Tampere University of Applied Sciences

Meta

Exploring the potential of a minimally adaptive intelligent tutoring system to support comprehension of grade appropriate texts

Laura Killips

MASTER'S THESIS
December 2019

Master of Business Administration
Educational Leadership

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Master of Business Administration
Educational Leadership

Laura Killips:
Meta: Exploring the potential of a minimally adaptive intelligent tutoring system to support comprehension of grade appropriate texts

Master's thesis 66 pages, appendix 2 pages
December 2019

Background: In spite of evidence that demonstrates the impact explicit strategy instruction can have on students' ability to comprehend text, few educators include these lessons in their classes. As a result, students graduate without the appropriate literacy foundation. Given the time commitment necessary to develop teacher practice and the urgency of this issue, it is useful to consider alternative methods of support.

Purpose: The study's objective was to explore the potential of a minimally adaptive intelligent tutoring system that provides explicit strategy instruction for comprehension monitoring. The question was whether or not such a system could raise student awareness of their understanding as they read, which would give evidence as to the system's potential to support comprehension.

Methodology: The research was designed as a proof-of-concept, randomized control study. Students in both the control group and the intervention group were given a pre-test that assessed metacognition and their ability to paraphrase text. Half of those students then participated in the intervention where they received explicit instruction in comprehension monitoring through an intelligent tutoring system prototype. Then students in both groups completed a post-test to measure growth.

Results: Results were mixed. A paired-sample t-test was used to determine that there was not a statistically significant difference in the change from pre-test to post test in students' ability to note a planted error; however, there was a statistically significant difference in the students' ability to notice vocabulary breakdowns. Additionally, there was evidence of accelerated growth on several measures for the intervention students as compared to the control students.

Conclusions: There is evidence to suggest that a minimally adaptive intelligent tutoring system has the potential to support comprehension of grade level text. In order to further explore the extent to which such a system can provide support, the student model should be built out further in order to capture all parts of the metacognitive process.

Key words: comprehension monitoring, explicit instruction, intelligent tutoring system, metacognition, strategy instruction

CONTENTS

GLOSSARY

Intelligent Tutoring System (ITS) — A computer system that adapts to student input in order to provide targeted instruction, practice, and feedback.

Metacognition — Often referred to as “thinking about thinking,” the term metacognition is generally used to describe one’s awareness and control of their thought process.

Comprehension Monitoring — Also known as metacognitive reading, comprehension monitoring refers to the active process of reading that encompasses having an awareness of one’s understanding and the strategic use of strategies to address comprehension breakdowns.

1 INTRODUCTION

A literate society is a healthy society. The human, economic, political, cultural, and social benefits of literacy have been well documented. As such, literacy is considered a fundamental human right that is protected by many international treaties (The United Nations, Educational, Scientific, and Cultural Organization [UNESCO], 2005). And yet, the World Literacy Foundation (2018) estimates that three billion people across the world struggle with even basic literacy skills (World Literacy Foundation, 2018).

Literacy encompasses a number of cognitive functions, which include both reading and writing. This current study is concerned primarily with those functions involved in reading comprehension. These functions can be taught through explicit strategy instruction (Fielding and Peierson, 1994; Grossman, Loeb, Cohen, & Wyckoff, 2013). However, a study recently published in Reading Research Quarterly revealed that teachers spend less than half their instructional time on explicit strategy instruction for comprehension (Magnusson, Roe, & Blikstad-Balas, 2019). Furthermore, very little of that instructional time would qualify as high-quality instruction (Magnusson et al., 2019; Grossman et al., 2013). What is more is that of the comprehension strategies taught, there were only two instances where a teacher explicitly instructed students in how to monitor their comprehension, a skill that is critical to understanding text (Magnusson et. all, 2019).

Improved professional development for teachers in this area is an important step, especially when one considers the significant potential expert teachers have to improve student outcomes (Hattie, 2012; RAND Corporation, 2012). However, The Education Commission of the States (2005) concludes that teacher change only comes after signifiacnt time investment—up to 50 hours of instruction, practice and coaching of the new strategies (p. 19). Other research breaks this down to a goal of 3-4 hours per week and 10 days per year spent developing one's professional practice (Killion, 2013, p. 6). In places like the United States, where 79% of a teacher's work week is spent in front of students, finding time to devote to continued professional development is a challenge (Organization for Economic Co-operation and Development [OECD], 2018). Given this reality, it could take years before educators develop the expertise necessary to better support comprehension instruction. Thus, improved professional development cannot be the only solution.

It is possible that technology can be leveraged to address the issue. And indeed, many intelligent tutoring systems (ITSs) for literacy skills already exist. The challenge to these systems lies, at least part, in the fact that there are numerous ways to solve any given breakdown in comprehension (Jacovina & Mcnamara, 2017). And in a system that aims to improve students' comprehension, the problem itself changes depending on the specific knowledge and skills a student brings to the text. Thus, creating a highly adaptive system to support comprehension through improved monitoring presents significant hurdles. Still, the evidence that even minimal levels of adaptivity can lead to improved outcomes in systems that support general comprehension is promising (Jacovina & Mcnamara, 2017). It is therefore worth exploring the potential of such a system to support comprehension monitoring specifically.

2 Theoretical Framework

ITS design generally includes four components: the knowledge base, the student model, the pedagogical model, and the user interface, (Sani & Aris, 2014; Sedlmeier, 2001; US Army Research Laboratory, 2015). The knowledge base refers to the specific learning goals as they relate to content and skill and the pedagogical model refers to the instructional practices that are employed to achieve these goals. The student model refers to the adaptability of the system, or how it adjusts based on student inputs. And user interface refers to the specific technology and formatting of the system itself.

To create their system, developers for the Intelligent Tutoring System of the Structure Strategy (ITSS) followed a process laid out by researchers for the United States Army Research Institute for the Behavioral and Social Sciences. Their first step was to build the knowledge base by identifying the appropriate learning tasks. They then determined the ideal learning approach for their pedagogical model. This was followed by developing the student model where they created possible tutor-student interactions. Finally, they created and programmed the user interface. (Anderson, Corbett, Koedinger, Pelletier, 1996; Wijekumar & Meyer, 2006).

The theoretical framework for this study is informed by this process. In order to examine the potential a minimally adaptive ITS has to support student comprehension through improved monitoring, a prototype must first be built. To do that, the design process used by the ITSS and US Army Research developers is followed. The following section is organized by this process. For each step, there is first a review of the relevant literature that relates to the component addressed (e.g. the knowledge base, the pedagogical model…). After the review, findings are synthesized in order to inform the design for each component of an ITS that could support