history. This is a phenomenon which concerns the "noise" caused by intervening variables not related to the intervention or training between pretest and posttest. However, in this study the participants finished the pretest and posttests in about five hours, during two consecutive days just before the semester began. It is unlikely that two non-school days would produce learning experience related to strategy training, self-explanation, or reading comprehension that would carry over to their performance in the posttest. The immediate posttest to measure short-term learning in about 3 hours was the goal of this study, so it was felt that carryover from history would not be a confounding problem as in other within-subject design. It can also be argued that an intermediate to high level second language speaker cannot improve his/her reading comprehension within such a short time (less than 30 hours, including necessary sleep time, detailed in the Procedure section), without strategy intervention, so a within-subject design was considered to suffice.

Additionally, the time schedule, namely, 1-hour pretest and presurvey, 1-hour posttest and postsurvey, and 3-hour training, was decided in consideration of the feasibility for recruiting incoming international L2 students to participate in this study. It was also decided according to the feedback and results of a pilot study with six Chinese students studying in English-speaking universities who volunteered to pilot the design. They were a convenient sample composed of the researcher's friends, 3 graduate students and 3 undergraduates.

Survey Design

Design rationale. A survey methodology is used to describe trends, attitudes, or opinions in a quantitative or numeric way (Creswell, 2011). The survey method as the second strand in this research aimed to provide a numeric description of the participants' demographic features, motivations to study with iSTART-ME, their individual metacognitive strategy application in reading before the training intervention, and their attitudes towards the iSTART-ME after the training. These were part of the data collected to study the factors that might be related to the interventional results. Together with the interview design and the intervention, this research design intended to mainly answer the following question:

2.2. Did participants hold a positive attitude towards iSTART-ME?

3.2. Is motivation change correlated with participant's self-explanation quality change?

3.4. Are prior metacognitive strategies related to self-explanation score change?

3.5. Are prior metacognitive strategies related to reading comprehension scores?

3.6.What individual characteristics predict self-explanation gain scores?

3.7.What individual characteristics predict short answer comprehension gain scores?

Survey research design is used to describe the attitudes, opinions, behaviors, or characteristics of the population based on data collected from a sample or a population. In this study, the purpose was to describe the sample's attitudes towards and perceptions of their learning process, behavior, and personal experience, in addition to collecting the sample's democratic features. Specifically, 1) using this method the sample's learning motivation level prior and post to the iSTART-ME intervention was collected to study if their motivation change through the intervention. A common way to reflect a person's learning motivation is by the use of a good survey. In this study, the Motivated Strategies for Learning Questionnaire (MSLQ) was selected. Details about the survey were described in the Materials section. 2) Using adapted Metacognitive Awareness of Reading Strategies Inventory (MARSI) to collected the metacognitive strategies that the participant employed in reading the expository text in the pre-intervention phase. Details about the survey are described in the Materials section. 3) Using Adapted Attitude toward Tutoring Agent Scale (ATTAS) to collect the participants' perceptions of their experience with the animated conversational agent built in iSTART-ME after the intervention. Details about the survey are described in the Materials section. It is believed those participants' attitudes towards iSTART-ME and perceptions of their learning process, performance, and results would be related to their learning effect.

Error control. This method's primary interest is to survey the participants' perceptions and attitudes to reflect the sample's characteristics and learning experiences which might be related to their learning effects. Factors that may limit survey design

include coverage error, sampling error, measurement error, and non-response error. All of the participants who completed the study did the surveys, so coverage and sampling errors were not a concern in this study. Measurement error may originate from the respondent, the mode of data collection, the data collection instrument, or the interviewer (Groves, 2004). In this study, all the participants were supposed to be good at typing, because they took the TOEFL test online. Additionally, the surveys only requested them to select the number for each survey item. Moreover, before the study, all the computers were tested and set on the right instrument page for participants to take the survey. Therefore, the error related to mode of data collection was controlled. Additionally, the data collection instruments were slightly adapted from established instruments, and their reliability indices in this study were specifically calculated and evaluated. The three survey instruments were valid and reliable tools based on the published literature, which is discussed in the material section, so the error related to instrument was controlled. Regarding to the respondent error, the author emphasized orally before the participants took the survey that all the information collected was for collective analysis, and nobody would be reported individually. It was also emphasized that all their answers were confidential, and honest sincere answers were highly valued. The informed consent letter was also presented to the participants. Some of the words in instruments were adapted to the actions that participants just completed, such as pre metacognitive strategy survey and post motivation survey. Details are shown in the Materials section of this chapter. In this way, the error related to the respondent would have been minimized. Moreover, the survey results were not used to be generalized to a different or broader population. All the results were described as the sample's characteristic. Non-response error was not a big

concern of this study, because very few participants missed one or two items unsystematically. Missing data and their treatment were described in the data analysis section.

Interview Design

Design rationale. Qualitative research attempts to understand phenomena from participants' subjective views; in other words, to understand individual's personal interpretations of a concept or phenomenon. Interviewing as a qualitative research method is designed to elicit extended responses and gain insight. Through interviews researchers can understand things from the interviewees' perspective and discover the meaning of their experiences. Interviewing, as other qualitative research methods, helps researchers to understand a phenomenon, a concept, a setting, or a context so as to create an agenda for change or reform. This process may help to improve iSTART-ME design to better serve international university students. Additionally, it created an outlet for participants to reflect and express their feelings and ideas, which may enhance their learning gains as well.

The interview design in this study was designed to answer research question 2:

RQ2. What are participants' perceptions of their learning experience with iSTART-ME?

Specifically, the goal was to uncover the participants' voices, feelings, and ideas regarding their learning experiences and results with iSTART-ME, such as their learning gains, future application plans, complaints, suggestions, etc. This design also served to help triangulate the MSLQ and ATTAS survey results and pretest - posttests comparison results (comparison and self-explanation score change, gain scores). The results were also

designed to help improve the design of iSTART-ME. The interview protocol is provided in the Materials section of this Chapter.

Subject bias control. In the interview process, the researcher attempted to ask neutral questions without showing any favor for any answers. While there is promise in using technologies, including iSTART-ME, to improve L2 language skills, there are also significant limitations of technology such as the lack of flexibility to answer users' questions. To make the ITS design more effective, it is just as important to have the feedback from users, educators, instructional designers, and subject matter experts as including significant technology breakthroughs. Based on the research, I, the interviewer, conjectured that reading strategy training and self-explanation might be good for readers, which might boost readers' reading skills, especially when readers came across challenging materials or sentences; however, no evidence has been collected for L2 learners. Therefore, I held a neutral positon, expectation, or attitude towards the strategy training intervention. Finally, it was not at all clear that the 3-hour training through iSTART-ME would be enough for L2 students to learn these strategies. Therefore, the researcher held a sincere desire to learn from the users' feedback without preference for one set of answers or another, and thus the potential for researcher subject bias could be considered low. Moreover, the findings of these interview data allowed for triangulation with quantitative data, including the data collected from the learning motivation survey (MSLQ), the within subject pre and post t-test , and the attitude towards iSTART-ME conversational agent survey (ATTAS), thus strengthening the design.

Participants

Recruitment

The recruitment targeted at incoming international students at a major university in the Southwestern United States, which had a total of 73,378 enrolled undergraduates and graduates in the fall 2012, including 59,382 undergraduates and 13,996 graduates[3] on four different campuses. Among them, 6,645 were international students from 120 countries; a number which places the institution in the top 11th positon among the American universities hosting international students in the 2012-1013 academic year (The Institute of International Education, 2014). The study recruitment was a continuous process beginning on August 1 and continuing through August 30 (one day before the last training sessions), on the most populated places of the four campuses. The training sessions were held on August 20 and 21, 2013, right before the fall semester starting date of August 23rd, and again on August 30 and September 1, which was the weekend following the first week of classes. This study schedule was set at the beginning of the semester because the aim was to reach incoming international students before they started their formal studies.

The recruitment was done in a variety of ways. First, flyers about this study were posted on public notification boards and distributed in dining halls, the Students Service Building, and student dormitories. Secondly, after the author introduced the potential of iSTART-ME for international students to improve their reading and writing abilities, she got permission from the school orientation organizer to set up a recruitment table when the new international student orientation was conducted one week before the semester

[3] From the university's Office of Institutional Analysis. The fall 2013 data were not available at the writing time, but the university news report showed the total registered students increased .

started. During that orientation break, and before and after the orientation, the author recruited volunteers from the orientation attendees, with one of her colleagues as her assistant.

In addition, research overview and recruitment forms were emailed to 39 instructors of English 101 who taught at the campus. English 101 was a first-year composition course in which both international and non-international freshmen registered, totaling more than 100 sessions on four campuses. Fourthly, the author went to dining halls to personally recruit incoming international students with her son's assistantship.

Sample

Because the incoming international students had a very busy schedule due to various orientations, class preparation, and other tasks associated with preparing for their first semester on campus, this study recruited 118 students in total to offset the anticipated attrition rate. Fifty-six students appeared in the lab for their first sessions; the rest excused themselves because of schedule conflicts. Among those who attended the first session, 36 students attended the second session, in other words, completed the entire training process. The other 20 students did not show up in the second session with a notice of conflicted schedule or without any notice. Because two of the 36 students did not complete most of the post-intervention tasks, this study excluded these two students' data, using only the remaining 34 students' data for the analysis. Of these 34, 11 (32.4%) were undergraduates and 23 (67.6%) were graduates; 20 (58.8%) were females, and 14 (41.2%) were males. Their ages ranged from 17 to 28. Except for 4 (11.8%) students

from Brasilia, Colombia, Iran, and Japan, the other (88.2%) students were from China. Their majors included engineering, business, science, economy, health, and education.

The participants for the interview were selected from these 34 students according to their registration sequence in the lab on day 2 and by selecting only the Chinese participants. The researcher purposely did not include participants from other countries in order to control the variables of L1 characteristics and educational background. Another reason for selecting only Chinese students was that the majority (88.2%) of participants were from China. All the Chinese participants were assigned a random number from one to three according to the time they arrived in the lab on day two. The persons assigned a "three" were targeted for interviews. All interviewees were interviewed for 10 minutes, with the exception of 3 interviewees, who participated and completed the study at almost the same time by chance. These participants were interviewed as a focus group for 25 minutes. In total, seven participants were interviewed individually and three were interviewed as a focus group.

Materials

This study was composed of these three phases: pre-intervention, intervention, and post-intervention. The intervention part was done within iSTART-ME system online; the other parts were done within *Qualtrics*, an online survey software. All the participants were expected to complete the whole process. The pre-intervention included pretests and pre-surveys. The pretests included a standardized reading test, a short-answer reading test, and a self-explanation test. Details are shown below. The pre-survey included a strategy usage survey, a motivation survey, and demographic information. The intervention was for participants to learn reading strategies from iSTART-ME through

self-explanation. iSTART-ME training consisted of introduction module, demonstration module, practice module. The post-intervention included posttests, post-surveys, and interviews. The post-tests included adapted PSAT test, a short-answer reading test, and a self-explanation test, corresponding to the pretests. The post-surveys included MSLQ survey and ATTAS survey. The semi-structured interviews included one-on-one interviews and a focus group interview. The following section describes the pre-interventional materials, interventional materials, and post-interventional materials including their origins and validity and reliability information. The reliability indices of these adapted surveys used in this study were provided in the Results Chapter.

Pre-Intervention Materials

Summary. Table 2 shows the materials used in the pre-intervention period. During the pre-intervention period, the pretest and pre-survey were conducted through *Qualtrics*.

Table 2: Materials Used in the Pre-Intervention Phase

Materials	Purpose
1. Adapted PSAT Critical Reading test, including vocabulary and reading comprehension	To assess prior reading comprehension ability
2. Short-answer reading comprehension test	To assess prior reading comprehension ability
3. Strategy application survey (MARSI)	To assess strategy application
4. Self-explanation test	To assess prior self-explanation ability
5. Demographics	To collect participants' demographic data

The pretests were used to measure participants' English vocabulary, reading comprehension, and self-explanation skills, including adapted PSAT test, short-answer reading comprehension test, and self-explanation test. All these tests had alternate versions in the post-intervention phase. The vocabulary and standardized reading comprehension tests were taken from the Critical Reading Sections of The Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) test 2011 Form S. The short-answer reading comprehension test and self-explanation test were provided by the Science of Learning and Educational Technology Lab (SoLet Lab) led by Danielle McNamara, which had been used in other studies related to iSTART-ME. These two texts were similar in terms of length, content difficulty (Flesch-Kincaid grade level 8–9), and linguistic features (Jackson, Varner, Boonthum-Denecke, & McNamara, 2013). These two types of comprehension tests were used in consideration that the effect of reading strategy training on L2 comprehension has not been equally significant in different types of tests. For example, Carrell, Pharis, and Liberto (1989) found that strategy training improved L2 learners' scores in open-ended recall protocol (a task scored for the number of ideas), not on multiple-choice questions; Barnett's (1988) study did not find any significant difference in the open-ended recall protocol between the experimental strategy training group and control group.

The pre-survey included reading strategy survey and learning motivation survey. The reading strategy survey was used to survey what metacognitive strategies the participants applied in doing the reading comprehension tests, which was adapted from Metacognitive Awareness of Reading Strategies Inventory (MARSI). The learning motivation survey was used to measure the participants' motivation for taking this

training program, which was adapted from MSLQ. This adampted MSLQ had an alternate version in the post-intervention phase. Below is a detailed description for each instrument.

The PSAT 2011 Form S. The PSAT/NMSQT is a standardized test that provides practice for the SAT as well as for American citizens to enter NMSC scholarship programs. International students usually do not take the PSAT. The PSAT measures critical reading skills, math problem-solving skills, and writing skills. Because the focus of this study was to test participants' vocabulary and reading comprehension skills, this study only used the two critical reading sections which focus on measuring these constructs.

The author was given permission from College Board to use PSAT 2011 and PSAT 2012 for the pre- and posttest, but the author had not received the PSAT 2012 at the time this research began. Considering the two critical reading sections were of the same structure, including sentence completions (vocabulary test) and passage-based reading, with almost the same number of words in each passage, the author used the first critical reading section of PSAT 2011 Form S for the pretest, and the second critical reading section for the posttest to compare students' performance. The two passages included in the pretest and posttest were expository texts; the third text in each section was a literature text with more than 500 words which was excluded from the pretest and posttest. See Table 3. The answers to the PSAT 2011 were found at the website of Answer Explanations.

Table 3: Words of Reading Comprehension Test

	Passage	Words
Pretest	1st Passage	110
	2nd Passage	88
Posttest	1st Passage	110
	2nd Passage	113

The College Board (2013) reported the reliability coefficient of the critical reading test as .88. Each critical reading section is composed of sentence completion and passage-based reading questions. In terms of the sentence completion questions and passage-based reading questions in the critical reading sections, the College Board explained as follows:

> Sentence completion questions require students to follow the logic of an idea expressed in a fairly complex sentence. Sentences are given with one or two words omitted. The correct answer is the word or set of words that, when placed in the blank(s), best fits the meaning of the sentence as a whole.
>
> These questions:
>
> - Test a student's ability to recognize logical relationships among elements of a sentence
> - Measure vocabulary in the context of the sentence

Passage-based reading questions measure students' ability to read, understand, and interpret reading passages. These passages are:

- Drawn from a variety of fields, including the humanities, social studies, and natural sciences. They may also be excerpted from works of fiction.
- Varied in style and may include narrative, argumentative, and expository elements
- About 100 to 850 words and will often include an introduction and/or footnotes (College Board, 2014)

Each of the critical reading sections of PSAT includes sentence completion, two short expository reading passages, and one long literary narrative reading passage. Because the main purpose of this research is to improve students' ability in reading expository passages, which represents the most frequent style used in the current academic fields of the study, the long literary narrative reading passage in each critical reading section was excluded.

In order to make the pre- and post-reading tests match, the last sentence completion question and one vocabulary question after the passage reading were removed from the first critical reading section as the pretest. Therefore, both the pretest and posttest included six vocabulary questions and three comprehension questions. The purpose of the tests was to compare students' vocabulary and comprehension before and after the iSTART-ME training.

Due to the confidentiality of the PSAT, the critical reading sections used in this study were not attached. Below is a sentence completion example, and "E" is the correct answer.

Hoping to -------the dispute, negotiators proposed a compromise that they felt would be ------- to both labor and management.

A. enforce …useful
B. end ... divisive
C. overcome ... unattractive
D. extend ... satisfactory
E. resolve ... acceptable

Short-answer reading comprehension test. The short-answer reading comprehension tests were provided by the Solet Lab and were used in the research of iSTART-ME group. The pretest passage titled Heart Disease has 305 words and 8 questions; the posttest passage titled Red Blood Cells has 282 words and 8 questions. See Appendix A and Appendix C. The detailed grading rubric was provided by the Solet Lab. These pre- and post-comprehension tests were used to test students' reading comprehension in reading science materials, beyond the multiple- choice reading comprehension tests. The purpose was to test students' reading comprehension skills in reading science materials before and after the iSTART-ME training. For each passage, half of the questions were bridging inference questions, and half of the questions were text-based questions. According to Mcnamara, O'Reilly, Best, and Ozuru (2006), "Bridging inference questions required the reader to bridge information across two or more sentences to form a correct answer. In contrast, the text-based questions could be correctly answered using information from a single sentence" (p. 154).

The detailed scoring rubrics were provided by the SoLet Lab. Two ELL language teachers scored the students' answers independently with the rubrics. These two graders

gave the same scores for each student, except for the pretest of one student. The difference came from the grading of the answers to one question. The two graders were asked to read the grading rubric on this question again and confirm their grading on this student. One grader changed her grading, and the final grades given by the two graders were the same. Therefore, the grading was of 100% agreement

Self-explanation test. The above described non-standardized reading comprehension passages were also used to test participants' self-explanation ability before and after the iSTART-ME training. After finishing the reading comprehension questions, the participants were asked to write down their self-explanations to four sentences in each passage. For students' reference, detailed directions with examples of good self-explanation were put at the beginning of this section. As described above, these two texts had similarity in terms of length, content difficulty (Flesch-Kincaid grade level 8–9), and linguistic features (Jackson et al., 2013). The sentences chosen for self-explanation in the pretest and posttest were similar in sentence length and linguistic features. See Appendix B and D. The purpose was to compare students' self-explanation skills before and after the iSTART-ME training.

***Assessment algorithm*.** All self-explanations were scored by the algorism built in iSTART-ME. Jackson, Guess, and McNamara (2010) explained the algorithm below:

> Several versions of the iSTART evaluation algorithm have been tested and validated with human performance (McNamara et al., 2007). The resulting algorithm utilizes a combination of both word-based approaches and latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007). The word-based approaches provide a more accurate picture of the lower level

explanations (ones that are irrelevant, or simply repeat the target sentence). They are able to provide a finer distinction between these groups than LSA. In contrast, LSA provides a more informative measure for the higher level and more complex explanations. Therefore, a combination of these approaches is used to calculate the final system evaluation. (p.130)

According to Jackson, Guess, and McNamara, the agreement between the algorithm built in iSTART-ME and human experts was significant, $\kappa = 0.646$, $p < .001$. The scores of self-explanation range from 0 to 4. A too short or irrelevant explanation was scored "0"; a sentence-based self-explanation relating to only the target sentence was scored "1"; a text-based incorporating part of text information beyond the target sentence was scored "2"; a global –based self-explanation incorporating world knowledge or relating to the overall theme of the whole text was scored "3". Table 4 shows the examples of self-explanation score categories.

Table 4: Examples of self-explanation categories

Score Category	Example
Target sentence	*''Energy-storing molecules are produced on the inner folds.''*
Irrelevant (0)	''Hello, I am a taco.''
Sentence-based (1)	''The molecules holding on to the energy are created on the inner folds.''
Text-based (2)	''These sentences say that the mitochondria's inner membrane produces energy-storing molecules.''
Global-based (3)	''The inner folds develop energy-storing molecules that help store more energy for the plant and help it grow, survive, and reproduce."

Notes: This table was adapted from the Table 1 of Jackson, Guess, & McNamara (2010, p. 132).

Adapted MARSI. The items on this survey were adapted from Mokhtari & Reichard's (2002) Metacognitive Awareness of Reading Strategies Inventory (MARSI).

See the Appendix E for this instrument. On the original MARSI there are 30 items. Among them, 13 are about global reading strategies; 8 are about problem solving strategies; 9 are about support reading strategies. The Likert scale ranges from 1 to 5, anchored by "I never or almost never do this" (1) to "I always or almost always do this" (5). According to Mokhtari & Reichard (2002), MARSI's reliability index Cronbach's alpha was .93, indicating a reliable measure. In this study 24 items were used as the options for participants to select. The other 6 items were not used, because they did not apply to the reading context in this study, such as discussion and reading aloud strategies.

The purpose of this survey was to investigate what strategies the participants had been using in their reading comprehension process before the iSTART-ME training. After completing the above mentioned comprehension tests provided by SoLet Lab, the participants were asked to select the strategies they used from the adapted MARSI survey when they did not understand a sentence, a section, or even an entire passage while reading those passages. It asked the participants to select all the strategies, (i.e., items on the adjusted adapted MARSI survey), that they had applied in their reading process. Because the use of strategies depends on readers' reading ability in English, the type of material read, and the purpose for reading it (Oxford 1990), the author changed the present verb tense on MARSI to the past tense so as to ask the students to recall how they applied these strategies in their just-finished reading comprehension process, instead of surveying their overall strategy usage, regardless of the reading materials.

Adapted MSLQ. MSLQ (Motivated Strategies for Learning Questionnaire) is a self-report instrument designed to measure college students' motivational orientations and self-regulated learning as related to a specific course (Pintrich, Smith, Garcia, &

McKeachie, 1991). The MSLQ had been used in at least 363 studies (Taylor, 2012). Students' motivation may change from course to course due to various factors like self-interest or course nature. The MSLQ consists of two broad categories of motivation and learning strategies. The learning strategies scales were not used in this study, because it surveyed what strategies a person uses in learning, separate from motivation. . The motivation scales include six subcategories. See Table 5. These strategies were not used because they did not fit the current study.

Table 5: Components of the MSLQ Motivation Scales

Motivation Scales	
Scale	# of Items (31 total)
Intrinsic Goal Orientation	4
Extrinsic Goal Orientation *(not used)*	4
Task Value	6
Control of Learning Beliefs	4 (discarded in analysis)
Self-Efficacy for Learning & Performance	8
Test Anxiety *(not used)*	5

This study used only four of these motivation subcategories. The subcategories of Test Anxiety and Extrinsic Goal Orientation were not applicable in this study context and were excluded. For example, an item under Test Anxiety subcategory is "I am so nervous during a test that I cannot remember facts;" an item under Extrinsic Goal Orientation subcategory is "The most important thing for me right now is improving my overall grade point average." At each session of this study, the participants were informed that this

study had nothing to do with their college grades, and their participation was simply to learn knowledge. They were asked not to regard any of the tasks as a test.

The selected subcategories were Intrinsic Goal Orientation, Task Value, Control of Learning Beliefs, and Self-Efficacy for Learning & Performance. According to the MSLQ Manual, the values of Cronbach Alpha of these four subcategories were .74, .90, .68, and .93 respectively. The Cronbach Alpha value of the adapted scale used in this study is presented in the Results Chapter. The Control of Learning Beliefs were discarded in the analysis, due to the low reliability of the construct in this study.

Considering the applicability of each item in this research context, three items from each selected subcategory were selected to be used in this study, totaling 12 items, which represented the different dimensions of each subcategory. To compare the motivation change before and after the iSTART-ME training, the adapted MSLQ were used in the presurvey and postsurvey, and the wording was changed to fit the precontext and postcontext. See Adapted MSLQ for Pre-Survey and Adapted MSLQ for Postsurvey in Appendix F and Appendix G.

Interventional Materials

The intervention included three modules: introduction, demonstration, and practice. Under introduction and practice modules, there were several components. See Table 6. All the intervention was implemented online in a computer lab. Details are described in the below Procedure section of this Chapter.

Table 6: Interventional Materials

Materials	Purpose
Introduction Module	
1. Overview	To introduce the concepts of self-explanation and reading strategies
2. Monitoring	To introduce comprehension monitoring strategy
3. Prediction	To introduce prediction strategy
4. Paraphrasing	To introduce paraphrasing strategy
5. Elaboration	To introduce elaboration strategy
6. Bridging	To introduce bridging strategy
7. Summary	To summarize the previous 5 strategies
8. Demonstration Module	To generate and discuss the quality of self-explanation
Practice Module	
9. Initial Coached Practice	To practice self-explanation with qualitative feedback
10. Coached Practice	To practice self-explanation with qualitative feedback
11. Bridge Builder Game	To practice strategy identification in the game
12. Showdown Game	To practice self-explanation in the competition
13. Balloon Bust	To practice strategy identification in Balloon Bust
14. Map Conquest Game	to practice self-explanation in Map Conquest

The introduction module included Overview, Monitoring, Paraphrasing, Prediction, Elaboration, Bridging, and Summary, which were instructional videos. At the end of each video were practice questions to ensure that the participants had understood the concepts and could identify the strategy in application. The participants were required to finish this section in sequence from left to right. When a video was playing, they could

play it back or pause it but could not speed it up. This was to ensure that all participants went through the same content training materials. At the top of the screen, participants could see their point and iBuck totals. Because this study only provided one hour for assigned extended practice, these points and iBucks were not usable by the participants. The extended practice module covered Coached Practice, Bridge Builder Game, Showdown Game, Balloon Bust Game, and Map Conquest Game. This following section sequentially described the introduction module, demonstration module, and practice module.

Overview video. For each participant, all the videos were locked except for the Overview in the beginning. Only after students completed one video, was another video unlocked. After the Overview was clicked, a virtual instructor in the video began to introduce self-explanation and the five reading strategies. See Figure 3 below and Figure 1 in Appendix I for the Overview.

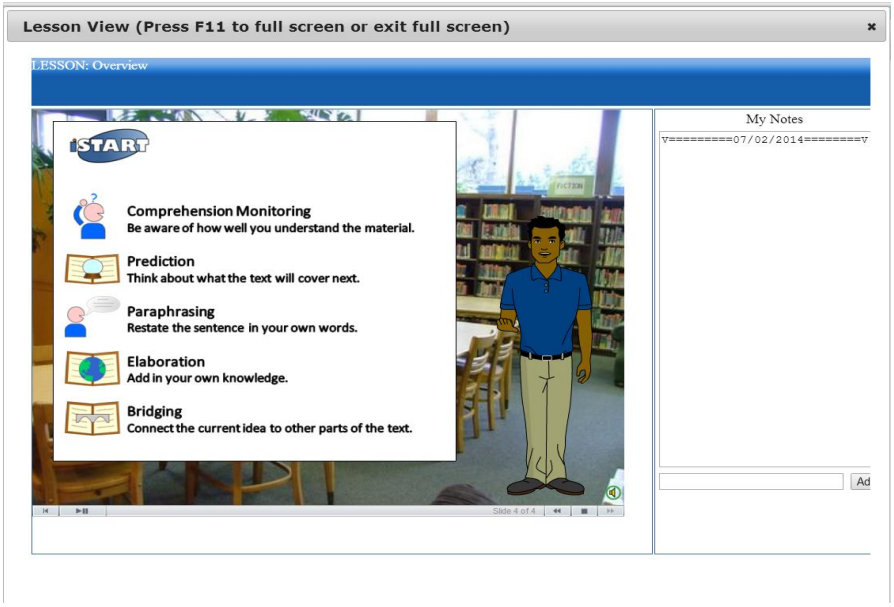


Figure 3: Screenshot of Overview

Monitoring video. After finishing the Overview video, the next video Monitoring was unlocked. The video began with an introduction to how to monitor one's comprehension, followed by modeling this strategy with examples. See Figure 4 below and Snapshots in the Appendix I. After the instruction video was over, participants were presented with a "Checkpoint" to test participants' comprehension of each strategy. Participants could choose the Quiz or Game Show format for each checkpoint.

The quiz checkpoint included four multiple-choice questions. After submitting an answer, participants could see their score and the correct answer in green (see Snapshots in Appendix I). At the end of the quiz, they could go to the scoreboard to see their total scores. See the Snapshots in Appendix I.

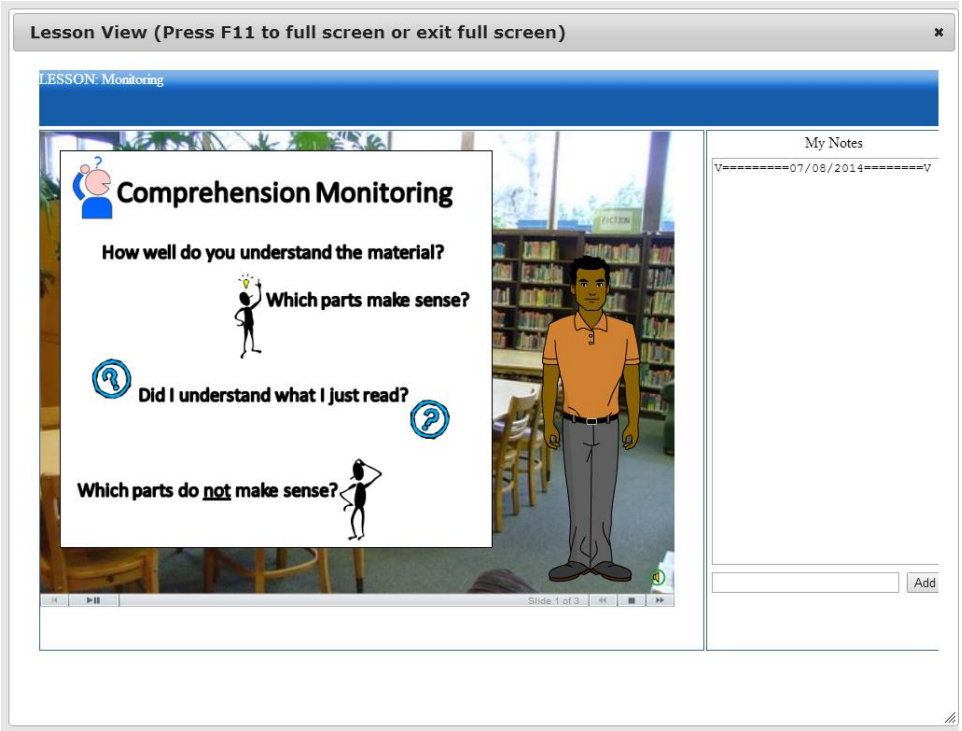


Figure 4: Screenshot of Monitoring

The game checkpoint included the same 4 multiple-choice questions but differed in terms of the interface and the immediate display of the scores of the current game

player and two other competitors' records which were chosen randomly with a certain likelihood of getting a correct answer. See snapshots in Appendix I.

Paraphrasing video. After finishing the Monitoring video, the Paraphrasing was unlocked. This video followed the same design as Monitoring video.

Prediction video. After finishing the Paraphrasing video, the Prediction video was unlocked. This video followed the same design as Monitoring video.

Elaboration video. After finishing the Prediction video, the Elaboration video was unlocked. This video followed the same design as Monitoring video.

Bridging video. After finishing the Elaboration video, the Bridging video was unlocked. This video followed the same design as Monitoring video.

Summary video. This video comprised a video to summarize how to use self-explanation and the five strategies to help reading comprehension. It emphasized the importance of applying these strategies in reading and summarized the key four points in practice. See Figure 5 below, and snapshots in Appendix I. At the end of this video, the Demonstration Module and Initial Practice Module were introduced so as to direct participants to watch the self-explanation and strategy application demonstrations and to practice these strategies with self-explanation.

Demonstration Module. With examples, this module demonstrated how to combine these strategies to improve self-explanation quality and reading comprehension. See Figure 6 below, and snapshots in Appendix I. When the participant was asked to explain a sentence, the virtual instructor first asked the participants to think how to self-explain this sentence and paused for the participants to think. After the pause, the instructor gave his own self-explanation, and asked the participants what strategies he

had used, which was followed by a short pause. After the pause, the instructor explained what strategy he had used with an elaboration of the definition of the strategy. He usually used more than one strategy.

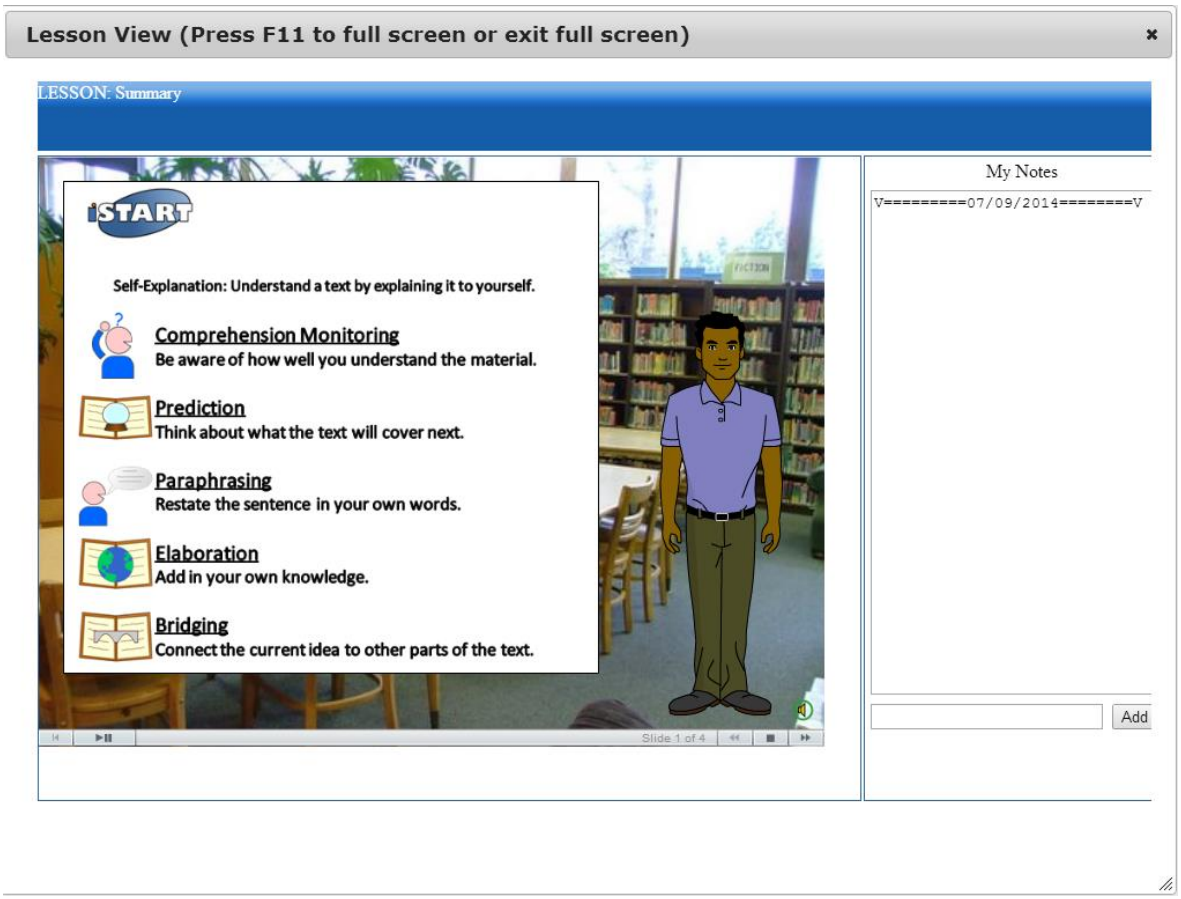


Figure 5: Screenshot of Summary

During the process, participants could choose to make each pause longer, if needed. After self-explaining each sentence in this paragraph, the video introduced the next initial practice module, and emphasized the importance of practicing these strategies together. Throughout the whole video, participants were asked to think how to explain a sentence and what strategies were used in the instructor's self-explanation, but were

never asked to write down or speak out their self-explanation.

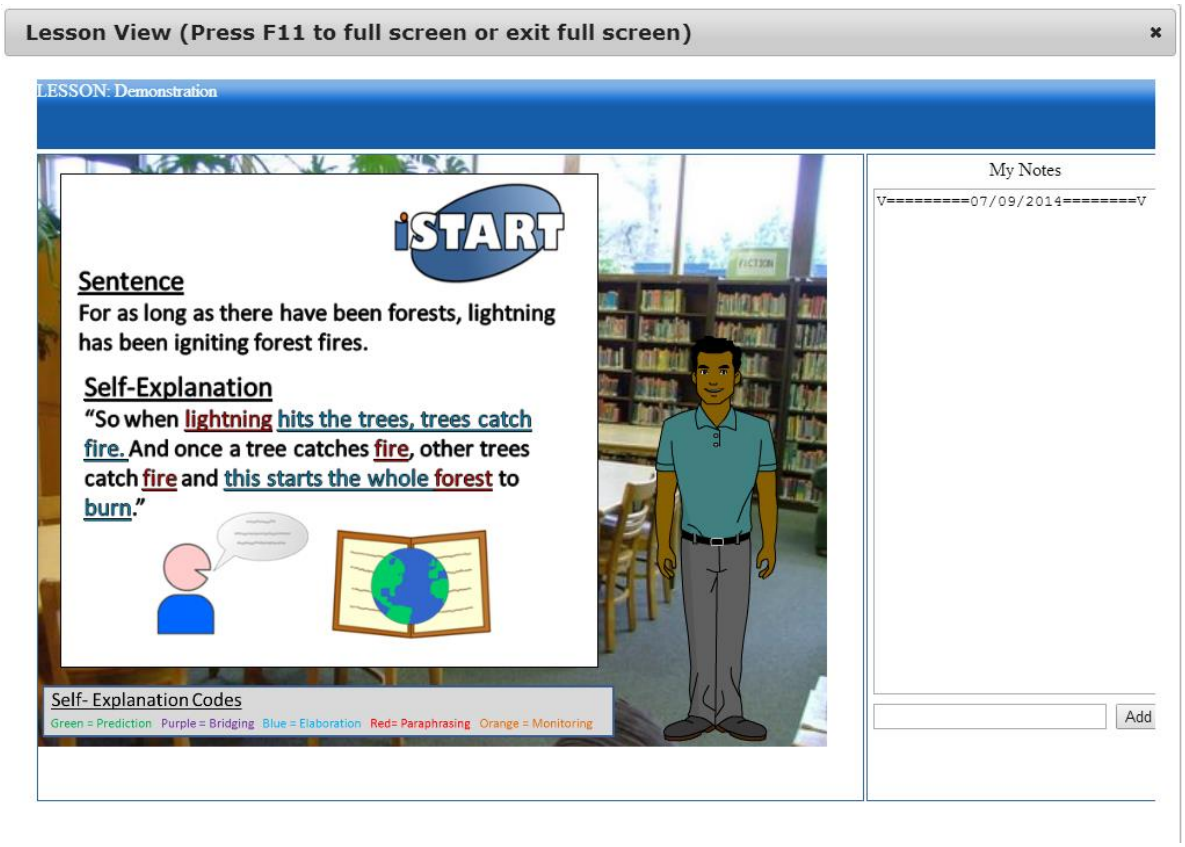


Figure 6: Screenshot of Demonstration Module

Initial Practice. The initial practice was a Coached Practice of the practice module. See Figure 7 and snapshots in Appendix I. During the initial practice or extended practice (Coached Practice and games), the participants were encouraged to apply any strategies or all strategies appropriate for self-explaining the sentences. These texts in practice were taken from middle school students' science textbooks. Participants were asked to explain sentences which were highlighted, one-by-one. After submitting a self-explanation, the participants could see their performance rated as poor, fair, good, or great. If the answer was not good, the instructor would push the participants to read previous sentences and get more information to help their comprehension and self-

explanation. Sometimes the instructor asked what strategies the participants used in their self-explanation. When the task of explaining one sentence was over, a "NEXT" button at the bottom appeared, and the participants could continue to the next sentence to explain. The total number of sentences to be explained was shown at the "Turn No:" on the left side of the screen, and varied among different texts, ranging from 6 to 10 sentences. When the practice was over, the participant could go to see his or her scores by clicking on "go to score." See snapshots in Appendix I.

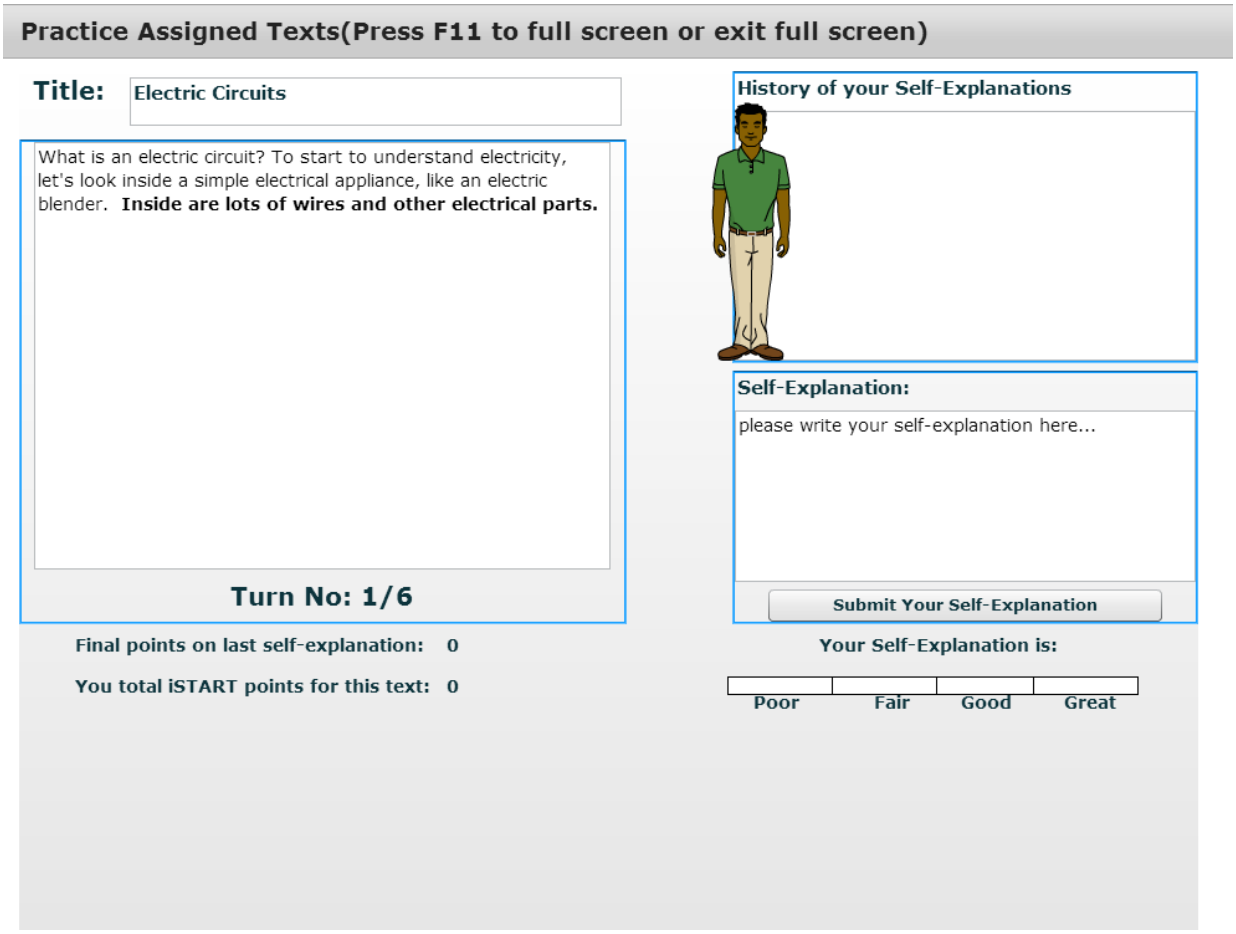


Figure 7: Screenshot of Coached Practice

Bridge Builder. This was the first game in the extended practice. See Figure 8 for the interface of extended practice. This game requested the participants to drag the

correct strategy symbol to the bridge, after they decided what strategy was used in the explanation of the target sentence. See the below Figure 9 and snapshots in Appendix I.

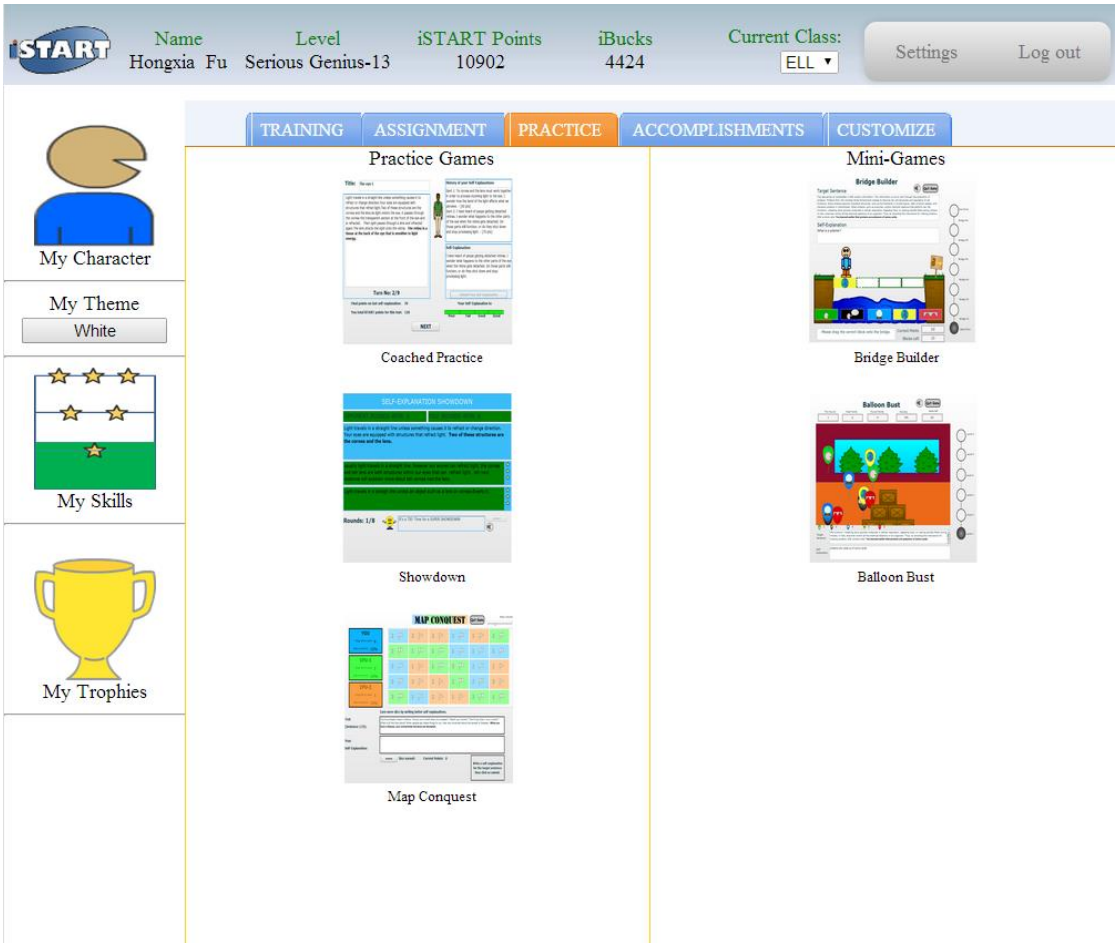


Figure 8: Screenshot of Extended Practice Interface

Participants had 30 strategies/blocks to drag into place. If they answered them all correct, they would see 30 different self-explanations, but if they answered some wrong, they would see fewer than that because they could get up to three chances per self-explanation. The participants could choose to have the background music or not by clicking on the speaker icon at the top of the screen. Whenever the participant selected the correct answer the character moved one block and a next self-explanation sentence appeared. If the character had successfully reached the right, and the game was not over,

the character would move back to the start point and move forward again until they had dragged all 30 strategies/blocks to the place. Once the character crossed the bridge, the participant gained one level up. At the end of the game, a result page showed the participant's score, level, iSTART Points, and iBucks.

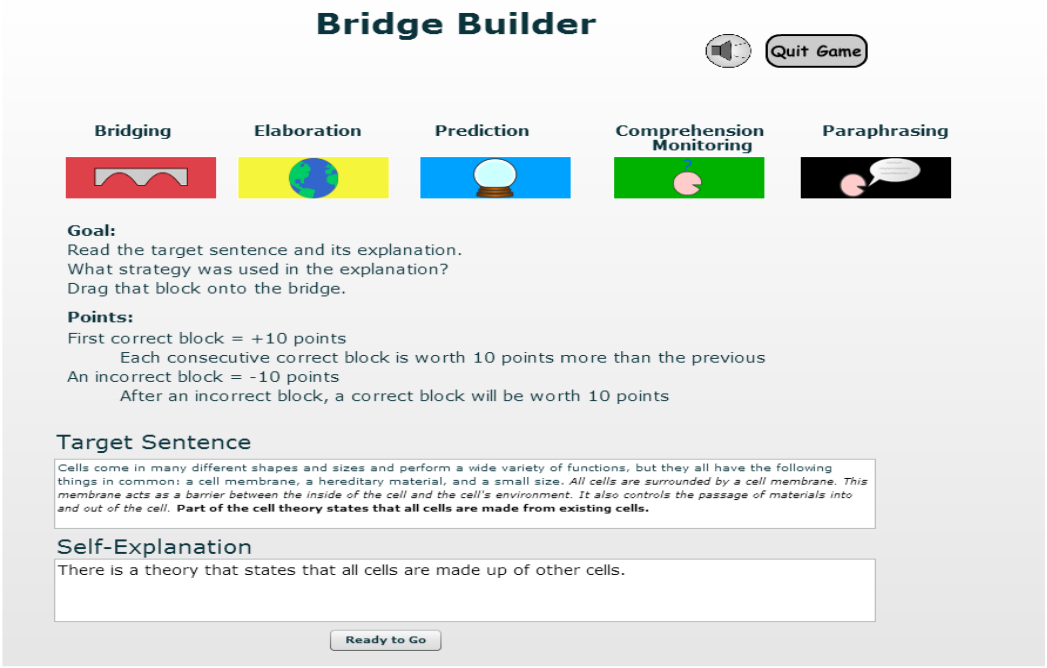


Figure 9: Screenshot of Bridge Builder Game

Showdown. This was a game that allowed the participant to compete with another randomly chosen player from the iSTART-ME database. The scores were based on each player's self-explanation quality. The text was randomly suggested by the system, and each text had 6-9 sentences for participants to do self-explanation. There was no time limit for each self-explanation and the scores were decided by the self-explanation quality as judged by the algorism built in iSTART-ME. After the participant submitted his/her self-explanation for one sentence, the opponent's self-explanation and the participant's own self-explanation showed up in the windows, and their scores also showed up. In the

meantime, a new sentence was highlighted for the next round of self-explanation competition. See Figure 10 below and snapshots in Appendix I.

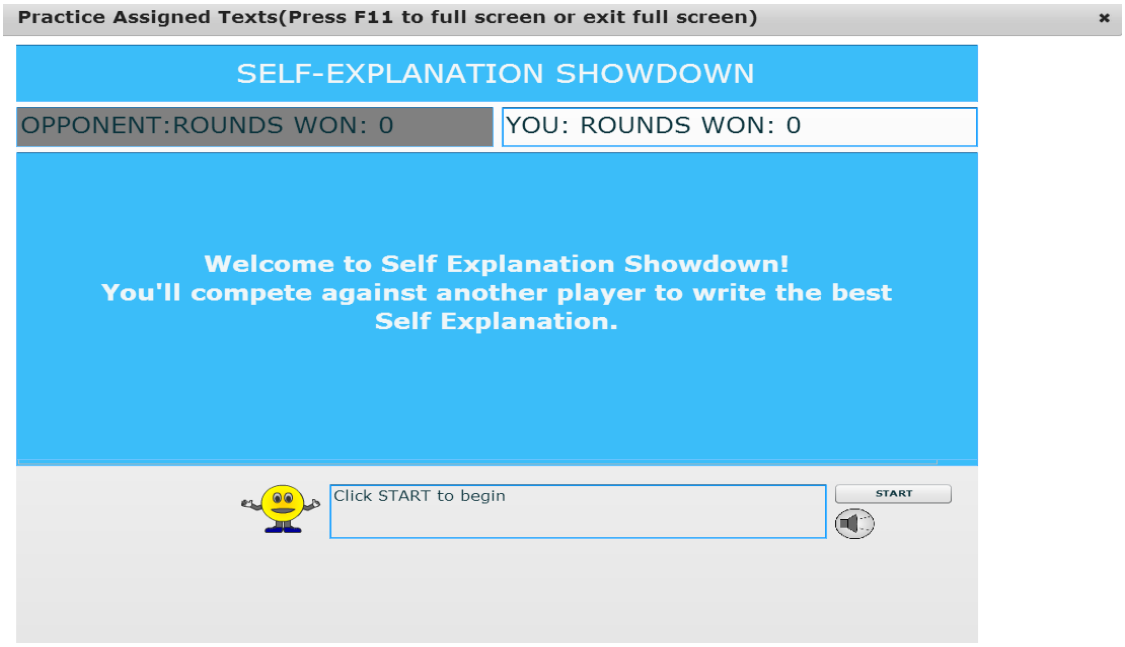


Figure 10: Screenshot of Showdown Game

Balloon Bust. This game required the participants to read the target sentence and its self-explanation to decide which strategy was used in the explanation, and then to click on a moving balloon with the corresponding strategy symbol. See Figure 11 below and snapshots in Appendix I. Participants could see their growth on the right during the game play. For Balloon Bust, participants got 60 darts. If they were always correct, they would see 30 self-explanations (because they needed to pop the balloons twice per self-explanation). However, because there were 8 "incorrect" balloons per self-explanation, they might actually spend up to 10 darts for a single self-explanation. At the worst, students would see six different SEs .Because balloon bust was hard for non-gamers, participants could choose to play Balloon Bust or Map Conquest described below. This

design also aimed to provide an environment with some freedom for making the learning more enjoyable.

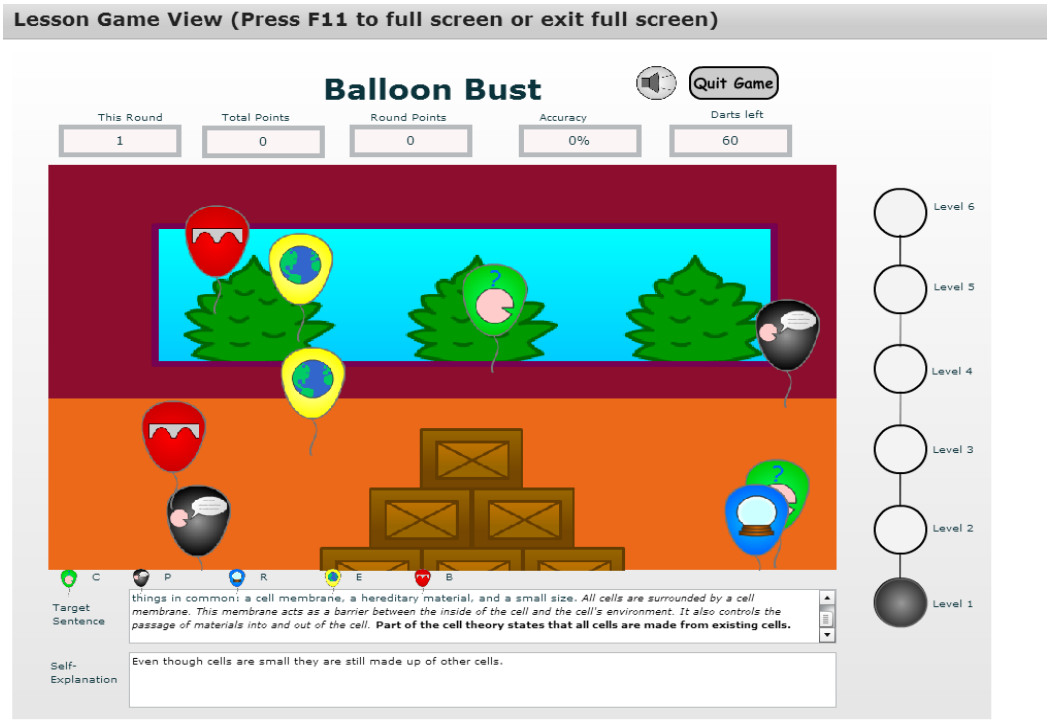


Figure 11: Screenshot of Balloon Bust

Map Conquest. This game required participants to conquer territory on a map by writing better SEs than their opponents. See Figure 12 below and snapshots in Appendix I. The text was randomly suggested by the system and each text had 6–9 sentences for participants to do self-explanation. There was no time limit for each self-explanation, and the scores were decided by the self-explanation quality. After explaining two sentences, a participant could obtain scores which determined how many flags the participant could place on his/her territory. The map ownership was displayed on the left. With more than 1 flag in one territory cube, the participant could conquer an enemy's territory. After the participant finished conquering their territory, the two enemies began to conquer their territory. Then another round of self-explanation began. When the competition was over,

the results were shown, including the iSTART Points and iBucks the participant earned. The earned iSTART Points decides the student's level, and iBucks could be used to purchase rewards in the system. However, in this study, participants were not allowed time or chances to play with these features.

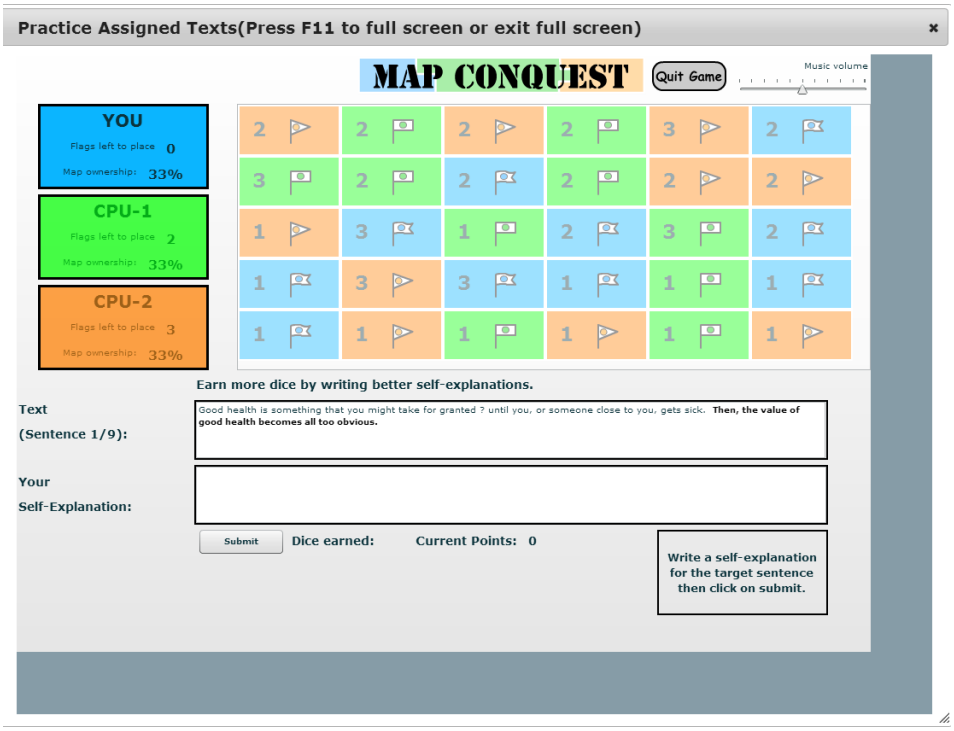


Figure 12: Screenshot of Map Conquest Game

Post-Intervention Materials

The post intervention materials used in this study are shown in Table 7.

Table 7: Post Intervention Materials

Materials	Purpose
1. Adapted PSAT test	To assess post reading comprehension ability
2. Short-answer reading comprehension test	To assess post reading comprehension ability
3. Self-explanation test	To assess post self-explanation ability
4. Adapted ATTAS survey	To assess the attitude towards the conversational agent
5. Adapted MSLQ motivation survey	To assess post motivation level

Adapted PSAT Critical Reading Test. See the relevant description in the section of Pre-Intervention Materials.

Short-answer reading comprehension test. See the relevant description in the section of Pre-Intervention Materials.

Self-explanation test. See the relevant description in the section of Pre-Intervention Materials.

Adapted MSLQ motivation survey. See the relevant description in the section of Pre-Intervention Materials.

Adapted ATTAS scale. ATTAS (Attitude toward Tutoring Agent Scale) is a scale to measure users' perceptions of an animated conversational agent, developed by Adcock and Van Eck (2005). This scale was validated using the interactions between college students and AutoTutor *(*Adcock and Van Eck), an intelligent tutoring system. It includes these three constructs of conversation/pedagogy, attitude toward student, and student interest/attention. This scale was derived from student rating systems of higher education faculty. Each construct includes 5 items with Cronbach alphas of .84, .87, and .89, respectively (Adcock & Van Eck). Considering the actual iSTART-ME learning environment, this study selected three applicable items from Conversation/Pedagogy construct and maintained the original five items of the other two constructs as in the original survey. See the Adapted ATTAS in the Appendix H.

Interview prompts. Both the one-on-one and focus group interviews used these same four prompts, as shown below. Follow-up questions were asked according to the interviewee's answer to each question.

1. “Please talk about what you have learned through this program.” If the interviewees did not mention their future plan or intention to apply these strategies, they were asked the follow up question “Have you thought about how to apply these strategies in your future study?”
2. “What reading strategies have you learned before this program?”
3. “Please talk about what you like this program, and what you dislike this program.”
4. “Any suggestions to improve the program?” Depending on the interviewees’ answers to each question, different follow-up questions were asked.

Procedures

On day 1, pretests, presurveys, and the iSTART-ME introduction module, demonstration module, and initial practice were completed. On day 2, iSTART-ME training Overview, Coached Practice, extended practice, posttests, postsurveys, and interviews were completed. Through the whole study, the participants were allowed to ask the study administrator questions, but were not allowed to consult with their peers. Table 8 presents descriptions and time estimations for each of the activities.

Table 8: Research Procedure

Timeline	Participants' Tasks	Purpose
Day 1(approximately 3hrs, task based), computer lab	*Pre-interventional tests & surveys (about 1hr)*	
	1. Adapted PSAT Critical Reading 1 test	To assess prior reading comprehension ability
	2. Short-answer reading comprehension test	To assess prior reading comprehension ability
	3. MARSI survey	To assess strategy application
	4. Self-explanation test	To assess prior self-explanation ability
	5. Demographic survey	To collect participants' demographic data
	6. Adapted MSLQ survey	To assess prior motivation level
	Refreshment and Break(10mins)	To energize the participants
	iSTART-ME training (2hrs)	
	iSTART-ME Training, including 8 modules of overview, 5 strategies training, summary, and demonstration	To acquire the knowledge of the strategies
Day 2 (approximately 2.5 hrs, task based), computer lab	*iSTART-MEPractice (1 hr)*	To learn strategies
	1. Overview (4 mins)	To review the strategies learned
	2. Coached Practice (task-oriented, about 18 mins)	To practice self-explanation with the instructional feedback
	3. Bridge Builder Game (10 mins)	To practice strategy identification in the game
	4. Showdown Game (15 mins)	To practice self-explanation in the competition
	5. Balloon Bust or/and Map Conquest Game (13 mins)	To practice strategy identification in Balloon Bust; to practice self-explanation in Map Conquest
	Refreshment and Break (10 mins)	
	Post-interventional tests & surveys (50 minutes)	
	1. Adapted PSAT test	To assess post reading comprehension ability
	2. Short-answer reading comprehension test	To assess post reading comprehension ability
	3. Self-explanation test	To assess post self-explanation ability
	4. Adapted MSLQ survey	To assess post motivation level
	5. Adapted ATTAS survey	To assess the effectiveness of the system
	One-on-one interview(10mins) & focus group interview (25 mins)	To collect individualized feedback beyond the surveys

Preparation. Two computer labs (66 computers each) with Internet access were reserved on the campus from 8:00 AM to 6:00 PM during the study days. One lab served as the main space, and the other was supplementary in case there were more participants or problems with the computers in the main lab came up. In order to accommodate participants' schedules, at every hour a new session could start. Because the study took participants more than 2 hours each day, there would be participants who were doing different sessions at the same time in the lab. In consideration of this supervising challenge, four Chinese graduate students in the engineering school were hired as the author's assistants to run the study. All of them were familiar with troubleshooting computer problems. Two days before the study began, these four assistants went through the whole study process as a study participant would do. They were also given the written instructions and directions for all the tasks and procedures for each day. During Day 1, the researcher and the assistants were all present at the back of the lab, and each person was responsible to answer questions or monitor the process for one assigned area of the lab. On Day 2 the assistants were all present at the back of the lab, and the researcher was there too until the researcher started to do interviews. To ensure the study process would go as planned, the iSTART-ME programmer and researchers at the Solet Lab were on phone and online to provide support during this study time. Fortunately, such a need never arose. Furthermore, during the study days every computer was tested and set on the right page for the participant to start.

Day 1. The author and her four assistants directed and supervised the whole process. Participants were informed upon arrival that the participation was voluntary and they could quit at any time. They read the consent form and signed it. They were also

encouraged to ask any questions before they began the session. All participants were entered in a lottery for 10 debit cards of $50 each. The participation rewards and rules were clarified. They were provided free bottled water and snacks during the research time.

After the above short clarification, the participants were assigned to every other computer, and the participants who came in at the same time were assigned to the same area of the lab. For each session, no more than 10 new participants showed up. The study Consent Form page on the computer was the start place for each participant. After they agreed to participate in this study, participants sequentially completed the Adapted PSAT test, short-answer reading comprehension test, strategy application survey (MARSI), self-explanation test, demographics questions, and adapted MSLQ motivation survey. Participants finished all instruments in between 50 to 65 minutes. After doing this, they took a 10-minute break to have water and snacks. After the break, they began the training section of the iSTART-ME. The strategy training was done on the website http://istart.soletlab.com, where each participant logged in to a computer with their own username and password in the lab. After login, they could see the interface of the training materials, as shown in Figure 2. Participants took about 2 hours to complete the intervention. The time varied because some participants did Coached Practice slower while some were faster, but the time variance was within 10 minutes.

Day 2. The first 5 participants who came to the lab on day 2 drew the lottery for the day 1 winners; the last 5 participants who came to the lab on day 2 drew the lottery for the day 2 winners for 2 debit cards of $200 each and 5 Broadway performance tickets.

The day 2 intervention began with the Overview video, followed by the extended practice, including a new Coached Practice, a Bridge Builder Game, a Show-down Game, and a Balloon Bust or/and a Map Conquest Game in sequence. This took the participants 1 hour total.

After the extended practice, the participants took a 10-minute break for water and snacks. After the break, the posttests and surveys were conducted in sequence (PSAT, short-answer comprehension test, self-explanation test, adapted MSLQ survey, and adapted ATTAS survey). Participants took between 50 and 60 minutes to complete the posttest and surveys.

After completing all the above tasks on computer, the selected interviewees were interviewed in a classroom next to the computer lab. The one-on-one interview took about 10 minutes, and the focus group took about 25 minutes. In order to have more connections with the interviewees and to get accurate expressions, participants were interviewed in Chinese. Notes for the one-on-one interviews and the focus group were written up immediately after each session.

Data Analysis

To answer the research questions, the data were merged for analysis with the quantitative research software SPSS and the qualitative research software of HyperResearch. To answer RQ1 below, the participants' comprehension scores were collected in both pretest and posttest, including scores on adapted PSAT test and short-answer reading comprehension test. The participants' SEWC (self-explanation word count) scores in pretest and posttest were collected, too. A within-subject t-test was conducted to analyze the difference between the pretest and posttest. All the negative

items on each survey were recoded to make the bigger number representing more positive effect for each item. The learning motivation survey of MSLQ was administered before and after the training intervention; the metacognitive strategy survey of MARSI was administered before the training; the attitude toward the animated conversational agent survey of ATTAS was administered after the training. The participants' demographic features were collected before the training.

RQ1. What are the effects of iSTART–ME on L2 reading comprehension and self-explanation?

1.1. Do participants improve their reading comprehension after the training?

1.2. Do participants improve their self-explanation quality after the training?

1.3. Do participants write longer self-explanations after the training?

To answer below RQ2, participants' scores on the MSLQ pre-survey and post survey were collected, and a within-subject t-test was conducted to compare the score difference. Participants' ATTAS score after the intervention was also collected to analyze their attitude towards the conversational agent built in iSTART-ME, and average score was calculated. The interview data collected after the intervention were coded and categorized with constant comparison method to infer the themes of the interviewees' point of views.

RQ2. What are participants' perceptions of their learning experience with iSTART-ME?

2.1 Has participants' learning motivation changed after the training?

2.2 Did participants hold a positive attitude towards iSTART-ME?

To answer below RQ3, a Pearson or Spearman correlation analysis was done to answer each correlation sub-question according to the data distribution. If a correlation existed, a regression model was applied to identify the predictors and relationship. According to Harris (1985) and Green (1991), the minimum sample size for multiple correlation is 50 + m (the number of independent variables) or 104 + m for testing individual predictors, assuming a medium-sized relationship. Because of the small sample size in this study, the power of the significance test might not identify a small or medium-sized relationship. The results of correlation and regression analysis in this study may work only as a reference. Interpretation of the results should be used with caution.

RQ3. What factors account for changes in reading comprehension and self-explanation?

3.1. Is there correlation between self-explanation word count change and self-explanation score change?

3.2. Is motivation change correlated to participant's self-explanation gain scores?

3.3. Are pre-vocabulary scores correlated to participant's self-explanation gain scores

3.4. Are prior metacognitive strategies related to self-explanation gain scores?

3.5. Are prior metacognitive strategies related to short answer reading comprehension gain scores?

3.6. What individual characteristics predict self-explanation gain scores?

3.7. What individual characteristics predict short answer comprehension gain scores?

Methods Limitations

The limitations of this study were as follows: 1) Self-explanation score or strategy survey is only an indirect way to reflect participants' strategy application ability, as other currently existed means. Because most of the strategy application activities happen in a reader's mind, it is hard to measure. 2) Self-explanation was scored through an algorithm built in the iSTART-ME system. Although the scoring correctness was as good as the average level of human raters, it might not be as good as an experienced excellent human rater. 3) The sample used in this study was not randomly selected from a population, so caution needs to be taken to make any generalization.

Delimitations

This study collected quantitative data from two designs, within-subject pre- and posttest design and survey design, and qualitative data from one-on-one interviews and a focus group interview. This methodological triangulation, by using multiple qualitative and quantitative methods, increased the validity of this study and uncovered the deeper meaning of the data. If the conclusions from each of the methods are consistent, the validity is established. Moreover, for each method, the selected instruments and design were discussed with professors working in instructional design, teaching and learning, and intelligent tutoring systems fields. In the administration of the study, the author carefully monitored the whole process and her own thinking, based on her trainings and experiences in qualitative and quantitative research methods. Therefore, with the current design, the conclusion of this study about this sample should be valid, due to the triangulated feature of this research design.

CHAPTER IV

RESULTS

Introduction

This chapter combines the data analysis results from different data sources to answer the research questions below:

1. What are the effects of iSTART–ME on l2 reading comprehension and self-explanation?
    1.1. Do participants improve their reading comprehension after the training?
    1.2. Do participants improve their self-explanation quality after the training?
    1.3. Do participants write longer self-explanations after the training?
2. What are participants' perceptions of their learning experience with of iSTART-ME?
    2.1. Has participants' learning motivation changed after the training?
    2.2. Did participants hold a positive attitude towards iSTART-ME?
3. What factors account for changes in reading comprehension and self-explanation?
    3.1. Is there correlation between self-explanation word count change and self-explanation score change?
    3.2. Is motivation change correlated to participant's self-explanation gain scores?

3.3. Are pre-vocabulary scores correlated to participant's self-explanation gain scores?

3.4. Are prior metacognitive strategies related to self-explanation gain scores?

3.5. Are prior metacognitive strategies related to short answer reading comprehension gain scores?

3.6. What individual characteristics predict self-explanation gain scores?

3.7. What individual characteristics predict short answer comprehension gain scores?

The data of the 34 participants who completed the entire research cycle were used for the following analyses unless otherwise noted. The majority of the participants were graduate students, and most of them were from China. See Table 9 for the demographic information and Table 10 for participants' majors.

Data were screened for outliers by calculating frequency and maximum and minimum values for each variable. Effect sizes are reported for individual questions using Cohen's d. The formula used is $d = (Mean_{post} - Mean_{pre})/Standard\ Deviation_{pre.}$

Table 9: Participants Demographics

Student Type		Gender		Age		Nationality				
Graduate (%)	Under-graduate (%)	Male (%)	Female (%)	Under 20(%)	20-30 (%)	China (%)	Brazil (%)	Colombia (%)	Iran (%)	Japan (%)
23 (67.6)	11 (32.4)	14 (41.2)	20 (58.8)	8 (23.5)	26 (76.5)	30 (88.2)	1 (2.9)	1 (2.9)	1 (2.9)	1 (2.9)

Table 10: Participants' Majors

	Frequency	Percent
Accounting	3	8.8
Biomedical Engineering	2	5.9
Business	5	14.7
Civil Engineering	2	5.9
Computer Science	1	2.9
Construction Engineering	1	2.9
Economics	2	5.9
Electrical Engineering	5	14.7
Geographical Information System	3	8.8
Global Health	1	2.9
Higher Education Administration	1	2.9
Industrial Engineering	3	8.8
Kinesiology	1	2.9
Mechanical Engineering	2	5.9
Physics	1	2.9
Urban Policy	1	2.9
Total	34	100

RQ 1. What Are the Effects of iSTART–ME on L2 Reading Comprehension and Self-Explanation?

RQ 1.1: Do the Participants Improve Their Reading Comprehension after the Training?

In this study reading comprehension was tested with two instruments. One was the adapted PSAT, and the other was a science passage reading comprehension test with 8 short-answer questions. The data collected from these two instruments had no missing data. Thus, this research question had two null hypotheses below:

1. Standardized reading comprehension scores do not change after this training intervention.

2. Short-answer reading comprehension scores do not change after this training intervention.

Hypothesis 1. This hypothesis was examined by comparing the pre and post-PSAT scores as the standardized reading comprehension.

Data distribution. The Shapiro-Wilk normality test showed that the adapted PSAT comprehension score difference between pre- and post-intervention was not in normal distribution ($w=.90$, $p<.05$). So a non-parametric test, the Wilcoxon Signed Rank Test, was used to test the pre- and post-comprehension score change. There were no missing data in these tests.

Results. The pre- and post-comprehension means are shown in Table 11. Wilcoxon Signed Rank Test resulted in a significance level $p >.05$. Therefore, the null hypothesis was confirmed, namely, no PSAT reading comprehension change happened after the intervention.

Table 11: Pre and Post PSAT Reading Scores

	N	Mean	Std. Deviation
Pretest score	34	50.98	24.94
Posttest score	34	39.21	33.30

Notes: At the end of the posttest, students reported that the post reading was harder than the pre reading, which might explain the lower mean in the posttest and the high standard deviation, although not statistically significant.

Hypothesis 2. This hypothesis was examined by comparing the pre and post short-answer reading comprehension scores.

Data distribution. The open-ended comprehension score difference distribution was tested with the Shapiro-Wilk normality test. The results showed that the violation of

the normality was slight (w = .94, p = .045). Because the p value was close to .05, the within-subject t-test was conducted. There were no missing data in these tests.

Results. The descriptive statistics are shown in Table 12. The t-test results showed that post-comprehension scores were significantly higher than the pre-comprehension scores ($t_{df=33} = 2.35$, $p<.05$). Thus, the null hypothesis was refuted, and the alternative hypothesis was confirmed. There was significant improvement in the short-answer reading comprehension scores after the intervention. The effect size of the improvement was medium, d =.41

Table 12: Short-Answer Comprehension Score Change

	Mean	N	Std. Deviation
Pretest Scores	76.75	34	13.08
Posttest Scores	82.17	34	14.40

RQ 1.2. Do the Participants Improve Their Self-Explanation Quality after the Training?

Self-explanation quality was indicated by the self-explanation scores, which were graded by the algorithm built into iSTART-ME. The null hypothesis was that learners' self-explanation scores had no changes after the training. To test this hypothesis, within-subject pre- and posttest t-test was conducted. Before doing the t-test, data screening was done and score difference data distribution were tested.

Missing data. S14 missed 2 post self-explanation scores. Because this accounted for 3% of the total subjects, less than 5, this subject was trimmed from the dataset for the data analysis.

Data distribution. Possible self-explanation scores ranged from 0 to 3, with 3 being the highest score. The pre- and post self-explanation mean scores are shown in Table 13. The Shapiro-Wilk test shows that the pre- and post self-explanation score differences were normally distributed, w=.96, df=33, p>.05. Therefore, it was appropriate to apply the t-test to test the null hypothesis.

Table 13: Self-Explanation Score Change

	Mean	N	Std. Deviation
Pre self-explanation score	1.49	33	.57447
Post self-explanation score	1.86	33	.67323

Results of the t-test. The t-test showed that post self-explanation scores were significantly higher than the pre self-explanation scores ($t_{df=32} = 2.63$, $p<.05$). Thus, the null hypothesis was rejected, and the alternative hypothesis was retained. In other words, self-explanation scores were significantly improved after the training intervention. The effect size of the improvement was medium (d = .63).

RQ 1.3: Do the Participants Write Longer Self-Explanation after the Training?

If a participant had a self-explanation score, he/she had a self-explanation word count (SEWC). On the contrary, if a participant did not have a self-explanation score, he/she did not have a SEWC. Word counting was conducted automatically within the iSTART-ME system. The null hypothesis for this research question was that learners' SEWC showed d no change after the training intervention.

Missing data. In the whole sample, only one subject (3%), S 14, had missing data values for this variable. S14 missed two post self-explanation score values and responsive SEWCs, so S14 was trimmed from the dataset for this analysis.

Data distribution. The smallest SEWC was 1; there was no maximum limit. The SEWC below in Table 14 refers to the word count (WC) of each self-explanation. The Shapiro-Wilk test showed that the pre- and post-SEWC difference was in normal distribution, w=.97, df=33, p>.05. Therefore, it was appropriate to apply the t-test to test the null hypothesis.

Table 14: SEWC Score Change (Gain Scores)

	Mean	N	Std. Deviation
Pre-SEWC	23.60	33	8.99
Post-SEWC	24.50	33	9.18

Results of t-test. Although the sample's post-SEWC was literally higher than the pre-SEWC, the within-subject pre and post t-test showed that post-SEWC scores were not significantly higher than the pre-SEWC ($t_{df=32} = .52$, p>.05). Thus, the null hypothesis was retained, namely, that there was no significant change in SEWC after the intervention.

RQ 2. What Are Participants' Perceptions of Their Learning Experience with iSTART-ME?

This research question was answered using three data sources: one-on-one interview data, focus-group interview data, adapted MSLQ and ATTAS survey data. The results are discussed separately below.

Interviews

Data Collection and Analysis. Using a systematic sampling method after each participant finished his/her post surveys and posttests, seven Chinese participants were interviewed one-on-one and three Chinese participants were interviewed as a focus group (see Table 15 for their demographics). For the one-on-one participants, 2 were undergraduates; 5 were graduates; 3 males, and 4 females. They were from schools of engineering, business, education, and health. For the focus group, all were graduate students from business and engineering schools. Two were females, and one was a male. These were semi-structured interviews. Participants were mainly asked to talk about what they had learned through this program, what they liked or disliked about this program, and any suggestions they had to improve this program.

Table 15: Interviewees' Demographics

Interview	Interviewee	Sex	School	Student Type
One-on-One	1	male	Engineering	Graduate
	2	female	Education	Graduate
	3	male	Business	Graduate
	4	female	Health	Graduate
	5	male	Engineering	Graduate
	6	female	Business	Undergraduate
	7	female	Engineering	Undergraduate
Focus Group	8	female	Business	Graduate
	9	female	Business	Graduate
	1	male	Engineering	Graduate

These interviews were conducted in their native language of Chinese within a very relaxed and informal environment. I, the interviewer of the same race as the interviewees, did not take notes or record the interviews. After each interview, I immediately recalled and translated the interviewees' words as accurately as possible, which were the data for analysis. All the "citations" of the interviewee's comments in this

study were translated by the interviewer as closely as possible to the interviewee's Chinese words. The initial data analysis was completed in the HyperResearch Software. After finishing the coding and categorizing process in HyperResearch, the code frequencies and categories were exported to an Excel spreadsheet to further check and revise coding and categorization, with constant reference to the original notes. Through this constant repeated coding and codes sorting process, two themes below were uncovered.

Theme 1: The learning with iSTART-ME was interesting and successful. All of the interviewees said iSTART-ME was not dull at all, and it was a great program. One interviewee told me that the program design was interesting, because it was different from normal tests or lectures.

They had heard or learned of testing strategies in China, but nobody had ever received any systematic strategy training and had never used the self-explanation method. One interviewee happily said,

> This not only helps me in reading, but also in writing. This program is great! If you have any similar programs, please let me know. Do you have a writing program? I have never got any systematic training in reading strategies in China, although I may have heard of a few strategies. I forget who told me those strategies. My English is not good enough, and this program really helps. I never felt bored during the program[4].

[4] All the comments or "citations" of the interviewees in this study were from the interviewer's translations of the interviewee's comments immediately after each 10-minute individual interview conducted in Chinese.

Interviewees said that they had learned the strategies and would like to consciously practice them in their future reading. They easily spoke out the names of these five strategies in English, and then spoke out their Chinese versions, which reflected their comprehension of the connotation and denotation of these strategies. They said these strategies were very useful and easy to learn.

Their motivation to learn more reading strategies and other strategies were inspired through this training program. Three interviewees directly asked about other, similar strategy training programs were available for them to learn; the other four interviewees stated clearly what they wanted to learn in the future. One participant wanted to learn fast reading strategies; one wanted to study vocabulary learning strategies; one wanted to learn reading strategies that could help her to recall the content she had read; one wanted to learn reading strategies at paragraph or passage level, and he said the current five strategies were of sentence level. All 7 (100%) interviewees wanted to learn other strategies.

On-on-one interview theme 2: Sometimes learners were confused and frustrated with iSTART-ME. One interviewee said he was not able to figure out how to play Map Conquest successfully, although he understood the instruction. Five (71.4%) of the interviewees said that the opponent in Showdown was too strong, so they were always beaten. One said sometimes the opponent's self-explanation was nonsense to him, but the opponent still won. Another interviewee said,

> Map Conquest is interesting, but it's an unfair game. The opponent is too strong; his area is continuous and big, and my area is scattered; so it is hard to beat the opponent. All the games are fun, but the showdown is too difficult. The

opponent is too strong, always beating me. I failed a lot in the Bridge Building game, though I thought my choice was correct. Don't know why. I like the program. Coached Practice is helpful. I like the Map Conquest most. I type words very fast, so my speed is fast. Showdown is good because it shows me how to do it with others' self-explanation.

In terms of the self-explanation quality, two interviewees wished to see the "correct answer," and wanted to figure out how to do self-explanation better. One interviewee said,

> I hope that the system can give me some feedback, like what I did well? What I did badly? Why was my explanation bad? Can the system have a discussion window where we can discuss our self-explanation results?

When they were followed up with what games they disliked, three (42.9%) of them mentioned Balloon Bust, because they were distracted by the moving balloon and words that were too small, and said that the game did not help their learning much.

Given the findings, I assert that when asked to do so, interviewees were able to name aspects of iSTART-ME that could be improved, overall they had very positive learning experiences with iSTART-ME. They were motivated to learn more strategies and improve their learning in the future.

Focus-group interview. The three-person focus-group interview results duplicated the results of one-on-one interview. The same two themes emerged. All of focus-group members said that iSTART-ME was a very good and successful program, and they were never bored.

RQ 2.1: Has the Participants' Learning Motivation Changed after the Intervention?

The null hypothesis was that learners do not change their motivation after the iSTART-ME training. Pre- and post-MSLQ learning motivation scores were used to answer this question using SPSS statistics software. Within subject t-tests were run to test this hypothesis.

Missing data. In the collected MSLQ pre-survey, two subjects, S12 and S25, missed item: "I think I will be able to use what I learn in this course in other courses," which was from the Task Value Construct on the survey. In the post-MSLQ survey, S5 missed item from the Control of Learning Beliefs Construct: "I tried to understand the content thoroughly;" S14 missed another item from the Self-efficacy Construct: "I'm confident I understood the basic concepts taught in this course." In total, 4 subjects, or 11 percent of the total sample, missed one data value each out of the survey. Because this sample was relatively small, the study used the multiple imputation method to examine the missing data pattern and found the data missing was random. Therefore, the study continually used the multiple imputation method to generate an expected value for each missing value to conduct the t-test below.[5] In this way, this dataset may better represent the whole sample than simply deleting them pairwise to make the dataset lose a total of 4 subjects.

[5] The Descriptive Statistics and Construct Internal Consistency test used the original data, excluding the missing data. This measure was taken because these calculations were based on each variable, not a pairwise variable. Therefore, the missing data value for one variable was 2.9% (1 missing) or 5.9% (2 missing), lower than or close to 5% of the total dataset.

Internal consistency test of the constructs[6]. Cronbach's alpha is recommended for measuring the internal consistency of each construct each time the survey is administered. The value of Cronbach's alpha is not only related to the correlation level of items in a construct, but also to the number of items in a construct. If the number of items is smaller, the value of alpha is reduced (Nunnally & Bernstein, 1994). This adapted survey included only 3 items for each construct, which was expected to have a lower Cronbach's alpha than the original survey; however, the final constructs' reliability were still acceptable except for the Control of Learning Beliefs Construct. Considering this construct had the lowest alpha (.68) among the four constructs in the original MSLQ manual, it seemed like a bad choice to delete one item from the construct which made the alpha even lower. The alpha value (.59) in the pre-survey was close to the value (.68) in the original survey, but the post-survey got a very low value (.13) which needs to be further studied in the future to find out the causes, like item adaption or the study features. Probably the third item in this construct was a different type of question, and the statement was inappropriate for this study. Specifically, the internal consistency test results showed that the Control of Learning Beliefs Construct had a low alpha value (α=.59 in the pre-survey; α=.13 in the post survey) in both the pre and post survey (see Table 16). Therefore, the Learning Belief construct was excluded in the further data analysis of this survey. The other constructs' alpha values were good (> .70), except for

[6] The t-test did not include the Learning Belief Construct, so it involved 3 (8.8%) missing data values. Therefore, the t-test used the pool values to replace the missing data values, which were calculated by the multiple imputation method.

the Intrinsic Goal Orientation construct in the post-survey which was .60. Because alpha values of.60 are considered acceptable, it was included in the further analysis.

Table 16: Reliability of the Adapted MSLQ

Construct	Presurvey		Postsurvey	
	α	Item	α	Item
Intrinsic Goal Orientation	.77	I prefer challenging materials so I can learn new things I prefer course material that arouses my curiosity, even if it is difficult to learn The most satisfying thing for me will be trying to understand the content as thoroughly as possible.	.57	The course material challenged me. The course material aroused my curiosity. I tried to understand the content thoroughly.
Task Value	.67	I think I will be able to use what I learn in this course in other courses. I like the subject matter of this course. Understanding the subject matter of this course is very important to me.	.77	I think I will be able to use what I learned in this course in other courses. I liked the content of this course. It is important to learn the content of this course.
Self-Efficacy for Learning & Performance	.71	If I study in appropriate ways, I will be able to learn the course material. It is my own fault if I don't learn the material in this course. If I try hard enough, I will understand the course material.	.85	I'm confident I understood the basic concepts taught in this course. I believe that I did well in this class. I'm certain I mastered the skills being taught in this class.
All Three Constructs, Above	.82		.76	

Table 16 cont.

Construct	Presurvey α	Presurvey Item	Postsurvey α	Postsurvey Item
Control of Learning Belief	.59	I'm confident I can understand the basic concepts I expect to do well in this class I'm certain I can master the skills being taught in this class.	.13	It was my own fault if I did not learn the material in this course If I tried hard, I could understand the course material. If I did not understand the course material, it was because of the computer system.

Data distribution. The descriptive statistics for these constructs are shown in Table 17. The Shapiro-Wilk test showed that the mean difference between the pre and post survey on each construct was normally distributed ($p>.05$); the mean difference for the overall survey items was also normally distributed ($p>.05$). See Table 18 for the normality test statistics.

Table 17: Descriptive Statistics of MSLQ

Construct		Mean	N	Std. Deviation
Pair 1	Pre_Intrinsic Goal	4.49	34	1.16
	Post_Intrinsic Goal	5.29	34	1.03
Pair 2	Pre_Task Value	4.82	32	.88
	Post _Task Value	5.60	34	1.02
Pair 3	Pre _Self-Efficacy	5.30	34	.54
	Post _Self-Efficacy	5.38	33	1.00
Pair 4	Pre _Mean	4.75	31	.78
	Post _Mean	5.42	31	.83

Note: MSLQ is a 7-level scale. The bigger of the Mean value, the stronger of the motivation is.

Table 18: Adapted MSLQ Pre and Post Mean Difference Distribution

	Shapiro-Wilk		
	Statistic	df	Sig.
Total Mean Difference	.95	34	.13
Self-Efficacy Difference	.97	34	.50
Task Value Difference	.95	34	.09
Intrinsic Goal Orientation Difference	.95	34	.14

Results of the t-test. T-test results showed that the subjects' average motivation level after the intervention was significantly higher than their level before the intervention ($Mean_{pre} = 4.75$, $Mean_{post} = 5.42$, $t_{df=33} = 4.86$, $p<.001$). The effect size d = .86. According to Cohen (1988), this was a large effect size. It indicated a nonoverlap of 47.4% in the two pre and post data distributions, and it also indicated that such a difference was clearly observable. There was also a significant difference between pre and post survey scores for the construct of Intrinsic Goal Orientation and the construct of Task Value (see Table 19 for details.). There was no significant difference between pre and post survey in the construct of Self-Efficacy for Learning & Performance, although the post-Mean score (M=5.38) was higher than the pre-Mean score (M=5.30).

Table 19: Comparison of MSLQ Scores

	Construct	Mean Difference	t	df	Sig (2-tailed)
Pair 1	Post-Intrinsic Goal *vs* Pre-Intrinsic Goal	.80	3.86	33	.00
Pair 2	Post-Task Value *vs* Pre-Task Value	.77	5.23	33	.00
Pair 3	Post-Self-Efficacy *vs* Pre-Self-Efficacy	.08	.46	33	.65
Pair 4	Post-total Mean *vs* Pre-total Mean	.67	4.86	33	.00

RQ 2.2 Did Participants Hold a Positive Attitude towards iSTART-ME?

Adapted ATTAS survey. This adapted ATTAS survey included 13 items, covering the three constructs of Conversation/Pedagogy, Attitudes towards Students, and Student Interest/Attention. The Likert scale was from Strongly Disagree to Strongly Agree with 6 levels.

Missing data. The data screening found that subject S28 missed Item 7 *The tutoring system made helpful comments*. According to the rationale stated above, the missing data value was less than 5%, so this subject was trimmed from the dataset in this analysis.

Reliability. The internal consistency of each construct was tested with Cronbach's alpha. The results showed that the Conversation / Pedagogy Construct and Student Interest / Attention Construct had good internal consistency with $\alpha = .87$ and $\alpha = .86$, respectively; the Attitude Towards Students Construct had acceptable consistency with $\alpha = .68$. Both the Cronbach's alpha and the Mean of each item are shown in Table 20.

Table 20: Construct Reliability of ATTAS

Construct	Survey		
	α	**Items**	**Item Mean**
Conversation / Pedagogy	.87	1. The tutoring system provided helpful feedback.	4.85
		2. The tutoring system responded effectively to my input.	5.00
		3. The tutoring system encouraged questions and answers.	4.82
Attitudes toward Students	.68	4. The tutoring system encouraged me to think for myself.	5.09
		5. The tutoring system encouraged the development of my knowledge.	4.94
		6. The tutoring system seemed impatient with me.	4.35(R)
		7. The tutoring system seemed friendly towards me.	5.09
		8. The tutoring system seemed discouraging towards me.	4.26(R)
Student Interest / Attention	.86	9. The teaching style of the tutoring system held my interest.	4.59
		10. The tutoring system made helpful comments.	4.76
		11. The tutoring system knows how to hold my attention	4.56
		12. when presenting material.	
		13. The tutoring system sensed when I needed help.	4.59
		14. The tutoring system increased my interest in the subject.	4.74

Notes: "(R)" means that the value of the item has been reversely scored. The bigger the value is, the more positive attitude the students hold towards the system.

Results. Participants showed positive attitudes toward the tutoring agent in iSTART-ME**.** Table 20 shows that the average score of each item on the survey was over 4.26, above the Somewhat Agree level (4). Three items had an average score over 5, approaching the highest degree of approval. These three items were:

- The tutoring system encouraged me to think for myself.
- The tutoring system seemed friendly towards me.

- The tutoring system responded effectively to my input

RQ 3: What Factors Account for Changes in Reading Comprehensionand Self Explanation?

RQ 3.1. Is There Correlation between SEWC Gain Scores and Self-Explanation Gain Scores?

This study used the Pearson Correlation test to test the null hypothesis that there is no correlation between learners' self-explanation score change (self-explanation gain scores) and SEWC score change (SEWC gain scores) through the training intervention.

Missing data. As in the above self-explanation and SEWC analysis**,** only one subject (3%) had missing data values in the sample, so the subject was trimmed from the dataset in this analysis.

Data distribution. According to Hatcher (2007), the assumptions underlying the Pearson correlation include interval-level measurement of variables, random sampling of the interested population, linear relationship between the two variables, and the bivariate normal distributions of the pairs of scores. Hatcher further described that "the Pearson correlation coefficient is robust against violations of this assumption [bivariate normal distribution] when the sample size is greater than 25" (p. 564). The sample met all but the random sampling assumptions. Considering the results of this study are not generalized to any whole international population, the violation of random sampling is not a concern.

Results of correlation test. The descriptive statistics of the two variables of self-explanation gain scores and SEWC gain scores are shown in Table 21. The Pearson Correlation test results showed that there was a significant positive correlation between learners' self-explanation gain scores and SEWC gain scores. Therefore, the test results rejected the null hypothesis, and retained the alternative hypothesis, namely that there

was a correlation between the learners' self-explanation gain scores and SEWC gain scores after the training intervention. This correlation was of medium size, r =.46, p<.05.

Table 21: Self-Explanation and SEWC Gain Scores

	N	Mean	Std. Deviation
Self-explanation gain scores	33	.36	.79
SEWC gain scores	33	3.60	39.59

RQ 3.2. Is Motivation Change Correlated to Participant's Self-Explanation Gain Scores?

This study used the Pearson Correlation test to test the null hypothesis that there is no correlation between learners' MSLQ gain scores and self-explanation gain scores through the training intervention.

Missing data. As explained in the "Missing Data" section of the of Research Question 1 results, multiple imputation methods were the same (involving 3 subjects)

Data distribution. This sample meets the assumptions of Pearson correlation study as described above on Page 121.

Results of the correlation test. A total of four pairs of correlation tests were conducted, including self-explanation gain scores and MSLQ total gain scores, self-explanation gain scores and Intrinsic Goal Orientation gain scores, self-explanation gain scores and Self-Efficacy gain scores, and self-explanation gain scores and Task Value gain scores. The results of all four correlation tests confirmed the null hypothesis, namely, there is no correlation between the subjects' self-explanation gain scores and MSLQ gain scores.

RQ 3.3. Is Pre-Vocabulary Score Correlated to Participant's self-explanation Gain Scores?

Missing data. The pre-vocabulary score variable had no missing data values in the sample. As discussed earlier, there was one subject who had missing self-explanation scores. Therefore, in this analysis, this one subject was trimmed from the dataset.

Data distribution. This sample meets the assumptions of Pearson correlation analysis as described above on Page 121.

Results of the correlation test. The results of Pearson correlation test ($p>.05$) confirmed the null hypothesis, namely, there is no correlation between pre-vocabulary scores and self-explanation gain scores.

RQ 3.4. Are Prior Metacognitive Strategies Related to Self-Explanation Gain Scores?

Descriptive statistics. This study used MARSI to survey what strategies the participants applied in reading a science passage, excluding five inapplicable Support Reading Strategies and one inapplicable Global Reading Strategy as described in the methodology section, previously. The application frequency and percentage of Global Reading Strategies, Support Reading Strategies, and Problem-Solving Strategies are shown in Tables 22, 23, and 24. In these tables, 12 (50%) of the total 24 strategies were used by half or more of the participants (indicated by the shaded items at the top of each table). The strategies used by the greatest number of students were three problem solving strategies: *I tried to get back on track when I lost concentration* (64.7%); *When text became difficult, I re-read to increase my understanding* (64.7%); *I tried to guess the meaning of unknown words or phrases* (64.7%).The least used strategies were *I critically analyzed and evaluated the information presented in the text* (14.7%), and *I asked myself questions to which I hoped to find answers in the text* (14.7%).

Table 22: Number of Participants Who Used Global Strategies

Item	Strategy	Frequency (N)	Percent (%)
1	I used context clues (e.g. other words around a difficult word, or an example) to help me better understand what I was reading.	21	61.8
2	I had a purpose in mind when I read.	20	58.8
3	I checked my understanding when I came across conflicting information.	20	58.8
4	I thought about what I had known to help me understand what I read.	19	55.9
5	I decided what to read closely and what to ignore.	19	55.9
6	I tried to guess what the material was about when I read.	19	55.9
7	I previewed the text to see what it was about before I tried to picture or visualize information to help reading it.	16	47.1
8	I skimmed the text first by noting characteristics like length and organization.	10	29.4
9	I used typographical aids like bold face and italics to identify key information.	10	29.4
10	I checked to see if my guesses about the text were right or wrong.	10	29.4
11	I thought about whether the content of the text fitted my reading purpose.	9	26.5
12	I critically analyzed and evaluated the information presented in the text.	5	14.7

Table 23: Frequency of Participants Who Used Support Strategies

Item	Strategy	Frequency (N)	Percent (%)
21	I went back and forth in the text to find relationships among ideas in it.	17	50.0
22	I summarized what I read to reflect on important information in the text.	9	26.5
23	I paraphrased (restated ideas in my own words) to better understand what I read.	8	23.5
24	I asked myself questions to which I hoped to find answers in the text.	4	11.8

Table 24: Frequency of Participants Who Used Problem-Solving Strategies

Item	Strategy	Frequency (N)	Percent (%)
13	I tried to get back on track when I lost concentration.	22	64.7
14	When the text became difficult, I re-read to increase my understanding.	22	64.7
15	I tried to guess the meaning of unknown words or phrases.	22	64.7
16	I adjusted my reading speed according to what I was reading.	20	58.8
17	When the text became difficult, I paid closer attention to what I'm reading.	20	58.8
18	I read slowly but carefully to be sure I understood what I was reading.	14	41.2
19	I tried to picture or visualize information to help remember what I read.	10	29.4
20	I stopped from time to time and think about what I'm reading.	8	23.5

Data distribution. This sample meets the assumptions of Pearson correlation analysis as described above on Page 121. No missing data.

Results of the correlation test. Pearson's correlation test refuted the hypothesis that there is no correlation between prior metacognitive strategy application and self-explanation gain scores, $r=.37$; $p<.05$. This is a positive medium effect size.

RQ 3.5. **Are Prior Metacognitive Strategies Related to Short-Answer Comprehension Gain Scores?**

Data distribution. This sample meets the assumptions of Pearson correlation analysis as described above on Page 121.

Missing data. One student, S9, missing one question score, but because this analysis was pairwise gain scores, one data point missing equals three data point missing, thus beyond 5% missing data. Therefore, this student's predicted score was calculated as explained above.

Results of the correlation test. Pearson's correlation test confirmed the hypothesis that there is no correlation between prior metacognitive strategy application scores (MARSI scores) and short-answer comprehension gain scores, r=.07; p=.69.

Pearson correlation test showed that the Motivated Strategies for Learning Questionnaire (MSLQ) scores prior to the intervention was also not correlated to short-answer comprehension gain scores, r=.18, p=.31

RQ 3.6. What Individual Characteristics Predict Self-Explanation Gain Scores?

Independent variables. These variables were included as individual characteristic variables: pre-SEWC scores, pre-motivation, pre-metacognitive strategies, pre-short answer comprehension scores, pre-multiple choice comprehension scores, pre-vocabulary, and game hours.

Data distribution. Because these variables and their relationship with self-explanation gain scores meets the assumptions of Pearson correlation analysis, the Pearson correlation coefficients were calculated. Missing data were implemented in the same way as explained before. If significant correlation existed between an independent variable and the self-explanation gain scores, the variable was added to the regression model. Because self-explanation gain scores were in normal distribution, linear regression model was applied.

Results of the correlation test. As shown above**,** SEWC gain scores and prior MARSI (metacognitive strategy) scores were significantly correlated with self-explanation gain scores. Among the rest independent variables, only the variable of pre self-explanation scores was significantly correlated with self-explanation gain scores, r=-.44, p<.05. This negative correlation means that participants with lower prior self-

explanation scores were more likely to get bigger self-explanation score increase, although as a whole group, the participants got positive self-explanation gain scores. Additionally, Pearson Correlation Test shows that the pre self-explanation scores were not correlated to the post self-explanation scores, r=.20, p=.27.

To eliminate the ceiling effect that skilled readers have less room to improve, relative SE gain scores were calculated as the criterion variable to test its relationship with the predictor of pre self-explanation scores. According to Jackson, Varner, Boonthum-Denecke, and McNamara, "Relative gain scores represent the amount of improvement achieved based on the amount of improvement possible [(Posttest proportion – Pretest proportion) / (1 – Pretest proportion)]" (p. 327). The Pearson correlation test shows that the variable of pre self-explanation scores was significantly correlated with the variable of relative self-explanation gain scores, r=-.40, p<.05. See Figure 13. Therefore, pre self-explanation scores significantly correlated with both self-explanation gain scores and relative self-explanation scores (a less biased measure). Additionally, Pearson correlation test showed that pre MSLQ scores, pre SEWC scores, pre metacognitive strategy scores, and pre short-answer comprehension scores were not correlated with the relative self-explanation gain scores, taking p=.05 as the significance level.

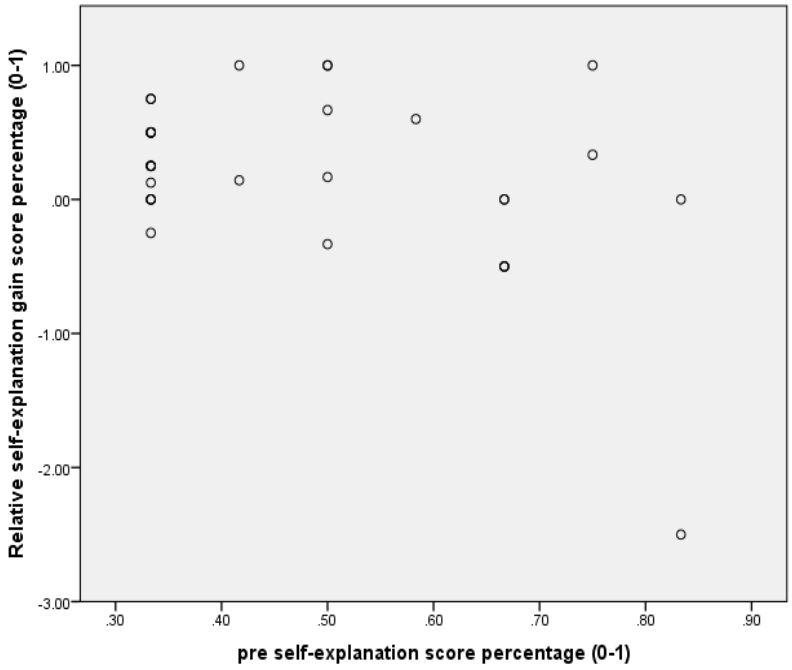


Figure 13: Scatterplots of pre SE scores and relative SE gain scores

Linear regression results. Because pre self-explanation score, SEWC, and pre-metacognitive strategies were significantly correlated with self-explanation gain scores, these three variables were entered to the regression model in a stepwise method. Only the variable of pre self-explanation scores was significant in the model, and the other two variables were insignificant in the model and were excluded from the model. Because only pre self-explanation scores were correlated to the relative self-explanation gain scores, no regression model was run.

A linear regression analysis revealed that Pre self-explanation scores was a highly significant predictor of regression scores (self-explanation gain scores), standardized β = -.56, p = .00. The adjusted R^2=.29, F(32) =13.82, p<.01, which demonstrated significant effects of pre self-explanation scores on self-explanation gain scores, accounting for 28.6% of the total variance of self-explanation gain scores.

RQ 3.7. What Individual Characteristics Predict Short Answer Comprehension Gain Scores?

Independent variables. These variables were included as individual characteristic variables: Pre-SEWC scores, pre- short answer comprehension scores, post short answer comprehension scores pre-motivation, pre-metacognitive strategies, pre-multiple choice comprehension scores, pre-vocabulary, and game hours.

Results of the correlation test. Because these variables and their relationship with self-explanation gain scores meets the assumptions of Pearson correlation analysis, the Pearson correlation coefficients were calculated. Pre-short answer comprehension scores were negatively correlated with short answer comprehension gain scores, $r = -.42$; $p<.05$, $N=34$. The pre- short answer test scores were significantly correlated to post short answer test scores, $r=.52$, $p=.00$. No significant correlation between other variables and the comprehension gain scores was found.

To eliminate the ceiling effect that skilled readers have less room to improve, relative short-answer comprehension gain scores were calculated as the criterion variable to test its relationship with the predictor of pre short-answer comprehension scores. The Pearson correlation test shows that the variable of pre short-answer comprehension scores was not significantly correlated with the variable of relative short-answer comprehension gain scores, $p>.05$. See Figure 14. Therefore, pre short-answer comprehension scores was significantly correlated with short-answer comprehension gain scores, but not significantly correlated with relative short-answer comprehension gain scores (a less biased measure).

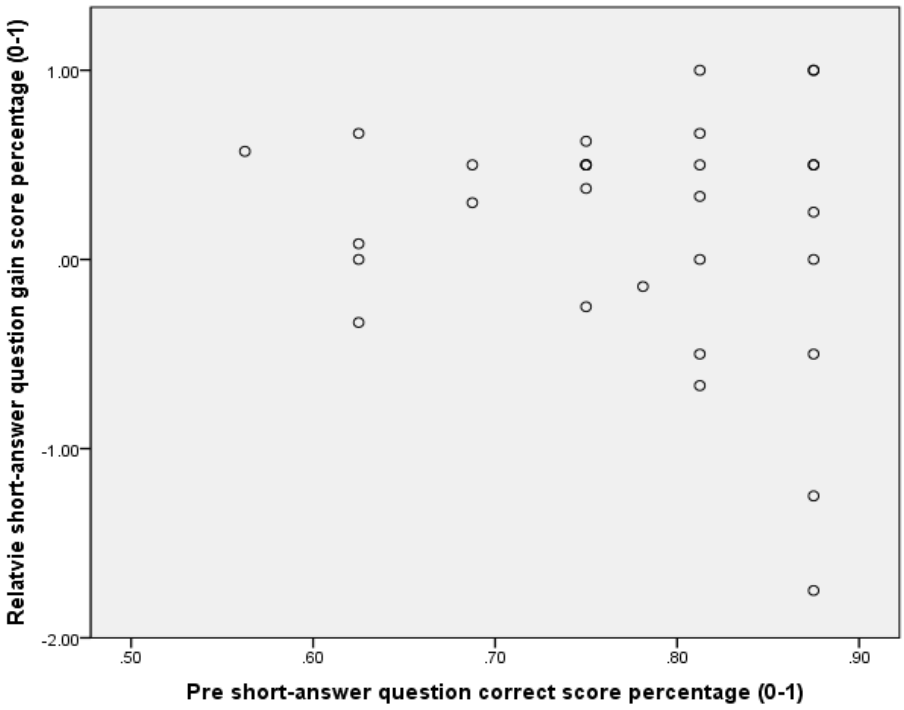


Figure 14: Scatterplots of pre short-answer scores and relative short-answer scores

Additionally, high and low pre SE score groups were created with a median split method (The two same pre SE median scores were classified to the high score group.) In the high score group (N=17), the relative short answer gain scores were significantly correlated to relative SE gain scores (r=.64, p=.01), and not significantly correlated to pre SE scores and pre short answer scores. In the same group, the relative SE gain scores were significantly correlated to pre SE scores (r=-.52, p=.03) and short answer gain scores (r=.55, p=.02). The short answer gain scores were significantly correlated to relative SE gain scores (r=.55, p=.02). The SE gain scores were significantly correlated to pre SE scores (r=-.69, p=.00). In the low score group (N=14), the pre SE scores were the same (no variance), and no corresponding significant correlation were found. Due to the small sample, caution should be taken in interpreting the results.

High and low pre short answer score groups were also created with a median split method. In the high group (N=16), the relative short answer gain scores were significantly correlated to relative SE grain scores (r=.54, p=.04). The relative SE gain scores were significantly correlated to pre SE scores (r=-.51, p=.04). In the low score group (N=16), no corresponding significant correlations were found. Due to the small sample, caution should be taken in interpreting the results.

Linear regression results. The variable of pre- short answer comprehension score was entered to the regression model. Other variables were not entered to the model because the previous correlation tests showed that they were not correlated to the shor-answer gain scores. A linear regression analysis revealed that pre- short answer comprehension score was a highly significant predictor of regression scores (short answer comprehension gain scores), standardized $\beta = -.42$, $p = .006$. The adjusted $R^2=.15$, $F(33) =6.77$, $p<.01$, which demonstrated significant effects of pre- short answer comprehension score on comprehension gain scores, accounting for 14.9% of the total variance of self-explanation gain scores. Additionally, the pre short answer comprehension scores explained 27% of the post short answer comprehension scores ($R^2 =.27$)

Summary

To summarize the statistically significant results, Table 25 shows the significant effect of the training on self-explanation, reading comprehension and motivation; Table 26 shows the correlational variable with self-explanation or reading comprehension; Table 27 shows the significant independent variables that predicted self-explanation and reading comprehension gain scores.

Table 25: Effects of iSTART-ME Training

Measuremnt	Significant Improvement	Effect Size (d)
Self-explanation scores	Y	.63
MSLQ scores/motivation	Y	.86
Short-answer test score	Y	.41
PSAT score	No	NA

Table 26: Variables Correlated to Self-Explanation or Reading Comprehension Gain Scores

Dependent Variables	Significantly Correlated Variables	Effect Size (r)
Self-explanation gain scores	SEWC gain scores	.46
	Pre self-explanation scores	-.44
	MARSI strategy scores	.37
Relative self-explanation gain scores	Pre self-explanation scores	-.40
Short-answer gain scores	Pre short-answer test scores	.418

Table 27: Predictors of Self-Explanation and Reading Comprehension Gain Scores

Dependent variables	Predictors(Independent Variables)	R^2	β
Self-explanation gain scores	Pre self-explanation scores	.29	-.56
Short-answer test gain scores	Pre short-answer test scores	.15	-.42

CHAPTER V

DISCUSSION

This chapter describes the findings related to each research question, followed by the implications of these findings for L2 teaching, L2 research, and ITS research, including future research directions. The limitations and contributions of this study are also discussed. At the end, recommendations are outlined.

RQ 1: What Are the Effects of iSTART-ME on L2 Reading Comprehension and Self-Explanation?

RQ1 Design Summary

The effects of iSTART-ME on L2 reading comprehension were measured by comparing participants' scores on pre and post standardized reading comprehension tests and a pre- and post- short-answer reading comprehension test. These two types of comprehension tests were used in light of the fact that that the effects of reading strategy training on L2 comprehension has not been equally significant in various types of tests (e.g. Barnett's,1988; Carrell, Pharis, and Liberto, 1989). The effects of iSTART-ME on self-explanation quality were measured by comparing participants' scores on pre- and post self-explanation tests and their self-explanation word count (SEWC) using a within-subject t-test. The correlation between SEWC gain scores and self-explanation gain scores was also calculated.

RQ1 Results Summary

The t-test on reading comprehension scores showed that there was statistically significant improvement in short-answer reading comprehension test scores after the intervention ($p<.05$), and the effect size of the improvement was medium ($d = .41$). These results indicate that in practice this improvement was observable. There was not a statistically significant change in the standardized reading comprehension test scores. The results also showed a significant improvement in self-explanation quality after the intervention ($p<.05$); the effect size of the improvement was medium ($d = .63$), which indicates that in practice this improvement was visible with the naked eye. While participants' SEs contained more words on the post-test ($M=24.50$) than in the pre-test ($M=23.60$), the difference was not statistically significant. However, the correlation between SEWC gain scores and self-explanation gain scores was significant ($r=.46$, $p<05$), indicating medium effect size.

Finding 1: Reading Ability Improvement.

The SERT instructional model built in iSTART-ME was effective in promoting L2 reading comprehension as measured by short-answer tests. These findings replicate prior research on metacognitive strategy training for L2 learners employing human teachers in classroom settings. These findings not only lend strength to the research on the use of such strategy training, but also extend research on ITSs like iSTART-ME. The effect size of score improvement on short-answer tests was exactly the same as the findings of Taylor, Stevens, and Asher's (2006) meta-analysis of the average effect of explicit metacognitive reading strategy training on L2 reading comprehension (.41). This finding is also consistent with Carrell, Pharis, and Liberto's (1989) finding that metacognitive strategy training significantly improved L2 learners' scores in an open-

ended recall protocol, although not on multiple-choice questions (three passages with multiple choice questions including factual information, direct inference, and indirect inference questions taken from popular ESL source materials). Taylor, Stevens, and Asher found the effectiveness of strategy instruction by traditional classroom teachers, including cognitive and metacognitive strategy training, was low to moderate, with a mean effective size of .54. Therefore, it seems that the training effect of iSTART-ME in this study equals the average effect of metacognitive strategy training for L2 learners in real classrooms, thus indicating that ITSs can be as effective as human tutors in this regard.

This study was also consistent with the findings of self-explanation studies in L1 learners, which has indicated that self-explanation can have a positive effect on short-answer comprehension assessment (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Crippen & Earl, 2007, Linderholm, Therriault, and Kwon, 2014). In a self-explanation eliciting study (provided human instruction that was equivalent to a one-time short training) with eighth graders (L1), Chi, Leeuw, & Chiu, M, Lavancher (1994) found that the gains were greater for the prompted self-explanation group (32%) than the unprompted group (22%) on short-answer questions, especially greater on the more bridging-inference complex questions (22.6% versus 12.5%). The current study replicated that study, prompted self-explanation with iSTART-ME having improved participants' gain scores on short-answer questions. Moreover, this current study extended the scope of self-explanation research to L2 learners.

McNamara (2004) found that significant comprehension improvement from self-explanation training (with humans doing the training) on a standardized test (Nelson

Denny Reading Test) occurred only for low-knowledge readers (L1) on text-based questions. McNamara, O'Reilly, Best, and Ozuru (2006) also found that the significant comprehension improvement of the self-explanation training with iSTART on a standardized test (Nelson Denny Reading Test) for adolescents only for low-knowledge readers (L1) at the text-based questions. While this study did not find any comprehension score improvement in the standardized test (PSAT), it did find that students with lower reading skills gained the most in the short-answer comprehension tests. The different results may be explained by several factors. One factor may have to do with differences between the different assessment tools. Another one may be related to the use of shortened reading sections in the PSAT rather than complete PSAT reading sections. The total training time and training design may be variables too. Of course, the difference between L1 and L2 may be another factor.

Furthermore, this study found that the readers using more metacognitive strategies tended to improve their self-explanation quality most, which prior studies rarely explored. Future research may explore this area further. The present study did not confirm the findings by Jackson, Varner, Boonthum-Denecke, & McNamara (2013) in two aspects: the relationship between prior reading strategy and self-explanation gain scores, and the relationship between pre self-explanation scores and self-explanation gain scores. Jackson, Varner, Boonthum-Denecke, & McNamara did not find that prior self-explanation scores and prior meta-cognitive strategy knowledge were related to self-explanation gain scores. The present study found they were correlated, and the effect size was medium. Furthermore, the pre self-explanation scores explained 28.6% of the self-explanation gain scores. Moreover, SEWC was also correlated to self-explanation gain

scores; pre short-answer comprehension scores were correlated to comprehension gain scores. These different results may be attributed to the different characteristics between L1 and L2 students. L2 students with lower prior self-explanation ability improved their self-explanation scores more than the students with higher prior self-explanation ability. Compared with L1 students, probably these L2 lower self-explanation ability students were more in lack of the strategy knowledge or lack of practice in using these strategies. This is indirectly consistent with the results that the L2 lower comprehension ability students improved their comprehension scores more than the students with higher comprehension ability did. Future investigations may explore these relationships or assumptions further.

To sum up, the different results about the effects of students' pre self-explanation, pre metacognitive strategies, and pre comprehension scores on self-explanation and comprehension gain scores between this study and previous studies may be caused by several factors. First, they could be due to the difference between L2 and L1 learners, such as the explicit strategy training may inspire the participant's L1 skill transfer. A difference in training time (minimum three hours in two sessions versus minimum eight hours in eight sessions) may also cause the difference. Different game conditions, or broadly speaking, study design may be another factor (e.g. self-chosen games and gaming time for each game versus assigned games and time).

Finally, the different assessment tools, such as multiple-choice questions and short-answer questions, may cause the difference. In answering short-answer questions, the writing as an exterior speech and the inner speech in the head with text to refine a meaning of the text may improve students' comprehension, according to Vygosky that

language is thought. Taylor, Stevens, and Asher (2006) found the test format is a significant moderating variable in their meta-analysis of strategy training effects. Nevertheless, the significant score improvement in short-answer questions in this study may indicate a reading ability improvement like the one in the study by Jackson et al. (2013). Therefore, more studies are called for to test the effects of strategy training on L2 learners.

Finding 2: Self-Explanation Quality Improvement.

The improvement in self-explanation quality scores from the pre-test to the post-test in this study confirmed the findings of previous studies using iSTART-ME (e.g. Jackson & McNamara, 2013a; Jackson, Varner, Boonthum-Denecke, & Mcnamara, 2013b; Jackson, Boonthum, & McNamara, 2010). In terms of training time, Jackson et al (2013a and 2013b) trained high school students (L1) in eight sessions (at least 1 hour/session), whereas this present study trained undergraduates for a total of three hours over the course of two sessions. Future studies may consider varying the training time to further study its impact. Jackson and McNamara also found that high school students (L1) maintained their increase in performance in a one-week delayed retention test; while this study did not measure delayed retention, future studies should be done to confirm similar results of iSTART-ME on L2 learners.

Finding 3: Self-Explanation Gain Scores and SEWC Gain Scores Correlation.

The medium sized significant correlation (r=.46) between SEWC gain scores and self-explanation gain scores in this study indicate that the participants who increased their self-explanation word counts improved their self-explanation quality scores. This finding is consistent with the finding of Chi et al. (1994) that high self-explanation score

explainers (L1) generated significantly more self-explanation propositions than low explainers, t (6) = 4.42, p < .0l., and Chi et al. (1989) also found that successful problem solvers (L1) generated significantly more self-explanation lines than less successful solvers (t(6)=2.16, p<.05). It is probable that self-explanation prompts readers to monitor their comprehension, and to therefore connect what they read to other ideas in the text or their world knowledge. This kind of reflection creates informative or valuable explanations instead of meaningless statements (Chi et al, 1989, 1994). In this sense, self-explanation may have a special value for helping L2 students write longer and meaningful content.

The Introduction, Demonstration, and Practice modules in iSTART-ME allow L2 learners to not only learn and apply reading strategies to expository reading materials, but also provide the opportunity for participants to practice their meaningful writing in L2 through self-explanation. They practice both their reading and writing meaningfully and simultaneously. This may be an evidence of Vygotsky's (2012, edited by Kozulin) findings in *Thought and Language* that the word and speech are dynamically related to thought and human consciousness. Writing the self-explanation might have helped the students to construct the meanings of the text, because both reading and writing share some cognitive process like thinking and constructing mental models (Stotssky, 1983, Tierney, 1992). The significant relationship between the self-explanation length increase and self-explanation quality increase in this study indirectly confirmed the finding of Carson, Carrell, Silberstein, Kroll, and Kuehn (1990). Carson, et al. studied 105 Chinese and Japanese adults with English proficiency from low-intermediate to advanced (around 400-525 in TOEFL) who were enrolled in pre-academic intensive English programs and

freshman composition courses at American universities. They found that L2 reading correlated significantly to L2 writing. Specifically, the L2 reading and writing coefficient was .49 for Chinese learners and .27 for Japanese learners. In this aspect, iSTART-ME training may have specific advantages for L2 learners. It may be valuable to explore the specific effects of self-explanation on writing gains in the future. For example, do participants improve their writing quality in terms of syntax complexity, vocabulary variety, idea organization, or creative ideas? Participants in this study did say that this training helped them in both reading and writing during the interview, and wanted to be notified of similar programs. Regarding the trend of integrating reading and writing in instruction, this study suggested that the game-based iSTART-ME environment with the built-in SERT model might be an effective tool in promoting L2 learners reading and writing ability, which needs further exploration in the future.

RQ 1 Summary

This study replicated and extended previously observed positive effects of 1) metacognitive strategy training by classroom teachers on L2 reading comprehension, 2) the effects of strategy training by classroom teachers on L1 reading comprehension, 3) the effects of self-explanation training delivered by human or technology on short-answer comprehension assessments for L1 learners, and 4) the effects of iSTART-ME training on self-explanation quality of L1 learners. This study's findings were inconsistent with prior research on the effects of iSTART-ME on improving comprehension scores on standardized tests. There are many possible explanations for this, but among the most likely is that results are probably due to different measurements, designs (e.g. different training hours and sessions), or the difference between L1 and L2 learners. The findings

of this study further suggested the possibility that iSTART-ME may be an effective self-explanation and metacognitive strategy training tool to boost incoming university-level international L2 learners' reading comprehension ability, as this study found that this population improved their self-explanation quality, learned the strategies, and improved their short answer comprehension scores. The interviewees agreed that these strategies were useful and were likely to apply them to their content area readings. This study also further supports the possibility that applying ITS to training L2 reading comprehension strategies at higher institutions can be effective, because this population's L2 proficiency qualifies them to communicate in L2 within the iSTART-ME environment, and they have strong desire to get explicit strategy training to improve their L2 comprehension. This study may also make the assumption that explicit strategy teaching tends to help the L2 adults with L1 literacy to transfer their high level (e.g. critical reading) skills or strategies to L2 reading. Future research may test this assumption. Finally, because of the relationship between L2 reading and writing, iSTART-ME may also have the potential to improve L2 learners' writing ability, as exhibited by the participants in this study. This effect needs to be explored further in future studies.

RQ 2: What Are Participants' Perceptions of Their Learning Experience with iSTART-ME?

RQ2 Design Summary

To answer this question, this study conducted a pre and posttest measure of motivation (MSLQ) to compare participants' motivation change before and after use of iSTART-ME. This question was further measured by use of the ATTAS survey after the intervention to understand participants' perceptions of the animated conversational agent

in iSTART-ME and was triangulated by the one-on-one and focus-group interviews to elicit participants' perceptions of their learning gains and experiences with the iSTART-ME system.

RQ2 Results Summary

Within-subject t-test results showed that participants' motivation level after the intervention was significantly higher than their level before the intervention ($t_{df=33} = 4.86$, $p < .001$), with a large practical significance ($d = .86$). This change was also reflected in the interview data. The ATTAS survey results further indicated that participants held positive attitudes toward iSTART-ME ($M = 4.73$, out of 6 points), indicating that participants tended to "agree" that the tutor was effective. This was also confirmed by the interviews. The themes of the interview were that the learning with iSTART-ME was interesting and successful, although occasionally the participants felt confused and frustrated with the system.

Finding 4: Learning Motivation Improved

This study confirmed the findings of previous studies related to iSTART-ME (e.g. Jackson & McNamara, 2013; Jackson, Dempsey, Graesser, & McNamara, 2011; Jackson & McNamara, 2011), which found that participants' motivation for learning was higher after interacting with iSTART-ME. The pre- motivation mean score in the three constructs of Self-Efficacy, Task Value, and Intrinsic Goal Orientation was 4.75, and the post Mean was 5.42. The effect size $d = .86$, indicating a large effect.

The Task Value survey construct in the post survey contained questions like "I think I will be able to use what I learned in this course in other courses" and "I like the content of this course". Higher scores in these items indicated that both the participants'

current learning gains and their potential future endeavor to apply these strategies. These findings were supported by comments from the interviews. All the interviewees recounted the strategy names and definitions, and commented that these strategies were easy to learn, and that they wanted to use the system to learn other strategies.

The Intrinsic Goal Orientation construct contains questions like "I prefer course material that arouses my curiosity, even if it is difficult to learn" in the pre-survey versus "The course material aroused my curiosity" in the post-survey; "I prefer challenging materials so I can learn new things" in the pre-survey versus "The course material challenged me" in the post-survey. Participants significantly improved their intrinsic goal orientation after the intervention. This is important because the participants' curiosity in learning the strategies increased and they wanted to take further challenging courses. This finding was confirmed by the interview results. The majority of the interviewees directly asked if we had similar programs and said they would like to learn more strategies.

Self-Efficacy scores on the MSLQ survey did not appear to change, although the construct mean score increased after the iSTART-ME training (M_{pre}=5.30, M_{post}=5.38, out of 7). Because both means on this scale were high, there may have been a ceiling effect. The prior self-efficacy mean score was the highest among the three constructs. Future studies may examine whether those with low self-efficacy scores might benefit from the training.

Although the construct of Control of Learning Beliefs on the MSLQ motivation survey was excluded from the above data analysis due to the low reliability of this construct in this study, the other three constructs of Intrinsic Goal Orientation, Task Value, and Self-Efficacy as individual predictors of motivation indicated that the

participants' motivation was enhanced through the training. These findings suggest that the materials and the tutoring methods were in these participants' zone of proximal development (Vygotsky, 1978), and that their curiosity on this topic was aroused. These findings are important because motivation, especially intrinsic motivation, not only impacts learners' current learning process, but also sets the foundation for continuous future efforts (e.g. Ryan & Deci, 2000; Pintrich & Schrauben, 1992; Benware & Deci, 1984).

Finding 5: Positive Perceptions of the Conversational Agent

The Attitude toward Tutoring Agent Scale (ATTAS) was administered after the intervention to investigate the perceived efficacy of the conversational agent built into iSTART-ME. The ATTAS includes statements such as "The tutoring system encouraged me to think for myself" and "The tutoring system responded effectively to my input." The average score in this study of 4.73 (out of 6.00) indicated that the participants held a positive attitude, and tended to agree with the statements. This result was confirmed by the interview results that participants liked the iSTART-ME program, and said that the reading strategies and self-explanation were easy to learn. Because people expect to follow human social interaction rules when interacting with computer characters (e.g., Reeves & Nass, 1996), learners' positive attitudes towards agents indicate the effectiveness of animated conversational agents in much the same way that student ratings of instructors do. High scores are evidence of enhanced effectiveness of dynamic tutoring conversations or interactions and learning gains (Moreno, Mayer, Spires, & Lester, 2001). This finding from the affective dimension reflected the appropriateness and effectiveness of iSTART-ME training for international L2 learners. This result was

also consistent with Jackson & McNamara (2013)'s finding that high school students (L1) held positive attitudes and enjoyment in learning with iSTART-ME. One of the interview themes (*learning with iSTART-ME was interesting and successful)* corresponded well to findings 1, 4, and 5. All seven (100%) interviewees wanted to learn other strategies with iSTART-ME or similar computer programs. These findings are important because if participants like it, they will interact with it actively in the intervention process, and they may develop automaticity in strategy application in their future content area readings.

Finding 6: Interactions to Be Perfected

Although participants were positive about the intervention overall, there were areas for improvement. Participants in the interviews wished to see a "perfect" self-explanation for each sentence that they explained so that they could compare and inductively reason out the methods to do a better self-explanation or to improve their self-explanation ability, even though they knew that the self-explanation was open-ended. Therefore, they liked the "Showdown" game the most, because they could see the opponent's self-explanation and they could compare their self-explanations with their opponents' and induce the methods to improve their own self-explanations. Participants further expressed the desire to know how and why they did well or poorly in their self-explanations. They reported that the current feedback of their self-explanation scores and the prompts to push them to connect the sentence to previous sentences and their common knowledge were insufficient. If they received the feedback on how and why they could improve their current self-explanation together with illustrations, they felt they would be learning more because they were learning from their own mistakes and worked

examples. The effect of this type of learning has been confirmed effective by previous studies (e.g. Chandler & Sweller, 1991; Cooper & Sweller, 1987; Ward&Sweller, 1990; Atkinson, Derry, Renkl, & Wortham, 2000; Kalyuga, Chandler, Tuovinen, & Sweller, 2001). This might be a direction for ITSs to develop to a more advanced level with new technologies available in the future.

The majority of the interviewees (71.4%) also complained that the opponent was too strong in the games and they were often beaten by the opponent. It is not clear what caused this imbalanced competition. Because the opponents were L1 high school students and college students, this complaint might indicate that the participants' self-explanation quality after three hours training was lower than that of the current L1 high school students and college students who received at least eight hours training. It is unclear how long would it take the L2 participants to create SEs that were as good as L1 participants' SEs, suppose there was a discrepancy between L1 and L2 students' self-explanation quality, as shown in Showdown game. Or such a discrepancy is hardly to be eliminated. Of course, there existed a possibility that interviewees came across strong opponents in the Showdown game by chance. More research needs to be conducted to explore these questions. One interviewee complained that the opponent won the game with a nonsensical self-explanation, and that it was not a fair game. Because one of the important features of a game is fairness, this complaint demands special attention. Considering the system scoring algorithm had a significant and substantial agreement level (inter rater agreement Kappa =.646) with human experts on the self-explanation assessment (Jackson, Guess, & McNamara, 2010), what caused this feeling of unfairness? What was the probability that an iSTART-ME user would come across an

unfair judgment in a game? What proportion of the users is at risk to experience such a case? Was it the participant's misconception that their victorious opponent's self-explanation was nonsense? All these questions need further qualitative and quantitative investigations to find the answers and further inform future L2 instruction and iSTART-ME improvement.

Another problem reported was about the administration of game playing. Perhaps because not all participants had had rich game playing experiences in the past, some participants reported they did not exactly know how to play Map Conquest well. In future studies, every participant should be trained to play a trial version of the games to make sure they understand how to play each game before the formal study begins. It was one of the limitations of this study that no formal game play training was conducted to ensure that every participant was familiar with the games, although the assistants did explain the rules of the game before each game started. The interview results indicated that the current game instruction itself might not be enough for those participant who had never played games.

RQ 3: What Factors Account for Changes in Reading Comprehension and Self-Explanation?

RQ3 Design Summary

Spearman and Pearson correlations were conducted to test the relationship between individual's characteristics and self-explanation gain scores and comprehension gain scores, including these independent variables: reading comprehension ability, self-explanation ability, metacognitive strategies, and game hours. Linear regression tests were conducted with self-explanation gain scores or comprehension gain scores as the

dependent variables, and the factors which were significantly correlated with self-explanation gain scores and comprehension gain scores as independent variables.

RQ3 Results Summary

Pre self-explanation scores were negatively correlated with self-explanation gain scores, explaining 28.6% of the total variance of the self-explanation gain scores. Pre self-explanation scores were also negatively correlated with the *relative* self-explanation gain scores. The SEWC gain scores were positively correlated with self-explanation gain scores, but this became non-significant when pre self-explanation scores were entered into the regression model. Pre- short answer comprehension scores were negatively correlated to the short answer comprehension gain scores, explaining 14.9% of the total variance of the short answer comprehension gain scores. However, pre- short answer comprehension scores were not significantly correlated to the *relative* short answer comprehension gain scores. Other variables were not significantly correlated to self-explanation or short answer comprehension gain scores in the regression model.

Finding 7: **Lower Reading Ability Students Benefited the Most**

Because the sample size was not over 50 (Harris, 1985; Green, 1991) as a rule of thumb, the research did not have enough power (>.80) to identify a small sized coefficient (≤.30; Vanvoorhis & Morgan, 2007). Therefore, the significance test results should be interpreted with caution because it may have not identified all the significant relationships. The negative correlation between prior self-explanation scores and self-explanation gain scores indicates that lower reading-ability students benefited most from the iSTART-ME training; higher reading-ability students also increased their self-explanation scores and comprehension scores, but not to the same extent.

Implications

Implications for L2 Reading Strategy Training

Necessity of offering strategy training. All the interviewees had previously learned about some reading strategies, but none of them had received systematic training and practice in applying these strategies. Their positive feedback and test results suggest the importance for higher institutions to provide strategy training to incoming international L2 learners. Reading support resources on campuses are scarce, yet college reading imposes significant challenges (Terry, 2007; Wilkins, Hartman, Howland,& Sharma, 2010; Snow, 2002). This study suggests that iSTART-ME can be a good resource to boost L2 leaners' interest in learning and their ability to apply reading strategies to improve their reading ability. Providing comprehensive strategy training in reading through effective online ITS may help the L2 learners better comprehend college course materials. In addition to the reading strategies covered by iSTART-ME, additional strategies used before reading and after reading may also benefit students further, as the participants expressed in the interview. Future ITSs may want to add additional strategies and test the cumulative and individual effects.

ITS' potential for strategy training. This study showed that incoming L2 college students enjoyed the training with iSTART-ME, and that they expressed the desire to learn more strategies and apply what they had learned through iSTART-ME to their future courses. As a result of iSTART-ME, participants successfully learned reading strategies and realized that self-explanation practice not only improved their reading ability but also their writing ability. This indicates that it is both practical and desirable to

use ITSs like iSTART-ME to provide strategy training for college-level L2 students, especially those students with lower reading ability.

For large-scale strategy training, iSTART-ME or other forms of ITS may provide L2 learners the convenience and motivation to improve their reading comprehension and strategies. Because the results of this study was as effective as prior strategy training for L2 students using humans (Taylor, Stevens, & Asher, 2006), iSTART-ME has the potential to improve reading comprehension for an unlimited number of L2s without the necessity of attending training in person. Furthermore, self-explanation training with iSTART-ME may benefit the L2 learners' writing ability in addition to their reading ability

Implications for Future L2 Reading Strategy Training Study

Individual differences. This study also examined the influence of an individual's prior metacognitive strategy applications, vocabulary knowledge, comprehension ability, self-explanation ability, motivation, and gaming hours on self-explanation and comprehension improvement. Prior self-explanation and reading comprehension ability were negatively related to self-explanation and comprehension gain scores respectively. They were also significant variables in explaining the variance of self-explanation gain scores and comprehension gain scores. However, due to the limited power of this study with 34 participants to identify the relationships, future studies need to examine these factors again to explore the relationships with bigger samples. Moreover, prior domain knowledge, L1 reading ability, and cognitive strategies may need to be examined too, taking into account the influence of prior domain knowledge and strategy applications in comprehension (McNamara, 2004; Pressley, 1998, 2000, 2002, 2006), and the linguistic

interdependence theory (Goodman, 1971; Cummins, 1979, 1991). L2 learners with various motivation levels and linguistic ability may also need to be further explored when they interact with each other.

More strategies. SERT is composed of five strategies (monitoring, paraphrasing, prediction, elaboration, and bridging) with the self-explanation approach, so this study only examined the effect of this approach and strategy training with iSTART-ME on reading comprehension and self-explanation quality improvement. These strategies and other strategies could be used not only during reading, but also before and after reading. This would include things, such as goal-setting, scanning titles and subtitles, summarizing the reading content, and evaluating the information for reading goals during before- and after-reading stage. Interviewees in this study explicitly expressed the desire to learn other strategies. Due to the limited research on L2 strategy training (Grabe, 2007; Taylor, Stevens, & Asher, 2006), more research is needed to further study the training effect of different strategies at different reading stages on reading comprehension.

Assessment. The effect of iSTART-ME training on score improvements in standardized reading comprehension assessments differed from that in short-answer reading comprehension assessments, but the interviewees in this study said that the training was effective, and expressed their desire to apply the strategies to their future reading. This different effect on different assessment instruments replicated the finding by Carrell, Pharis, and Liberto's (1989) on L2 metacognitive strategy training. Future research may further explore the effect of strategy training on improving students' scores on different comprehension assessments, which may reveal more information to understand comprehension assessment, because different reading comprehension

assessment results may be influenced by many factors. Secondly, this study only tested the immediate effect after iSTART-ME training; retention was not measured. It is not clear if the participant's higher self-explanation and reading comprehension ability after the training would remain stable over a period of time, or whether they would apply these strategies in future reading, although there is no reason to expect that it would not replicate the findings from studies of L1 learners. Future studies should examine these questions.

Thirdly, interviewees clearly stated that this study not only improved their reading ability, but also their writing ability. Although the post self-explanation quality was significantly better than the pre self-explanation quality, this study did not directly examine the effect of this training on participants' writing ability improvement. This should be explored in the future.

Fourthly, this study did not compare the self-explanation quality of college L2 learners and that of the L1 high school students who formed the basis of the competitors for games in iSTART-ME, but the majority of interviewees in this study said that the opponents (L1 high school students) always beat them in playing games. This could of course simply reflect the innate advantage L1 learners have over L2 learners in terms of vocabulary, domain knowledge, etc., but it may need to be further analyzed in future studies together with the analysis of possible linguistic feature changes in a longer training process. Therefore, future studies may design different assessments to reveal the effect of iSTART-ME training on short- and long-term reading comprehension and writing skills change.

L1 and L2 comparison. The majority of participants of this study were Chinese students (30 out of 34), and their L1 reading ability and strategy application in L1 were not studied. It is possible that their L1 reading strategy and ability had an impact on their L2 strategy training effect, despite them all meeting the same level of proficiency required by the high school graduation standards. Because of the impact of L1 on L2 reading (Coady, 1979, Yamashita, 2002, Van Gelderen, et al, 2004, 2007), future studies may need to include participants with different L1s and with different L1 abilities to test the impact of L1 on L2 reading and/or writing, and to further test the language interdependence theory and strategy training effect on L2 learners. For example, strategy studies with L1 and L2 learner comparison groups or L2 strategy studies with different L1 learner groups could be conducted. Van Geldereen (2007) found that metacognitive strategy had a significant contribution to L2 learners' reading ability, but that when L1 reading comprehension was added to the regression model, metacognitive strategy's contribution to L2 reading ability became insignificant. Therefore, including the variable of L1 reading skills in the future L2 reading strategy training study may help researchers to better understand the relationship of L1 reading skill and L2 reading strategy training effect.

Game-based ITS training design. This is the first study with iSTART-ME, a game-based ITS, to train L2 reading strategies, so more similar studies must be done to fully understand the potential of L2 strategy training and ITS in L2 learning so as to reveal more features of L2 learning. This study trained L2 learners for about three hours in two sessions. The time for each game was predetermined according to the average reported time for participants to become tired of a game during the previously conducted

pilot study, and the texts were automatically suggested to the participants by the system. The participants were not given an opportunity to explore all game features, like using iBucks to change their given avatar attires. The entire training time was limited to three hours (5.5 hours total, including pre- and posttests and interviews) due to the resources available for this dissertation. Future studies may want to extend the total training time with flexibility for participants to explore other features of the games, as Jackson and McNamara (2013) trained L1 learners for a minimum of eight hours, and allowing participants the freedom to choose their own games. By extending training time, the learners' L2 linguistic feature changes in self-explanation, along with the effect of training on self-explanation quality and comprehension ability improvement, could probably be identified through the training process. Studies may also want to vary the time for each game in order to test the effect of each game on L2 reading and motivation change. Thirdly, studies may test the effect of different kinds of games on L2 learners with different levels of L2 ability. Finally, studies may want to vary the text difficulty and genre to test the effects of strategy training on L2 reading, considering that McNamara (1996) found that the strategy training was more effective for L1 readers with less cohesive and challenging texts. Different research designs may lead to different training effects and help researchers to understand more about L2 reading comprehension and its interactions with ITS design, and instructional model design.

Limitations

Small Sample Size

This study had a sample of 34 L2 learners, which results in insufficient power to identify potential relationships among different factors such as prior linguistic ability and

comprehension ability. Only strong relationships were identified in this study, such as pre self-explanation scores with self-explanation gain scores, and pre-reading comprehension scores with comprehension gain scores. Moreover, among the 34 participants, 30 were Chinese students, so the results might be less likely to be generalized to other L2 learners with L1 backgrounds other than Chinese.

Mortality

In this study 56 L2 learners participated in the first session, but 20 did not show up for the second session for various reasons. The interviewees who completed the whole study gave positive comments on this iSTART-ME training system in terms of learning gains and interests, which was confirmed with survey and assessment results. However, it remains unknown how many of the dropped students lacked interest in this training system, and what they disliked about this training program. There could be common characteristics shared by this group that make interventions like iSTART-ME less likely to be universally adopted by L2s, although no patterns could be found in those that dropped out.

Limited Comprehension Assessment

The onsite PSAT for vocabulary and comprehension ability test was split for pretest and posttest as explained in the methodology chapter, which may arouse the doubts for the validity of the standardized tests in this study. Additionally, this study did not run a retention test, because incoming international L2 students were extremely busy at the beginning of a semester with many orientation activities and preparation work of school start. It is unknown, therefore, whether the improved reading and self-explanation

ability will transfer to college level text reading in terms of its different text length and complexity from the training materials.

Contributions

To L2 Strategy Training Field

The purpose of this study was to investigate the effect of L2 strategy training with iSTART-ME. The findings of this study significantly expanded the evidence of L2 training effect on reading comprehension through the self-explanation training with iSTART-ME. It further examined the L2 learners' motivations and perceptions of the learning process, which scholars had not previously studied. This is the first study to research the effect of teaching L2 participants the self-explanation approach with the application of five reading strategies with an ITS, which opens potential opportunities for training L2 reading strategies through ITSs to the entire L2 population by virtue of removing the need for a human-led strategy training class.

This study also suggests an opportunity for L2 learners to learn and practice self-explanation approach as one of their future college course learning techniques so as to deepen their comprehension of the course materials and problem solving abilities (Chi et al., 1989; 1994). Moreover, this study tried to connect personal features like intrinsic motivation, values, and self-efficacy to strategy training results, which few L2 studies had ever examined. Most previous studies in L2 only examined the relationship between reading comprehension scores and strategy training. This study expands previous conceptual approaches by examining more personal features. Furthermore, due to the rich data saved in ITS database, this research may stimulate more L2 or ITS researchers' interest to explore the available log data so as to understand the language learning better.

To ITS Field

This was the first study to examine the effect of iSTART-ME strategy training on L2 learners, which expanded the research with ITS beyond L1 field. The positive findings of this study relating iSTART-ME to promoting L2 learners' learning interest and learning gains has advanced the current ITS research to including L2 learners. The interview feedback from participants in this study also provided technical information as well as global visions for future ITS design from users' perspective, such as technical glitches, opponent's response level in a game, and the demands for more examples with trials in practice. Similar findings were impossible without the mixed methods approach employed in this study, which previous studies in ITS field seldom used. This approach also allowed this study to triangulate the research results and understand the stories behind the numbers, thus adding to the current ITS research and L2 research.

Recommendations

International L2 students admitted to American higher institutions face reading problems, because the amount of reading material is dramatically increased from the amount of L2 materials they have ever read in their home country; furthermore, these students are expected to understand the materials independently, without any assistance from their teachers. Teaching incoming international L2 students successful strategies for comprehending expository texts can be of great assistance for them in keeping pace with their L1 peers in learning domain knowledge. An ITS, such as iSTART-ME, is a good tool for reaching large numbers of students at low cost.

In the L2 research field, more studies are called for to test the effects of reading strategy training, and especially the lasting effects of specific strategy or strategy

instructional models (combinations of strategies) in terms of different first languages and different L2 proficiency levels. It is also important to involve more variables, such as prior knowledge and assessment tools, in studying the effect of strategy training. Likewise, involving L2 students in expressing their understanding of text through writing may be a tremendous facet of reading, writing, and learning, as suggested in content area learning in L1 (Irvin, Buehl, & Klemp, 2007). To sum up, the positive effect of self-explanation has been widely studied in different disciplines; it is time to study the effect of it in the L2 world.

APPENDICES

APPENDIX A

Reading Comprehension Test with Open-Ended Questions for Pre-Test

Please read the following passage, and answer the questions after it with one or two sentences. Please answer them as accurately as possible.

Heart Disease

The heart is the hardest-working organ in the living body. Any disorder that terminates the body's blood supply is a threat to life. More people are killed every year in the U.S. by heart disease than by any other disease.

A congenital disease is one with which a person is born. Most babies are born with perfect hearts, but something can go wrong for approximately one in 200 cases. Sometimes a valve develops the incorrect shape causing it to be too tight or fail to close properly. Sometimes a gap is left in the septal wall between the two sides of the heart. When a baby's heart is badly formed, it cannot work efficiently. The baby's blood does not receive enough oxygen and cannot eliminate carbon dioxide through the lungs. The blood becomes purplish, and the baby's skin looks blue. The baby is in danger of suffocating.

Diseases also cause the heart to form improperly. For example, the disease called rheumatic fever follows a sore throat caused by bacteria called streptococci. The tissues of the heart become inflamed and, if badly affected, can cause it to stop. Usual y the heart recovers, but the heart valves are left with scars. Years later, they may fail to work properly and cause the heart to stop.

The most common heart problem is a heart attack, or coronary thrombosis, which is caused when a coronary artery becomes blocked. The blood vessels that extend across the heart and supply it with blood are called the coronary arteries. They give the heart the oxygen it needs to carry on working. The blockage of a coronary artery is usual y caused by a thrombus, or blood clot. Whether heart disease is congenital, caused by other diseases, or the result of a blood clot, it is a very serious problem that requires medical attention.

1. If a person has a congenital disease, at what stage of life did it most likely occur?
2. How often does congenital heart disease occur?
3. What are the consequences of a malformed heart valve?
4. If a baby's skin begins to turn blue, what is the most likely physical cause?
5. What causes rheumatic fever?
6. Why is rheumatic fever considered such a serious condition?
7. What is the main function of the coronary arteries?
8. What causes a heart attack?

APPENDIX B

Self-explanation is one way that helps you to better understand the meaning of the entire text, as well as the meaning of the paragraph.

e.g. Polluted rain results from large amounts of sulfur oxides and nitrogen oxides combining with rainwater.

Self-Explanation:

This sentence is saying that some rain is polluted, and pollution comes from sulfur oxides and nitrogen oxides. These chemicals must be in the air and so, when the rain comes down, it must pick up the chemicals as it falls. The chemicals must come from things like cars.

As you can see, self-explanations are not mere restatements of sentences. Rather, self-explanations contain your understandings about the meaning of the information.

Please give a self-explanation for each of the following FOUR numbered and underlined sentences, and type your self- explanation in the numbered boxes.

Heart Disease

The heart is the hardest-working organ in the living body. Any disorder that terminates the body's blood supply is a threat to life (1). More people are killed every year in the U.S. by heart disease than by any other disease.

A congenital disease is one with which a person is born. Most babies are born with perfect hearts, but something can go wrong for approximately one in 200 cases. Sometimes a valve develops the incorrect shape causing it to be too tight or fail to close properly(2). Sometimes a gap is left in the septal wall between the two sides of the heart. When a baby's heart is badly formed, it cannot work efficiently. The baby's blood does not receive enough oxygen and cannot eliminate carbon dioxide through the lungs. The blood becomes purplish, and the baby's skin looks blue. The baby is in danger of suffocating.

Diseases also cause the heart to form improperly. For example, the disease called rheumatic fever follows a sore throat caused by bacteria called streptococci(3). The tissues of the heart become inflamed and, if badly affected, can cause it to stop. Usual y the heart recovers, but the heart valves are left with scars. Years later, they may fail to work properly and cause the heart to stop.

The most common heart problem is a heart attack, or coronary thrombosis, which is caused when a coronary artery becomes blocked. The blood vessels that extend across the heart and supply it with blood are called the coronary arteries. They give the heart the oxygen it needs to carry on working. The blockage of a coronary artery is usual y caused by a thrombus, or blood clot. Whether heart disease is congenital, caused by other diseases, or the result of a blood clot, it is a very serious problem that requires medical attention(4). [Numbered boxes omitted]

APPENDIX C

Reading Comprehension Test with Open-Ended Questions for Post-Test

Please read the following passage, and answer the questions after it with one or two sentences. Please answer them as accurately as possible.

Red Blood Cells

Red blood cells have the vital role of carrying oxygen to all of the cells in the body. They also pick up waste carbon dioxide for removal. These cells are the most numerous of the blood cells. The disk shape of the red blood cells results in a large surface area, which enables them to be efficient at gas diffusion.

Red blood cells contain a large, complex protein called hemoglobin. Hemoglobin binds to the oxygen and carbon dioxide that the red blood cells transport. Each red blood cell contains about 250 million hemoglobin molecules, each carrying four molecules of oxygen. Hemoglobin also contains iron, which gives blood its red color. Molecular oxygen can also be transported by another route, in dissolved blood plasma. However, oxygen is poorly soluble in water, so only about 1.5% is carried in dissolved form. Therefore, most oxygen is carried by hemoglobin.

Red blood cells lack a nucleus and the organelles found in other cells. Therefore, these cells cannot reproduce or repair themselves. Red blood cells live for about three or four months before being broken down in the spleen. Iron from the broken-down cells is returned to the bone marrow to be recycled into new hemoglobin.

Sometimes blood does not transport enough oxygen, resulting in a condition called anemia. This makes a person feel tired and weak. Anemia can result from too little iron in the diet, loss of blood due to injury or menstruation, or various medical conditions. One type of anemia, called sickle-cell disease, is characterized by red blood cells that are sickle-shaped instead of disk-shaped. The shape of the cells causes them to clog blood vessels, preventing oxygen from reaching muscles and other tissues.

1. How does sickle-cell disease get its name?
2. Explain why blood plasma is a poor carrier of oxygen?
3. Explain why the disk shape of red blood cells is advantageous for gas diffusion?
4. What causes a person to feel weak and tired in anemia?
5. How many oxygen molecules can be carried in each red blood cell?
6. What are the critical elements of regular body cells that enable these cells to reproduce or repair themselves?
7. In the production of hemoglobin, where does iron come from?
8. How does sickle-cell disease cause anemia?

APPENDIX D

Self-explanation is one way that helps you to better understand the meaning of the entire text, as well as the meaning of the paragraph.

e.g. Polluted rain results from large amounts of sulfur oxides and nitrogen oxides combining with rainwater.

Self-Explanation:

This sentence is saying that some rain is polluted, and pollution comes from sulfur oxides and nitrogen oxides. These chemicals must be in the air and so, when the rain comes down, it must pick up the chemicals as it falls. The chemicals must come from things like cars.

As you can see, self-explanations are not mere restatements of sentences. Rather, self-explanations contain your understandings about the meaning of the information.

Please give a self-explanation for each of the following FOUR numbered and underlined sentences, and type your self- explanation in the numbered boxes.

Red Blood Cells

Red blood cells have the vital role of carrying oxygen to all of the cells in the body. They also pick up waste carbon dioxide for removal. These cells are the most numerous of the blood cells. The disk shape of the red blood cells results in a large surface area, which enables them to be efficient at gas diffusion (1).

Red blood cells contain a large, complex protein called hemoglobin. Hemoglobin binds to the oxygen and carbon dioxide that the red blood cells transport(2). Each red blood cell contains about 250 million hemoglobin molecules, each carrying four molecules of oxygen. Hemoglobin also contains iron, which gives blood its red color. Molecular oxygen can also be transported by another route, in dissolved blood plasma. However, oxygen is poorly soluble in water, so only about 1.5% is carried in dissolved form. Therefore, most oxygen is carried by hemoglobin.

Red blood cells lack a nucleus and the organelles found in other cells. Therefore, these cells cannot reproduce or repair themselves(3). Red blood cells live for about three or four months before being broken down in the spleen. Iron from the broken-down cells is returned to the bone marrow to be recycled into new hemoglobin.

Sometimes blood does not transport enough oxygen, resulting in a condition called anemia. This makes a person feel tired and weak. Anemia can result from too little iron in the diet, loss of blood due to injury or menstruation, or various medical conditions. One type of anemia, called sickle-cell disease, is characterized by red blood cells that are sickle-shaped instead of disk-shaped. The shape of the cells causes them to clog blood vessels, preventing oxygen from reaching muscles and other tissues(4).

[Numbered boxes omitted]

APPENDIX E

Strategy Application Survey (Revised MARSI)

Please recall what strategies you have used (or describe what you did) in comprehending the above 3 passages. For example, when you didn't understand a sentence, a paragraph, or an entire passage, what did you do?

Please check ALL the strategies you have used in comprehending the above passages

Strategy	Selection	
1. I had a purpose in mind when I read.	Yes	No
2. I thought about what I knew to help me understand what I read.	Yes	No
3. I previewed the text to see what it's about before reading it.	Yes	No
4. I summarized what I read to reflect on important information in the text.	Yes	No
5. I thought about whether the content of the text fits my reading purpose.	Yes	No
6. I read slowly but carefully to be sure I understand what I'm reading.	Yes	No
7. I skimmed the text first by noting characteristics like length and organization.	Yes	No
8. I tried to get back on track when I lose concentration.	Yes	No
9. I adjusted my reading speed according to what I'm reading.	Yes	No
10. I decided what to read closely and what to ignore.	Yes	No
11. When text became difficult, I paid closer attention to what I'm reading.	Yes	No
12. I stopped from time to time and think about what I'm reading.	Yes	No
13. I used context clues to help me better understand what I'm reading.	Yes	No
14. I paraphrased (restated ideas in my own words) to better understand what I read.	Yes	No
15. I tried to picture or visualize information to help remember what I read.	Yes	No
16. I used typographical aids like bold face and italics to identify key information.	Yes	No
17. I critically analyzed and evaluated the information presented in the text.	Yes	No
18. I went back and forth in the text to find relationships among ideas in it.	Yes	No
19. I checked my understanding when I come across conflicting information.	Yes	No
20. I tried to guess what the material is about when I read.	Yes	No
21. When text became difficult, I re-read to increase my understanding.	Yes	No
22. I asked myself questions I like to have answered in the text.	Yes	No
23. I checked to see if my guesses about the text are right or wrong.	Yes	No
34. I tried to guess the meaning of unknown words or phrases.	Yes	No

APPENDIX F
Adapted MSLQ for Pre-Survey

Please select one that best describes your values and expectancy in THIS COURSE.

		Very untrue of me	Untrue of me	Somewhat untrue of me	Neutral	Somewhat true of me	True of me	Very true of me
1.	I prefer challenging materials so I can learn new things	1	2	3	4	5	6	
2.	I prefer course material that arouses my curiosity, even if it is difficult to learn	1	2	3	4	5	6	7
3.	The most satisfying thing for me will be trying to understand the content as thoroughly as possible.	1	2	3	4	5	6	7
4.	I think I will be able to use what I learn in this course in other courses.	1	2	3	4	5	6	7
5.	I like the subject matter of this course.	1	2	3	4	5	6	7
6.	If I study in appropriate ways, I will be able to learn the course material.	1	2	3	4	5	6	7
7.	It is my own fault if I don't learn the material in this course.	1	2	3	4	5	6	7
8.	If I try hard enough, I will understand the course material.	1	2	3	4	5	6	7
9.	I'm confident I can understand the basic concepts	1	2	3	4	5	6	7
10.	I expect to do well in this class	1	2	3	4	5	6	7
11.	I'm certain I can master the skills being taught in this class.	1	2	3	4	5	6	7
12.	Understanding the subject matter of this course is very important to me.	1	2	3	4	5	6	7

APPENDIX G
Adapted MSLQ for Post-Survey

Please select the one that best reflects you in this course.

		Very untrue of me	Untrue of me	Somewhat untrue of me	Neutral	Somewhat true of me	True of me	Very true of me
1.	The course material challenged me.	1	2	3	4	5	6	7
2.	The course material aroused my curiosity.	1	2	3	4	5	6	7
3.	I tried to understand the content thoroughly.	1	2	3	4	5	6	7
4.	I'm confident I understood the basic concepts taught in this course.	1	2	3	4	5	6	7
5.	I believe that I did well in this class.	1	2	3	4	5	6	7
6.	I'm certain I mastered the skills being taught in this class.	1	2	3	4	5	6	7
7.	I think I will be able to use what I learned in this course in other courses.	1	2	3	4	5	6	7
8.	I liked the content of this course.	1	2	3	4	5	6	7
9.	It is important to learn the content of this course.	1	2	3	4	5	6	7
10.	It was my own fault if I did not learn the material in this course	1	2	3	4	5	6	7
11.	If I tried hard, I could understand the course material.	1	2	3	4	5	6	7
12.	If I did not understand the course material, it was because of the computer system.	1	2	3	4	5	6	7

APPENDIX H

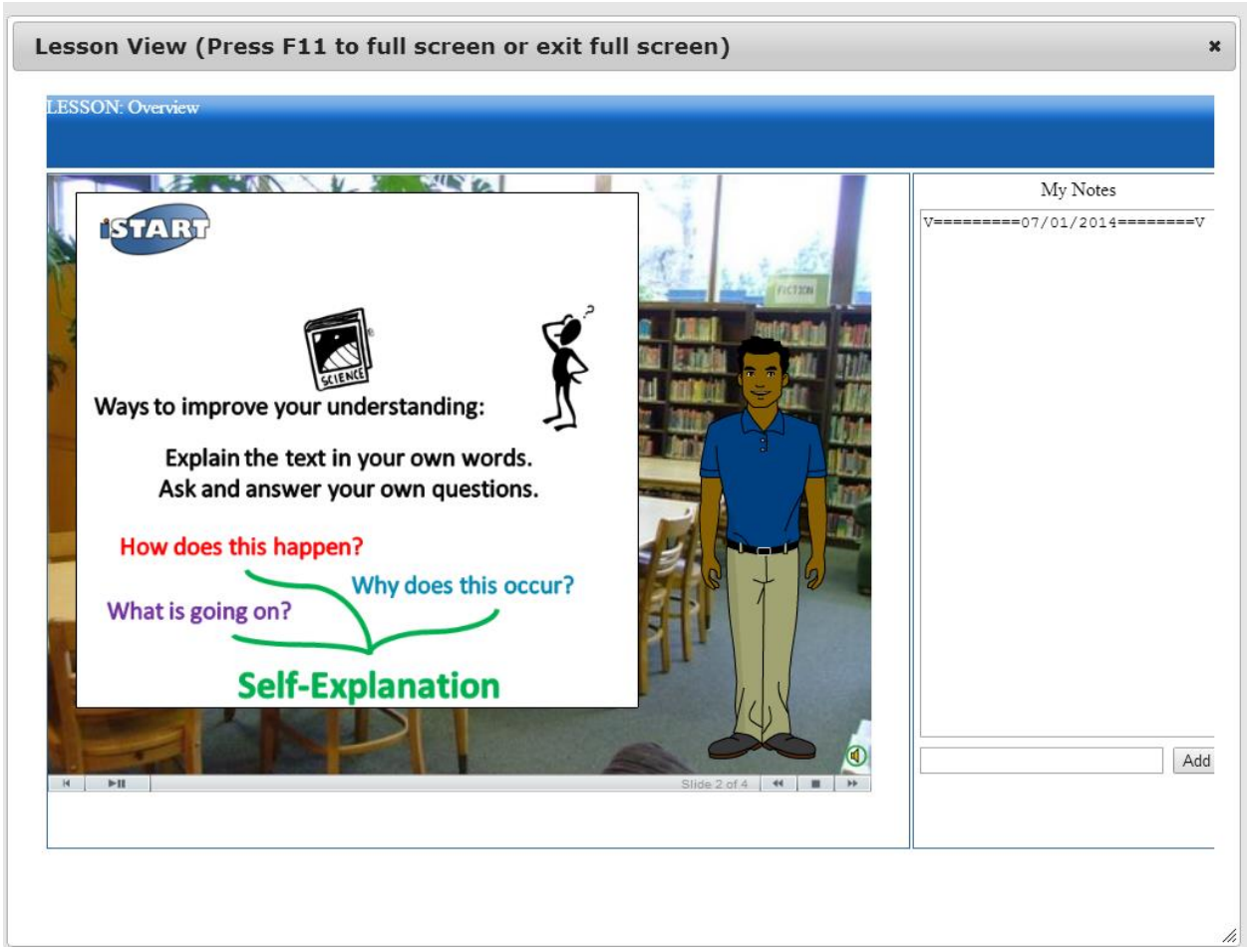
Adapted ATTAS (Attitude toward Tutoring Agent Scale)

Please select the one that best reflects you in this course.

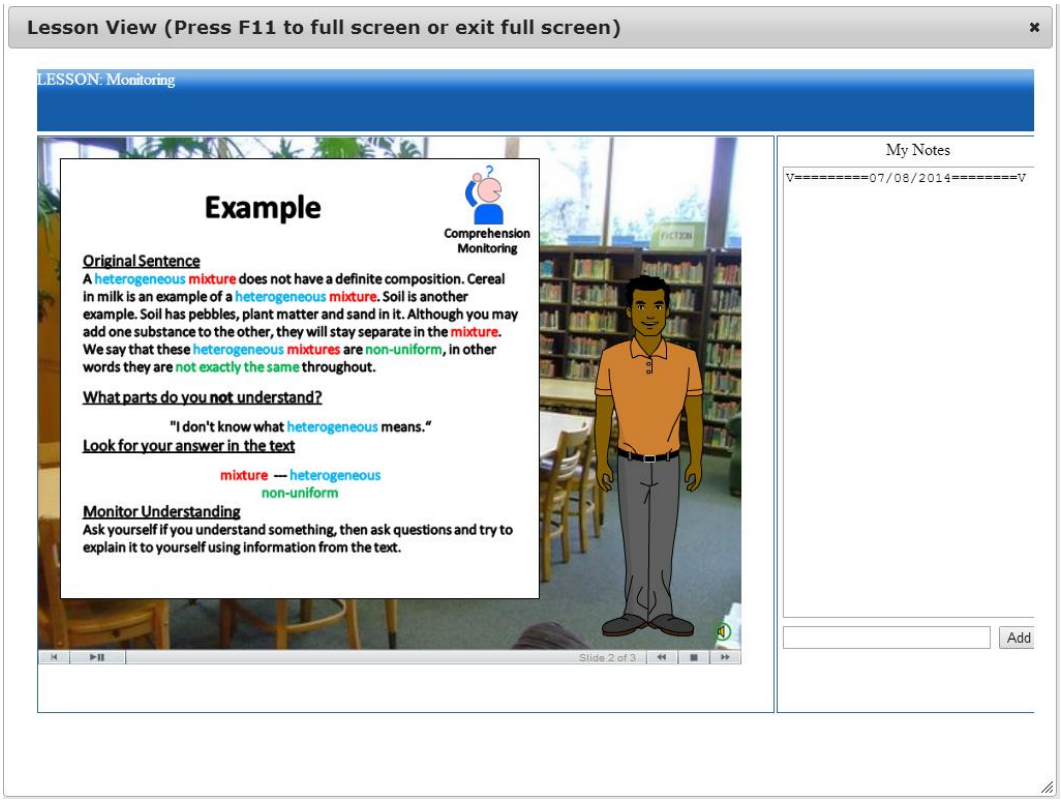
		Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1	The tutoring system provided helpful feedback	1	2	3	4	5	6
2	The tutoring system responded effectively to my input	1	2	3	4	5	6
3	The tutoring system encouraged questions and answers	1	2	3	4	5	6
4	The tutoring system encouraged me to think for myself	1	2	3	4	5	6
5	The tutoring system encouraged the development of my knowledge	1	2	3	4	5	6
6	The tutoring system seemed friendly towards me	1	2	3	4	5	6
7	The tutoring system seemed discouraging towards me	1	2	3	4	5	6
8	The tutoring system seemed impatient with me	1	2	3	4	5	6
9	The tutoring system made helpful comments	1	2	3	4	5	6
10	The tutoring system sensed when I needed help	1	2	3	4	5	6
11	The tutoring system increased my interest in the subject	1	2	3	4	5	6
12	The teaching style of the tutoring system held my interest	1	2	3	4	5	6
13	The tutoring system knows how to hold my attention when presenting material	1	2	3	4	5	6

APPENDIX I

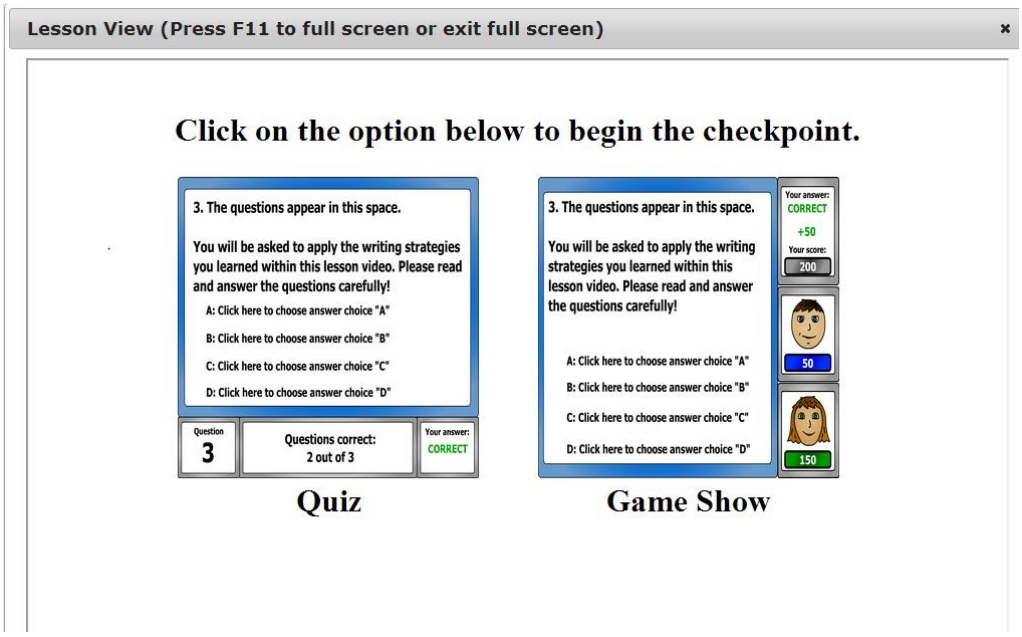
Snapshots in iSTART-ME



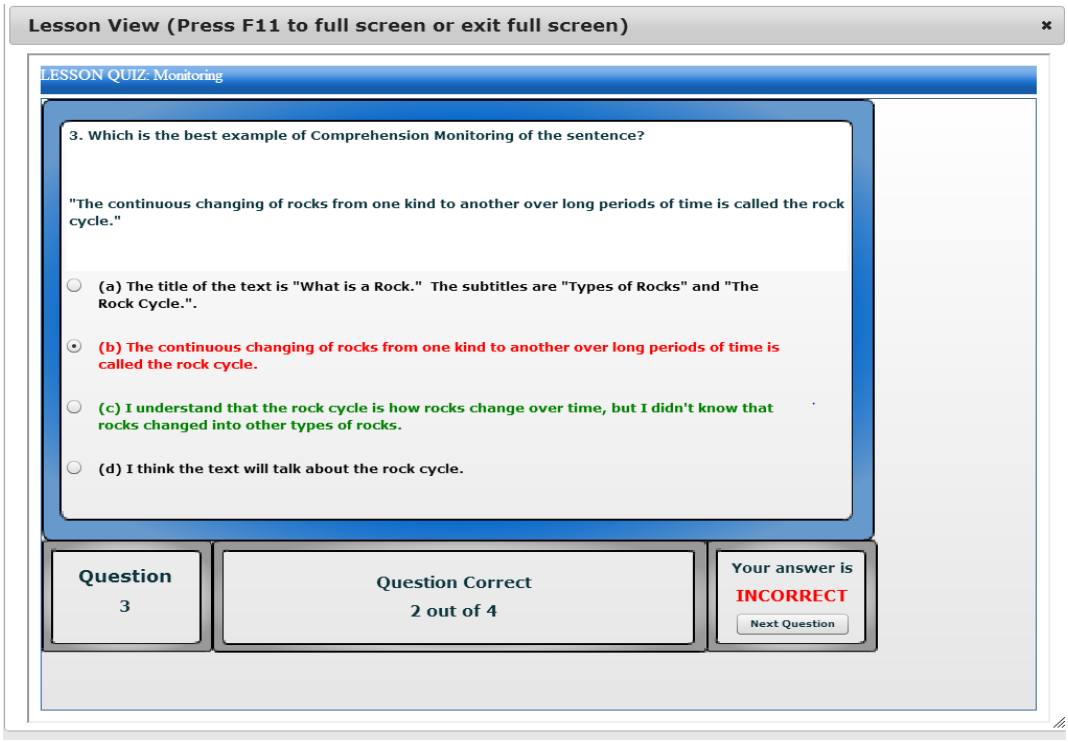
Appendix I. Figure 1. Overview Module Snapshot 1



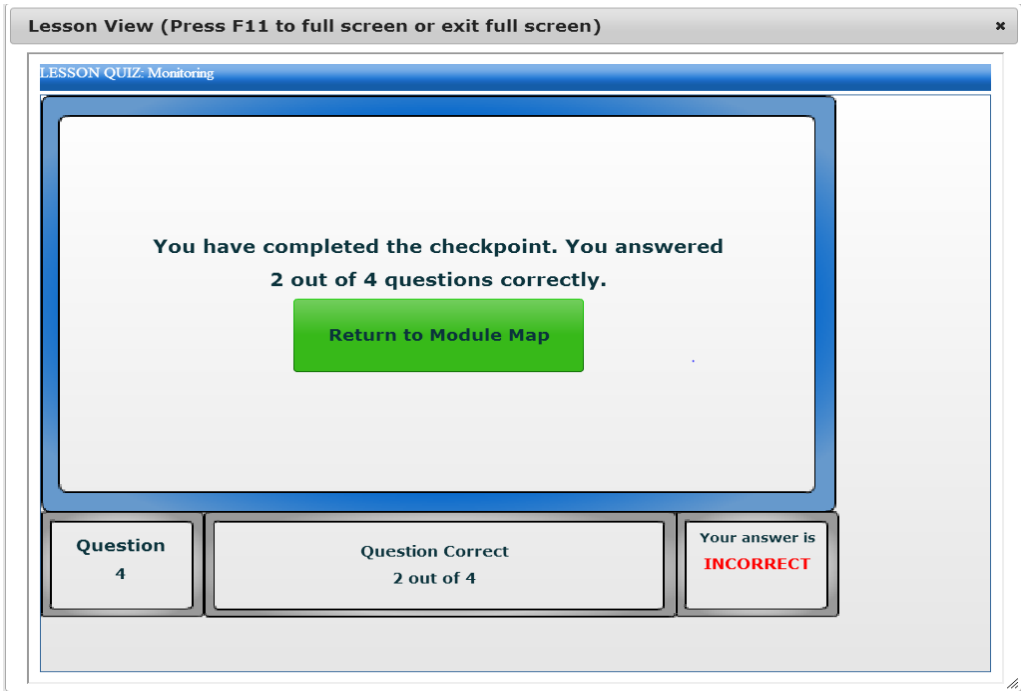
Appendix I. Figure 2. Monitoring Module Snapshot 1



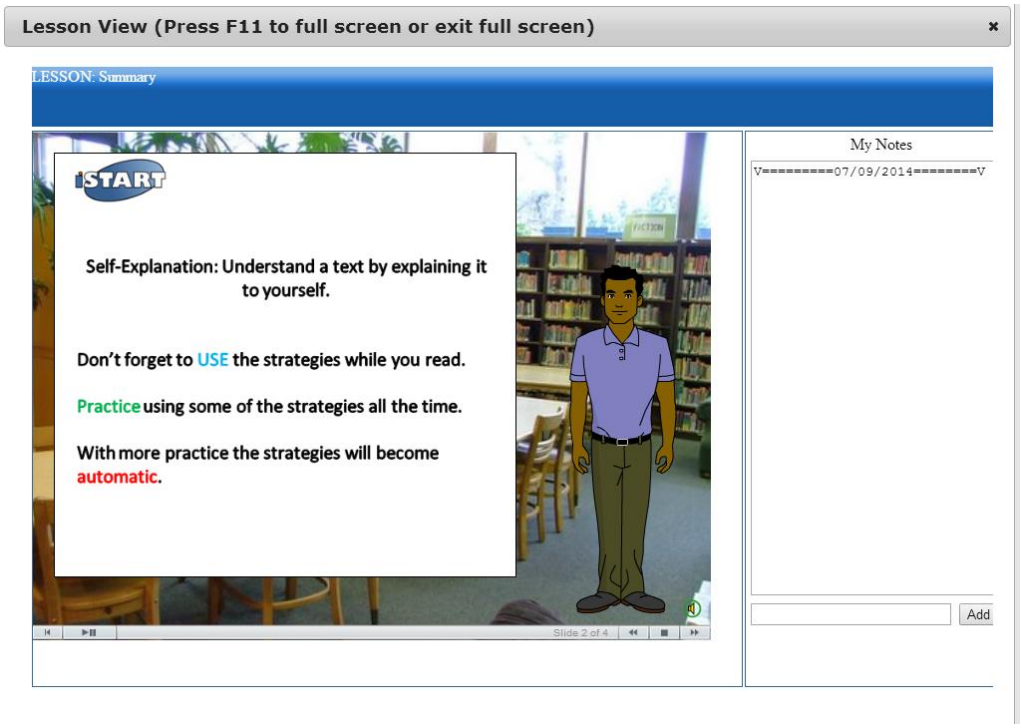
Appendix I. Figure 3. Monitoring Module Snapshot2



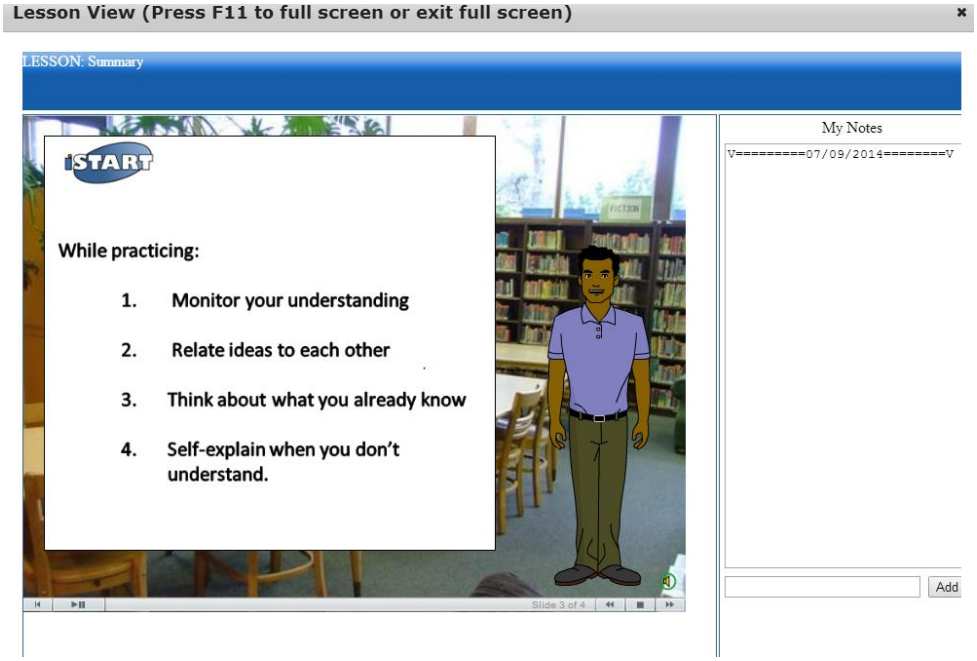
Appendix I. Figure 4. Monitoring Module Snapshot 3



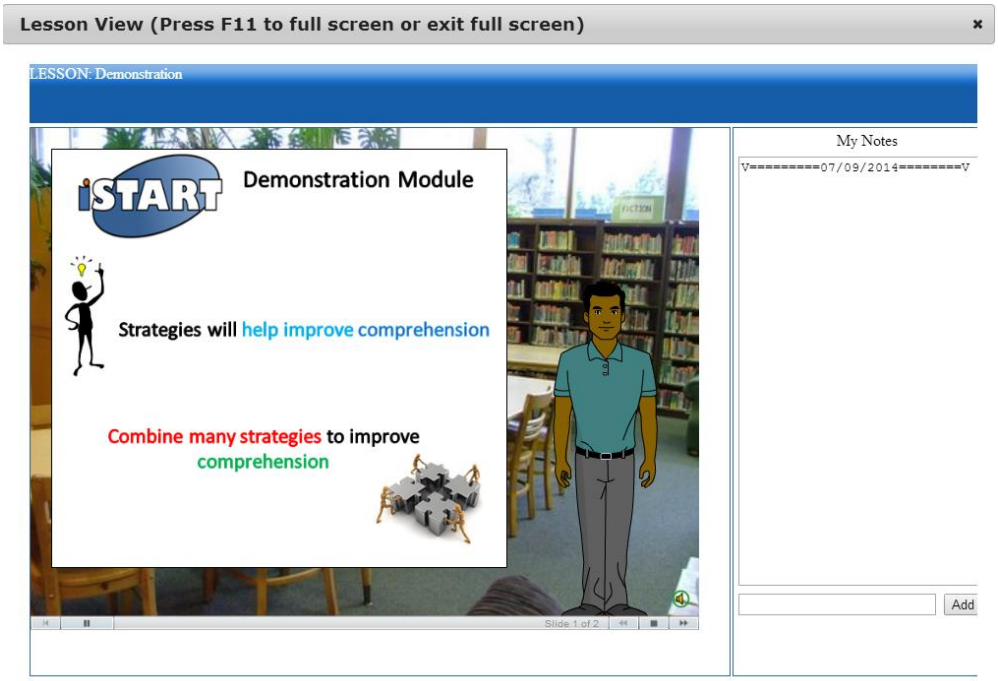
Appendix I. Figure 5. Monitoring Module Snapshot 4



Appendix I. Figure 6. Summary Module Snapshot 1

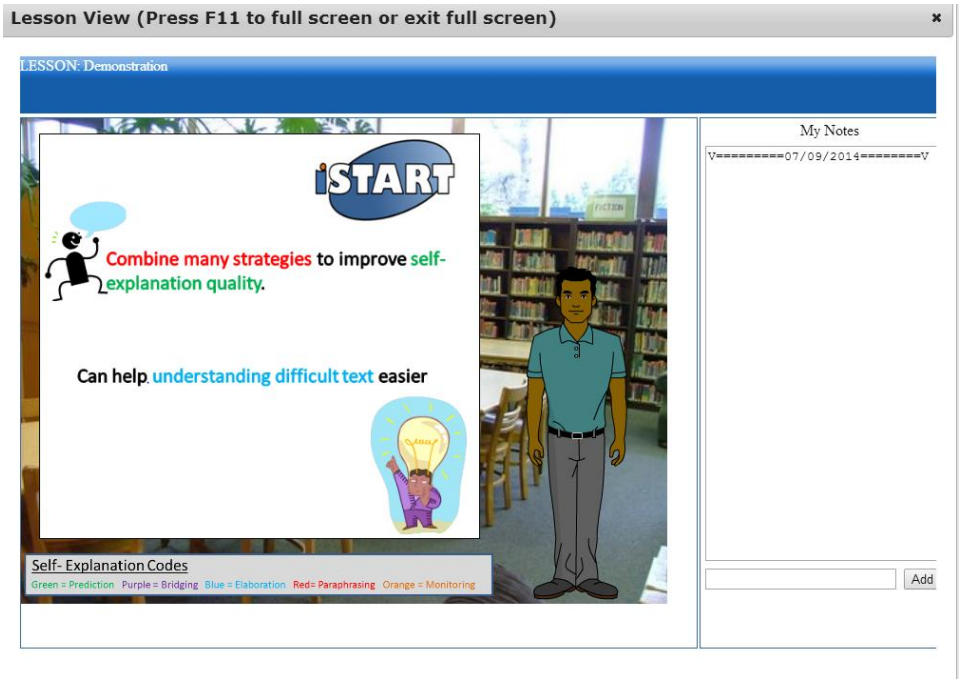


Appendix I. Figure 7. Summary Module Snapshot 2

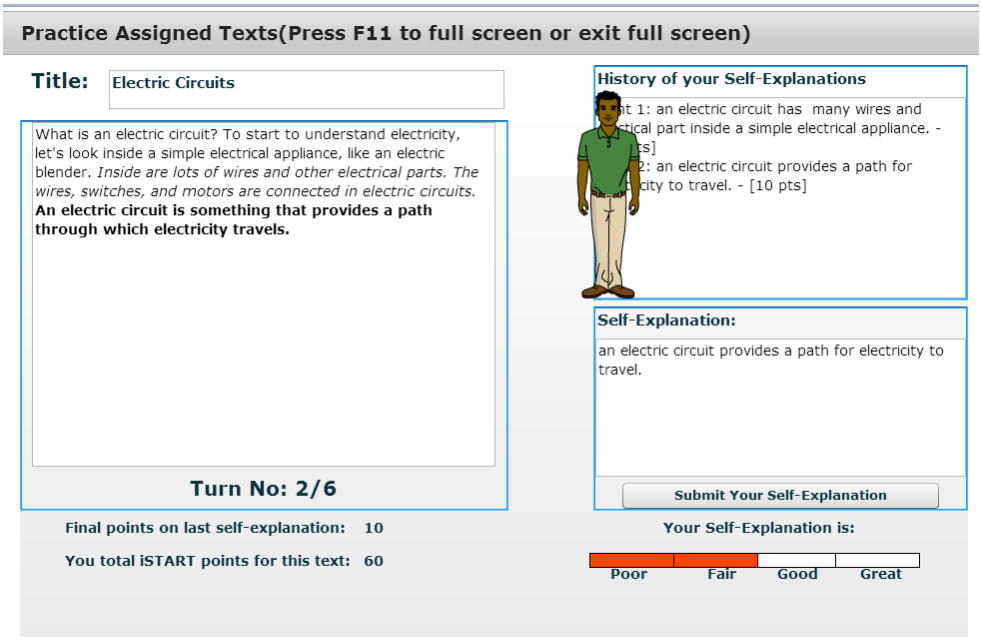


Appendix I. Figure 8. Demonstration Module Snapshot 1

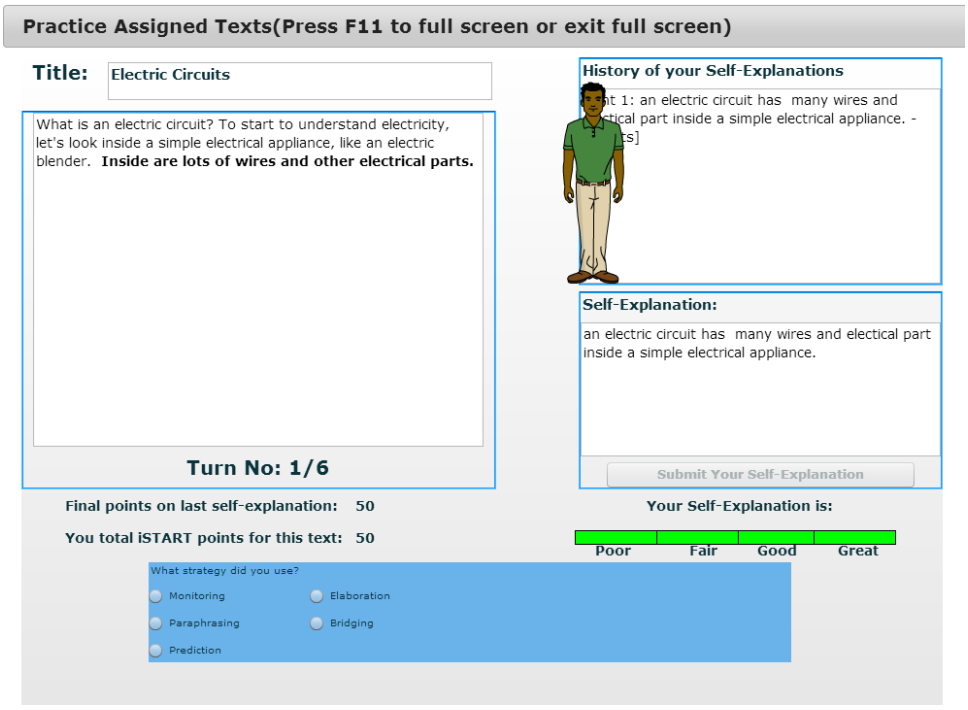
Appendix I. Figure 9. Demonstration Module Snapshot 2



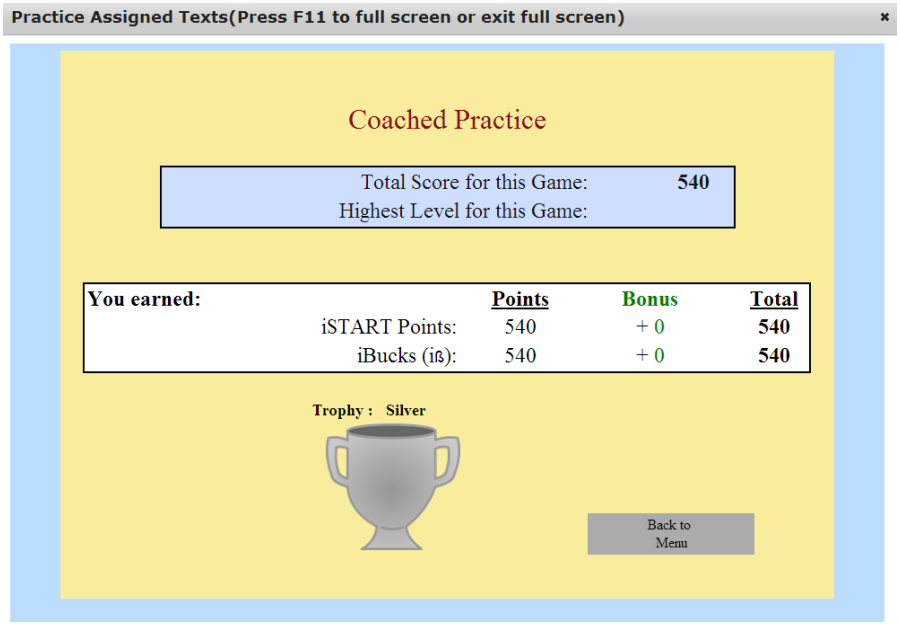
Appendix I. Figure 10. Demonstration Module Snapshot 3



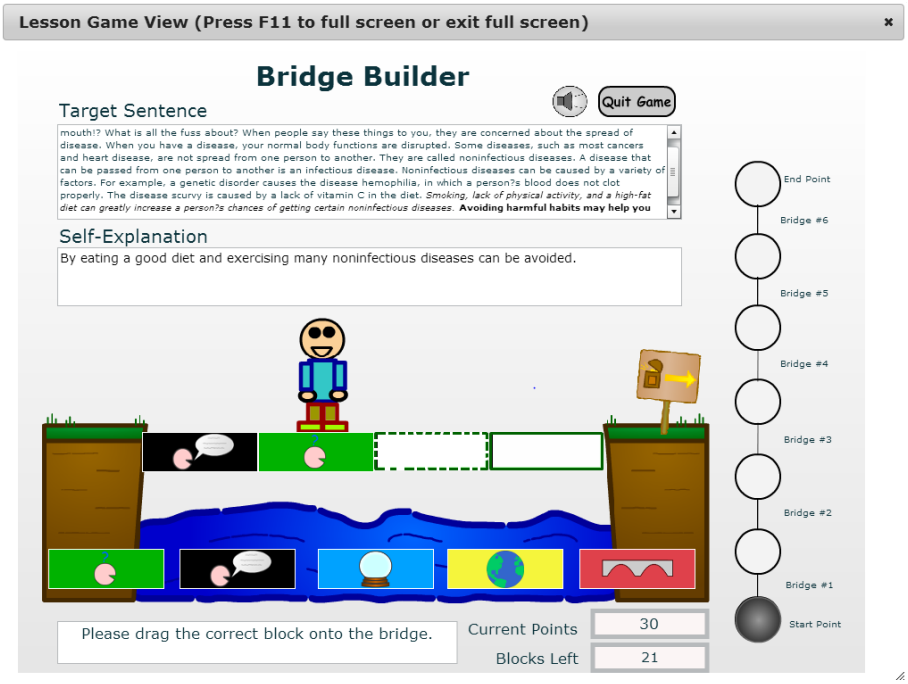
Appendix I. Figure 11. Coached Practice Module Snapshot 1



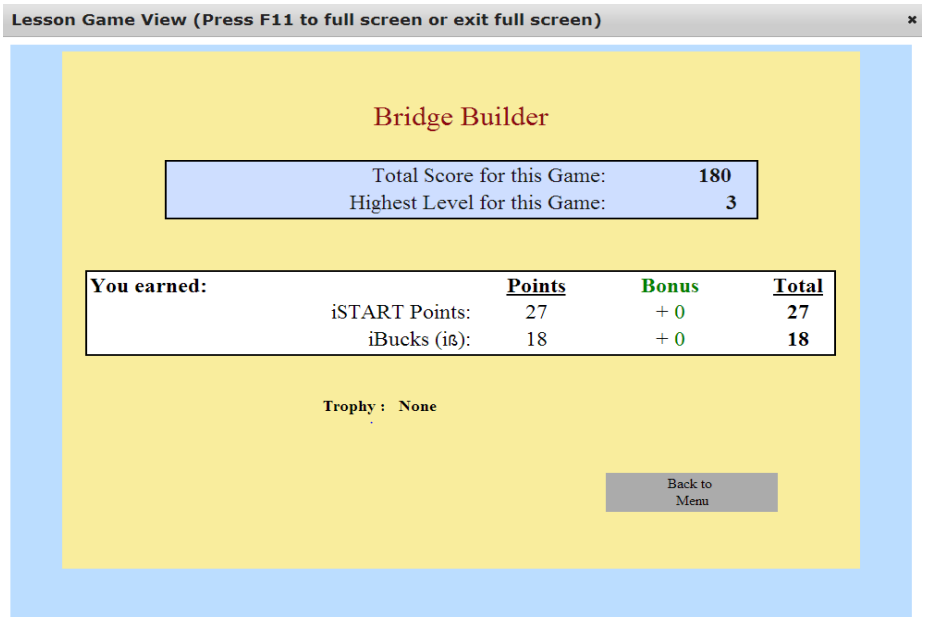
Appendix I. Figure 12. Coached Practice Module Snapshot 2



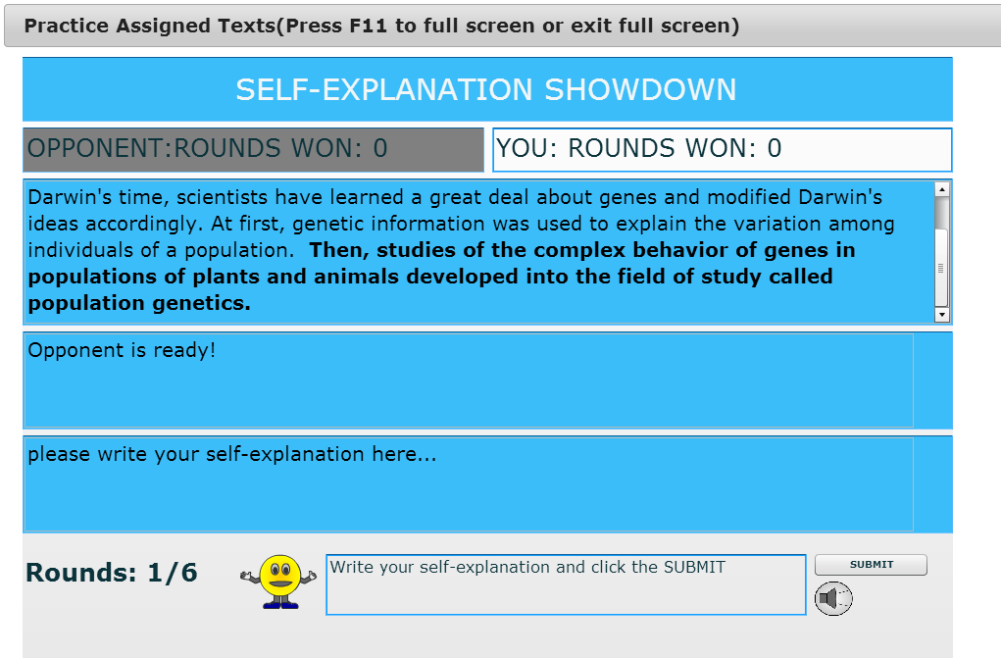
Appendix I. Figure 13. Coached Practice Module Snapshot 3



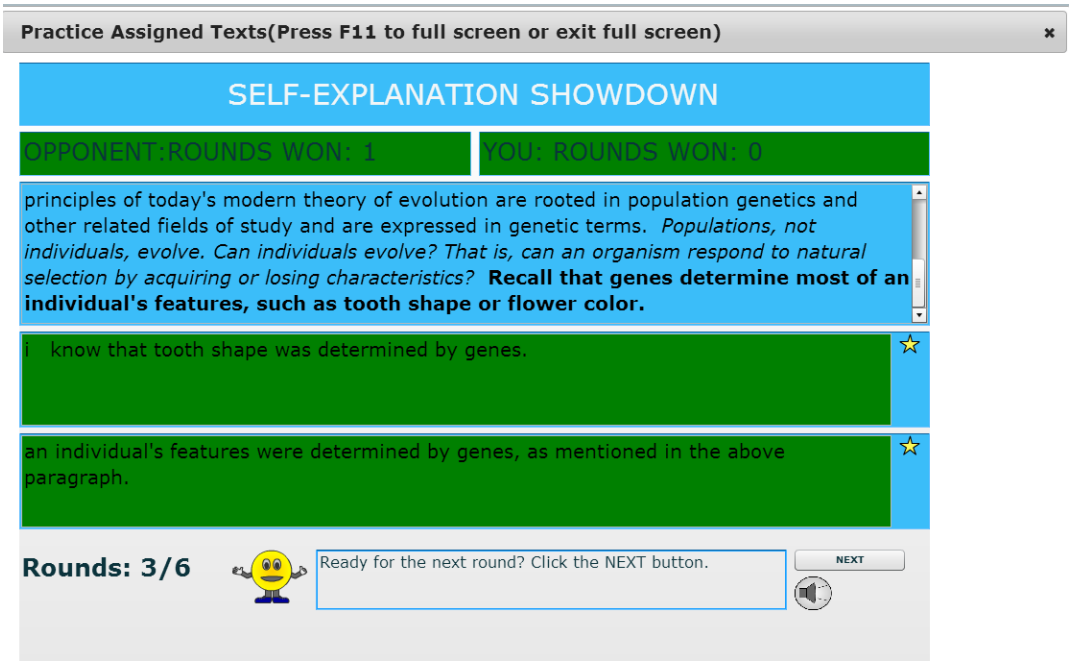
Appendix I. Figure 14. Bridge Builder Game Snapshot 1



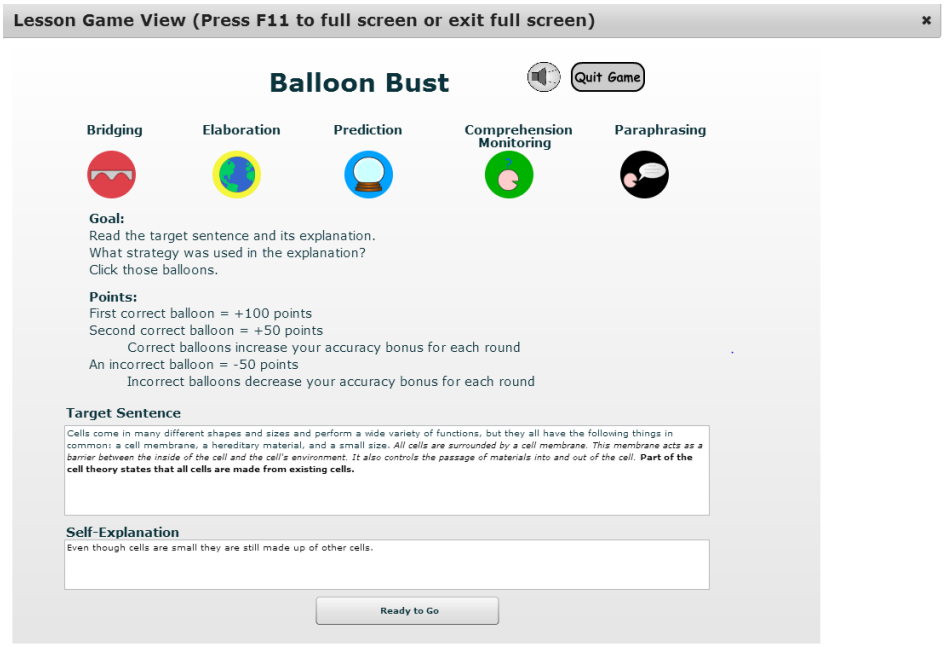
Appendix I. Figure 15. Bridge Builder Game Snapshot 2



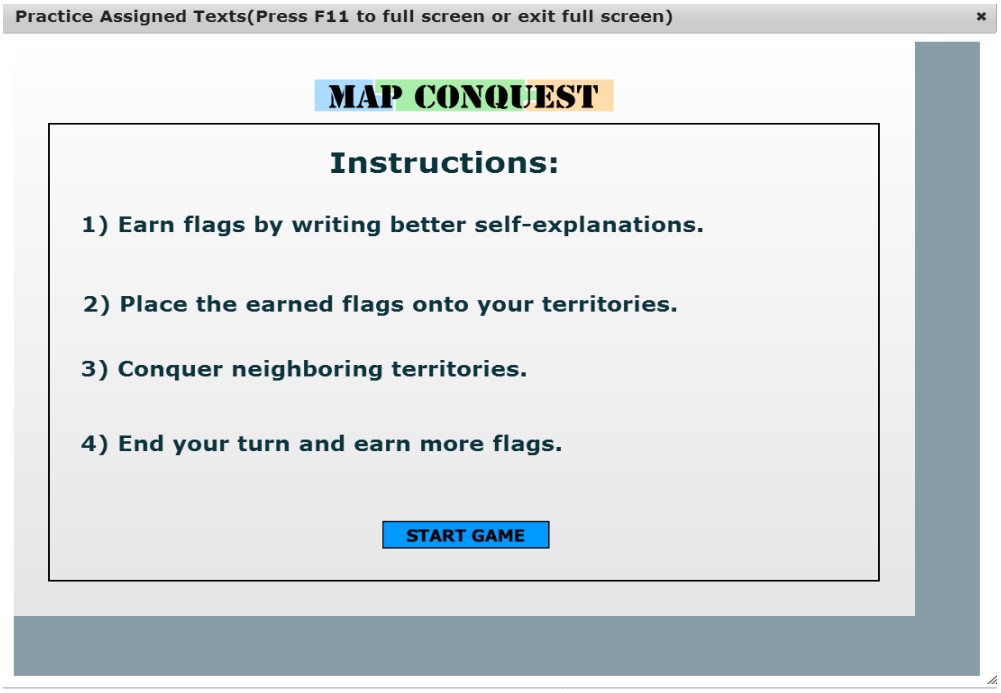
Appendix I. Figure 16. Showdown Game Snapshot 1



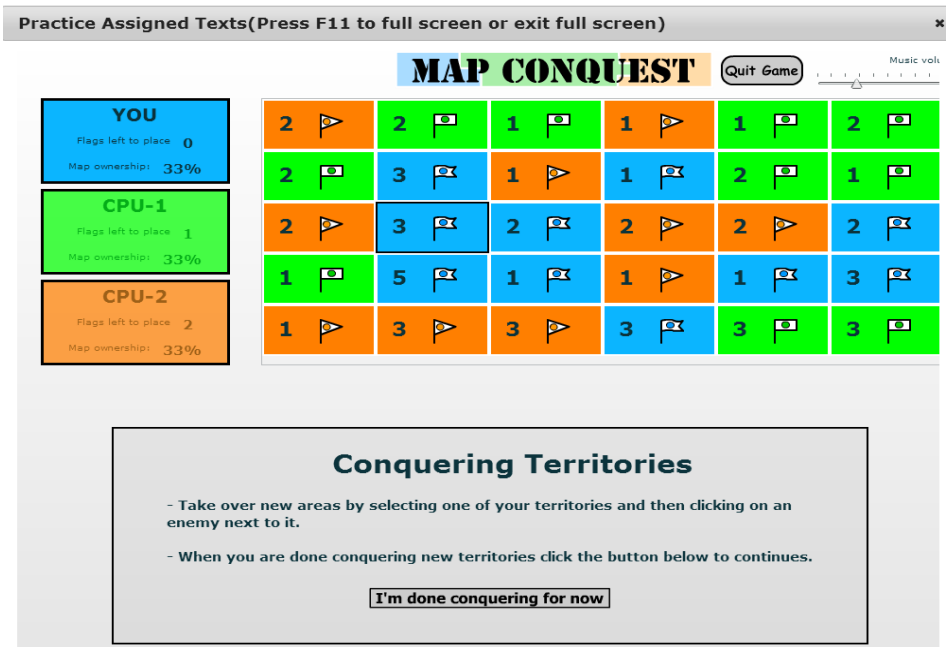
Appendix I. Figure 17. Showdown Game Snapshot



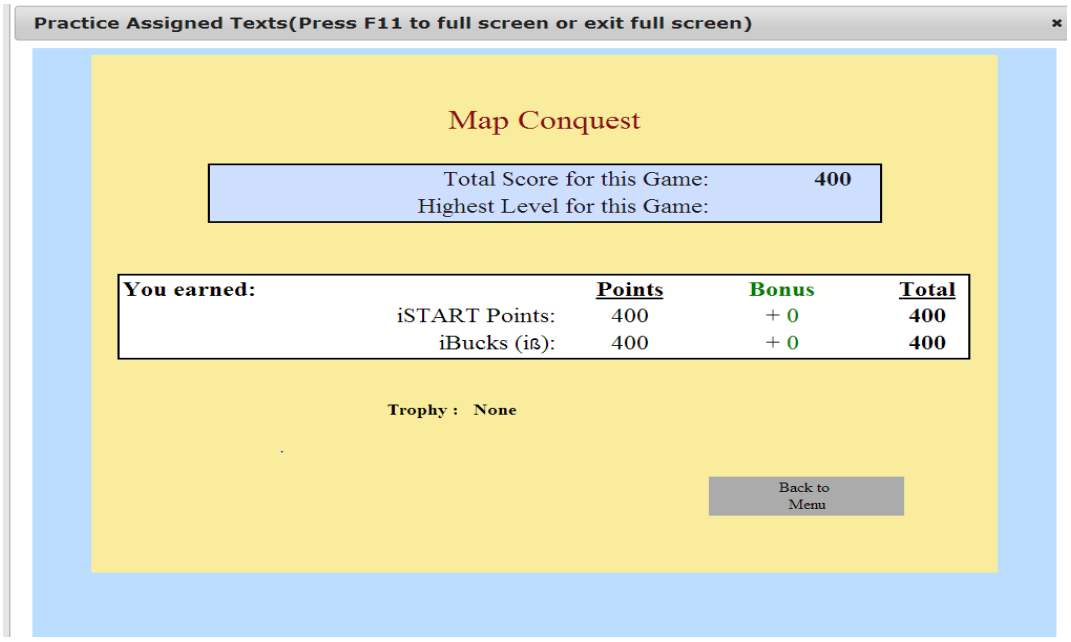
Appendix I. Figure 18. Balloon Bust Game Snapshot 1



Appendix I. Figure 19. Map Conquest Game Snapshot 1



Appendix I. Figure 20. Map Conquest Game Snapshot 2



Appendix I. Figure 21. Map Conquest Game Snapshot 3

REFERENCES

Alderson, J. C. (1984). Reading in a foreign language: A reading or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–24). London: Longman.

Alvermann, D. E., & Nealy, A. (2004). Professional development content for reading educators at the middle and high school levels. In D. Strickland & M. Kamil (Eds.), improving reading achievement through professional development (pp. 85-93), Norwood, MA: Christopher-Gordon.

Anderson, R.C. (1984). Role of the reader's schema in comprehension, learning, and memoir. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 469-482). Hillsdale, NJ: Erlbaum.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic process in reading comprehension. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), The handbook of reading research (Vol. 1, pp. 255-291). New York: Longman.

Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. *Teaching & Teacher Education*, *8*, 391-403.

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International Education*, *5*(2), 131–154. doi:10.1177/1475240906065589

Answer Explanations. (2014). Answer Explanations for: PSAT Saturday 2011 (Form S) – See more at: http://www.answerexplanations.com/psat/answer-explanations/psat-saturday-2011/#sthash.cKYfFuxx.dpufRetrieved from http://www.answerexplanations.com/psat/answer-explanations/psat-saturday-2011/

Baker, R. S. J. d., 'D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence and persistence of affect during interactions with three different computer-based learning environments. International Journal of Human–Computer Studies, 68, 223–241

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Bell, C., & McNamara, D. S. (2007). Integrating iSTART into a high school curriculum. In D. S. McNamara & G. Trafton (Eds.), Proceedings of the 29th Annual Meeting of the Cognitive Science Society, Nashville, Tennessee, August 1–4, 2007 (pp. 809–814). Austin, TX: Cognitive Science Society.

Best, R.M., Rowe, M., Ozuru, Y., & McNamara, D.S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, *25*, 65- 83.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in Self-Explanation and Self- Regulation Strategies : Investigating the Effects of Knowledge Acquisition Activities on Problem Solving. *Cognition and Instruction*, *13*(2), 221–252. doi:10.1207/s1532690xci1302

Blackmore, M. A., Pratt, C., & Dewsbury, A. (1995). The use of props in a syntactic awareness task. *Journal of Child Language*, *22*, 405–421.

Bloomfield, Leonard (1933). Language. New York: Holt Rinehart Winston

Bossers, B. (1992).*Reading in two languages: A study of*272 *The Modern Language Journal 87 (2003) reading comprehension in Dutch as a second language and in Turkish as a first language*. Rotterdam, the Netherlands: Drukkerij Van Driel.

Brantmeier, C. (2002). Second Language Reading Strategy Research at the Secondary and University Levels: Variations, Disparities, and Generalizability. The Reading Matrix, 2(3), 1–14.

Brown, A. (1982). Learning how to learn from reading. In J. A. Langer & M. T. Smith-Burke (Eds.), Reader meets author: Bridging the gap (pp.26 –54).Newark, DE: International Reading Association.

Brown, J., Goodman, K. S., &Marek, A. M. (1996). *Studies in miscue analysis: Annotate bibliography*. Newark, DE: IRA.

Carrell, P. L., & Eisterhold, J. C. (1983). Schema theory and ESL reading. *TESOL Quarterly*, *17*, 553–573.

Carlson , S. (1985) . The ethical appropriateness of subject-matter tutoring for learning disabled adolescents. *Learning Disability Quarterly, 8* , 310 – 314 doi:10.2307/1510594

Chamot, A.U., & O'Malley, J. M. (1994). Instructional approaches and teaching procedures. In K. Spangenberg-Urbschat & R. Pritchard (Eds.), Kids come in all languages: Reading instruction for ESL students (pp. 82-107). Newark, DE: International Reading Association.

Chi, M.T. H., Bassok, M., Lewis, M.W., Reimann, P.,&Glaser,R. (1989). Self-explanations:Howstu- dents study and use examples in learning to solve problems. Cognitive Science, 13, 145–182.

Chi, M. T. H., Leeuw, N. D. E., Chiu, M., & Lavancher, C. (1994). Eliciting Self-Explanations Improves Understanding, *477*, 439–477.

Clarke, M. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning, 29*, 121–150.

Clarke, M. A. (1980). The short-circuit hypothesis of ESL reading—or when language competence interferes with reading performance. *Modern Language Journal, 64*, 203–209.

Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, R.R. Jordan (Eds.), *Reading in a second language* (pp 5-12). Rowley, Mass.: Newbury House.

Corbett , A. T. (2001) . Cognitive computer tutors: Solving the two-sigma problem . In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva (Eds.) , User modeling 2001: 8th International Conference (pp. 137 – 147). Heidelberg , Germany : Springer .

Cohen , P. A. , Kulik , J. A. , & Kulik , C. C. (1982) .Educational outcomes of tutoring: A meta-analysis of fi ndings . American Educational Research Journal , 19 ,237 – 248 . CollegeBoard (2014), Critical reading, Retrieved from http://professionals.collegeboard.com/testing/psat/prep/section/critical-reading

College Board. (2006) College Board Standards for College success. Retrieved from http://www.collegeboard.com/prod_downloads/about/association/academic/english-language-arts_cbscs.pdf

Collins, C. (1991). Reading instruction that increases thinking abilities. *Journal of Reading, 34* (7), 510-516.

Cross, D. R., and Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology*, 80, 131–142.

Crede, M. (2011). A meta-analytic review of the Motivated Strategies for Learning Questionnaire. Learning and Individual Differences, 21(4), 337-346.

Cummins, J. (1979 a). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19,* 197–205.

Cummins, J. (1979 b). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*, 222-251.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly,14,* 175–187.

Cummins, J. (1981), Immigrant second language learning. Applied Linguistics, 11, 132-149. Cummins, J. (1991). Conversational and academic language proficiency in bilingual contexts. In J. H. Hulstijn and J.F. Matter (Eds.), *Reading in two languages: AILA review*. Vol. 8, 75-89.

Cummins, J. (2000). Language, power and pedagogy. Clevedon, England: Multilingual Matters.

Cziko, G. A. (1980). Language competence and reading strategies: A comparison of first and second language oral reading errors. *Language Learning, 30,*101–114.

D'Mello, S. K., & Graesser, A. C. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems, 2*(4), 23:1-38.

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist, 40,* 117-128.

Dunlosky, J., Rawson, K. a., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. Psychological Science in the Public Interest, 14(1), 4–58. Eccles, J. (1983). Expectancies, values and academic behaviors. In J. T. Spence, Achievement and achievement motives (pp. 75 - 146). San Francisco: Freeman.

Education Week, 2011. English-Language Learners. Retrieved from http://www.edweek.org/ew/issues/english-language-learners/ .

Farrar, M. T. (1986). Four conceptions of literacy. Reading Psychology, 7, 43-55.

Federal Bureau of Investigation, Language Proficiency Self-Assessment, retrieved from https://www.fbijobs.gov/1243.asp

Freedman, R. (2000). What is an intelligent tutoring system? Intelligence, 11(3), 15–16.

Fung, I., Wilkson, I., & Moore, D. (2003). L1-Assisted reciprocal teaching to improve ESL students' comprehension of English expository text. *Learning and Instruction, 13*, 1-31.

Gelderen, A., Schoonen, R. Stoel, R. Glopper, &K. Hulstijn. J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology,* 99(3), 477-491.

Gelderen, A. et al. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-Language reading comprehension: A componential analysis. *Journal of Educational Psychology,96*(1), 19-30.

Gelderen, A. V., Schoonen R., Glooper, K. D., Hulstijin, J., …, & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *The International Journal of Bilingualism*, *1*, 7–25.

Goodman, K. S. (1971). Psycholinguistic universals in the reading process. In P. Pimsleur & T. Quinn (Eds.), *The psychology of second language learning* (pp. 135–142). Cambridge, England: Cambridge University Press.

Goodman, K. S. (1973). Psycholinguistic universals of the reading process. In F. Smith (Ed.), *Psycholinguistics and reading* (pp. 21–29). New York: Holt, Rinehart and Winston.

Grabe, W. (2009). Teaching and testing. In L. Michael & C. Doughty (Eds.), *The handbook of language teaching* (pp. 441-462). West Sussex: Blackwell Publishing.

Graesser, A. C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), Reading comprehension strategies: Theories, interventions, and technologies (pp. 3-26). Mahwah, NJ: Lawrence Erlbaum.

Graesser, A. C., VanLehn, K., Rose, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. AI Magazine, 22(4), 39-51.

Hakuta, K. Educating language minority students and affirming their equal rights: Research and practical perspectives. *Educational Researcher, 40* (4), 2011.

Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. Journal of Educational Psychology, 75, 821–829

Hellekjær, G. O. (2009). Academic English reading proficiency at the university level: A Norwegian case study. *Reading in a Foreign Language, 21* (2), 198-222.

Holmes, P. (2004). Negotiating differences in learning and intercultural communication. *Business Communication Quarterly, 67*(3), 294–307

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, Textual coherence, and inferences. *Studies in Second Language Acquisition, 18,* 403–432.

Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes, 29,* 223–267.

Institute of education sciences, the condition of education, 2012.English Language Learners in Public Schools. Retrieved from http://nces.ed.gov/programs/coe/indicator_ell.asp

Institution of International Education. (2012). Open doors 2012: international student enrollment increased by 6 percent. Retrieved from http://www.iie.org/Who-We-Are/News-and-Events/Press-Center/Press-Releases/2012/11-13-2012-Open-Doors-International-Students

Institute of International Education. (2014). 2013 Opendoors 20 | 13 "Fast Facts". Retrieved from file:///C:/Users/hongxiaf/Downloads/Fast-Facts-2013.pdf

Irvin, J., Buehl, D. R., & Klemp, R. M. (2007). Reading and the high school student: Strategies to enhance literacy. Boston, MA: Pearson Education.

Jackson, G.T., Guess, R.H., & McNamara, D.S. (2010). Assessing cognitively complex strategy use in an untrained domain. Topics in Cognitive Science, 2, 127-137.

Jackson, G.T., & McNamara, D.S. (2011). Motivational impacts of a game-based intelligent tutoring system. In R. C. Murray & P. M. McCarthy (Eds.), Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference (pp. 519-524). Menlo Park, CA: AAAI Press.

Jackson, G. T., Dempsey, K. B., & McNamara, D. S. (2012). Game-based practice in a reading strategy tutoring system: Showdown in iSTART-ME. In H. Reinders (Ed.), Computer Games (pp. 115-138). Bristol, UK: Multilingual Matters.

Jackson, G. T., Varner, L. K., Boonthum-Denecke, C., & Mcnamara, D. S. (2013a). The impact of individual differences on learning with an educational game and a traditional ITS Chutima Boonthum-Denecke, 8(4), 315–336.

Kirajima, R. (1997). Referential strategy training for second language reading comprehension of Japanese texts. *Foreign Language Annuals, 30,* 84–97.

Koskinen, P. S., Gambrell, L. B., Kapinus, B. A, & Heathington, B. S. (1988). Retelling: A strategy for enhancing students' reading comprehension". *The Reading Teachers, 41*(9), 892-896.

Kuo, B. C. H., & Roysircar, G. (2004). Predictors of Acculturation for Chinese Adolescents in Canada: Age of Arrival, Length of Stay, Social Class, and English Reading Ability, *32*(July), 143–154.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.) (2007). Handbook of latent semantic analysis. Mahwah, NJ: Erlbaum.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, and words you can't guess. In T. H. J. Coady (Ed.), *Second language vocabulary acquisition* (pp. 20–34). Cambridge, England: Cambridge University Press.

Lee, D. S. (1997). What teachers can do to relieve problems identified by international students. *New Directions for Teaching and Learning 70*: 93–100.

Lewthwaite, M. (1996). A study of international students' perspectives on cross-cultural adaptation. *International Journal for the Advancement of Counselling 19*(2): 167–185.

Lin, J.G., & Yi, K. Y. (1997). Asian international students' adjustment: Issues and program suggestions'. College Student Journal 31(4), 473–479.

Linderholm, T., Therriault, D. J., & Kwon, H. (2014). Multiple Science Text Processing: Building Comprehension Skills for College Student Readers. Reading Psychology, 35(4), 332–356.

Maarof, N., & Yaacob, R. (2011). Meaning-making in the first and second language: reading strategies of Malaysian students. Procedia - Social and Behavioral Sciences, 12, 211–223.

Mathewson, G. C. (1994). Model of attitude influence upon reading and learning to read. In R. B. Ruddell, M.R. Ruddell, & H. Singer (Ed.), *Theoretical models and processes of reading* (4th ed.), (pp. 1131-1161). Newark, DE, US: International Reading Association.

McCardle, P., Chhabra, V., Kapinus, B. (2008). *Reading research in action: A teacher's guide for student success*. Baltimore, MD: Brookes.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes,38,* 1-30.

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

McNamara, D.S. (2006). Bringing cognitive science into education and back again: The value of interdisciplinary research. Invited editorial in *Cognitive Science, 30,* 1-4.

McNamara, D.S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34*, 147-171.

McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theory, interventions,and technologies*. Mahwah, NJ: Erlbaum.

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of latent semantic analysis (pp. 227–241). Mahwah, NJ: Erlbaum.

McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). The 4-Pronged Comprehension Strategy Framework. In D. S. McNamara (Ed.), Reading comprehension strategies: Theories, interventions, and technologies (pp. 465–496). Mahwah, NJ: Erlbaum.

Mcnamara, D. S. (2009). The Importance of Teaching Reading Strategies. Perspectives on Language and Literacy, 35, 34–40.

McNamara, D.S., & Magliano, J.P. (2009). Self-explanation and metacognition: The dynamics of reading. In J.D. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 60-81). Mahwah, NJ: Erlbaum.

McNamara, D.S., Jackson, G.T., & Graesser, A.C. (2009). Intelligent tutoring and games (iTaG). In H.C. Lane, A. Ogan, & V. Shute (Eds.), Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education*(pp. 1-10). Brighton, UK: AIED.*

McNamara, D.S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science*, *2*, 1-15.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for Social Agency in Computer-Based Teaching : Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents? *Cognition and Instruction*, *19*(2), 177–213.

Nassaji, H. (2003). Higher-Level and lower-Level test processing skills in advanced ESL reading comprehension. *The Modern Language Journal, 87* (2), 261-276.

National Accessible Reading Assessment Projects. (2006). Defining reading proficiency for accessible large scale assessments: Some guiding principles and issues. Minneapolis, MN: Author.

National Accessible Reading Assessment Projects. (2006). Defining reading proficiency for accessible large scale assessments: Some guiding principles and issues. Minneapolis, MN: Author.

National Center for Education Statistics. The condition of education. Retrieved from http://nces.ed.gov/programs/coe/indicator_cgf.asp

National Endowment for the Arts. (2007). To read or not to read: a question of national consequence. Retrieved from http://arts.gov/sites/default/files/ToRead.pdf

Nkambou, R., Mizoguchi, R., & Bourdeau, J. (2010). Advances in intelligent tutoring systems. Heidelberg: Springer

No Child Left Behind (2001). Act of 2001, 20 U.S.C. 6301 e seq (2001) PL 107–110.

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. Artificial Intelligence Review, 4, 251-277.

Nye, B.D., Graesser, A.C., & Hu, X. (in press). AutoTutor and family: A review of 17 years of natural language tutoring. International Journal of Artificial Intelligence in Education.

O'Malley. J., & Chamot, A. (1990). Leaning Strategies in Second Language Acquisition. Cambridge, UK: Cambridge University Press.

Oxford, R. L. (1990). Language Learning Strategies: What Every Teacher Should Know. New York: Newburry House/Harper& Row.

Ozuru, Y., Briner, S., Best, R., & McNamara, D. S. (2010). Contributions of Self-Explanation to Comprehension of High- and Low-Cohesion Texts. Discourse Processes, 47(8), 641–667.

Paas, F., Renkl, A., & Sweller, J. (2003) Cognitive load theory and instructional design: recent developments, *Educational Psychologist*, *38*(1), 1-4.

Pearson, P.D., and Tierney, R.J.. (1984). On becoming a thoughtful reader: Learning to read like a writer. In A.C. Purves and O. Niles (Eds.), Becoming readers in a complex society [Eighty-third Yearbook of the National Society of the Study of Education] (pp. 144-173). Chicago: University of Chicago Press.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Pintrich, P. R., & Schunk, D. H. (1996). Motivation in education: Theory, research, and applications. Englewood Cliffs, NJ: Prentice Hall Merrill.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. International Journal of Educational Research, 31, 4597470.

Pleasant Valley Community School District. (2014). Reading Strategies. Retrieved from http://www.pleasval.k12.ia.us/studyskills/parentreading.htm

Prensky, M. (2005). Computer games and learning : Digital game-based learning. In J. Rasessens and J. Goldstein (Ed.), *Handbook of computer game studies* (pp. 97-124). Cambridge (MA): MIT Press

Pressley, M., & Afflerbach, P. (1995). Verbal protocols of reading: The nature of constructively responsive reading. Hillsdale, NJ: Lawrence Erlbaum.

Pressley, M. (2000) What should comprehension instruction be the instruction of? In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), Handbook of reading research: Volume

III (pp. 545-561). Mahwah NJ: Erlbaum.

Pressley, M. (2002a). Comprehension strategies instruction: A turn-of-the-century status report. In C.C. Block & M. Pressley (Eds.), Comprehension instruction: Research-based best practices (pp. 11–27). New York: Guilford.

Pressley, M. (2002b). Metacognition and self-regulated comprehension. In A.E. Farstrup & S.J. Samuels (Eds.), What research has to say about reading instruction (3rd ed., pp. 291–309). Newark, DE: International Reading Association.

Pressley, M. (2006). Reading instruction that works: The case for balanced teaching (3rd ed.). New York: Guilford.

Ritter , S. , Anderson , J. R. , Koedinger , K. R. , & Corbett , A.(2007) . Cognitive Tutor: Applied research in mathematics education . Psychonomic Bulletin & Review , 14 , 249 – 255

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. The American Psychologist, 55(1), 68–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11392867

Saville-Troike, M. (1984), What really matters in second language learning for academic achievement?. TESOL Quarterly, 18: 199–219. doi: 10.2307/3586690

Schraw, G., & Moshman, D. (1995). Metacognitive Theories. Educational Psychology Review, 7(4), 351-371.

Schunk, D. (1985). Self-efficacy and school learning. Psychology in the Schools, 22,

Senyshyn, R. M., Warford, M. K., & Zhan, J. (2000). Issues of adjustment to higher education: International students' perspectives. *International Education 30*(1), 208-223.

Shneyderman, A., & Froman, T. (2012). Time to English reading proficiency. Research Brief. Retrieved from http://drs.dadeschools.net/ResearchBriefs/RB1201.pdf

Smagorinsky, P. (2009). The Cultural Practice of Reading and the Standardized Assessment of Reading Instruction: When Incommensurate Worlds Collide. Educational Researcher, 38(7), 522–527. doi:10.3102/0013189X09347583

Snow, C., & RAND (2002). Reading for understanding: Toward an R&D program in reading comprehension. Santa Monica, CA: RAND.

Strong American Schools (2008). Diploma to nowhere: report pegs college remediation cost in excess of $2 billion annually. Retrieved from http://www.edin08.com/diplomatonowhere.aspx

Summary, S. D. (2012). 2012 Test and Score Data Summary for TOEFL iBT Tests and TOEFL PBT Tests.

Taillefer, G. (1996). L2 reading ability: Further insights into the short-circuit hypothesis. *Modern LanguageJournal, 80,* 461–477.

Taylor, S. (1999). Better learning through better thinking: Developing students' metacognitive abilities. Journal of College Reading and Learning, 30(1), 34. Retrieved from Expanded Academic Index ASAP.

Taylor, A., Stevens, J., & Asher, J. W. (2006). The effects of explicit reading strategy training on L2 reading comprehension. In J. Norris & L. Ortega (Eds.), *Synthesizing research on Language Learning and Teaching* (pp. 213-244). Philadelphia: John Benjamins.

Taylor, R. T. (2012). Review of the Motivated Strategies for Learning Questionnaire (MSLQ) Using Reliability Generalization Techniques to Assess Scale Reliability. Retrieved from http://etd.auburn.edu/bitstream/handle/10415/3114/Dissertation%20-%20TaylorRobin%202012.pdf?sequence=2

Terry, B. D. (2007) The cost of remedial education. Retrieved from http://www.texaspolicy.com/sites/default/files/documents/2007-09-PP25-remediation-bt.pdf

TESOL. (2008). Position Statement on English as a Global Language. Retrieved from http://www.tesol.org/docs/pdf/10884.pdf.

Tierney, R. J. (1992). Ongoing research and new directions. Reading/writing connections: Learning from research, 246-259.

Trabasso, T., & Bouchard, E. (2002). In C. Block& M. Pressley (Eds.), *Comprehension instruction: Research based best practice* (pp.28-41).New York: The Guilford Press.

Trabasso, T., & Magliano, J. P. (1996). How do children understand what they read and what can we do to help them? In M. Grades, P. van den Broek, & B. Taylor (Eds.), The first R: A right of all children, (pp 160-188). NY: Columbia University Press.

Tunmer, W. E., & Hoover, W. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough., L. C. Ehri & R. Treiman (Eds), *Reading acquisition* (pp. 175–214). Hillsdale, NJ: Erlbaum.

Van Eck, R. (2006). Digital game-based learning. *Educause Review , 41* (20), 16-30.

Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, …, Snellings, P., et al. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: a componential analysis. *Journal of Educational Psychology*, *96*(1), 19–30.

van Gelderen, A., Schoonen, R; Stoel, R. D.; & de Glopper, K. (2007) . Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, *99*(3), 477-491.

Vanvoorhis, C. R. W., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50.

Verhoeven, L. (1990). Acquisition of reading in Dutch as a second language. *Reading Research Quarterly*, *25*, 90–114.

Vygotsky, L. (2012). *Thought and Language*. Edited by A, Kozulin. Cambridge, MA: MIT Presses.

Walker, J. (2009). World's English mania. Retrieved from http://www.ted.com/talks/jay_walker_on_the_world_s_english_mania.html

Walter, C. (2004) Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics* 25.3: 315–339.

Walter. C. (2007). First- to second-Language reading comprehension: not transfer, but access. *International Journal of Applied Linguistics, 17*(1), 14-35.

Wilkins, C., Hartman, J., Howland, N., and Sharma, N. (2010). How prepared are students for college-level reading? Applying a Lexile®-based approach. (Issues & Answers Report, REL 2010–No. 094). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Wilkinson, I., & HyeSon, E. (2011). A dialogic turn in research on learning and teaching to comprehend. In M. Kamil, P. Pearson, E. Moje, & P. Afflerback (Eds.), *Handbook of reading research* (pp. 359-387), 4th vol. New York: Routledge.

Willingham, D. T. (2006). The usefulness of brief instruction in reading comprehension strategies. *American Educator, 30* (4). 39-50.

Yamashita, J. (2002). Mutual compensation between L1 reading ability and l2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, *25* (1). 81-95. Schoenfeld, D. (2014). Statistical considerations for clinical trials and scientific experiments. http://hedwig.mgh.harvard.edu/sample_size/js/js_crossover_quant.html

Ying, Y. (2001). Migration and cultural orientation: An empirical test of the psychoanalytical theory in Chinese Americans. Journal of Applied Psychoanalytical Studies, 3,409-430.

Yuan, W. (2011). Academic and cultural experiences of Chinese students at an American university: A qualitative study. Inercultural Communication Studies, XX:1, 141-157.

Yuko, I. (2008). The perceptions of Japanese students toward academic English reading: implications for effective ESL reading strategies. Retrieved from http://www.freepatentsonline.com/article/Multicultural-Education/184800662.html

Yeh, C. J., & Inose, M. (2003). International students' reported English fluency, social support satisfaction, and social connectedness as predictors of acculturative stress. *Counselling Psychology Quarterly*, *16*(1), 15–28.

Zwan. R. F., Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation-model construction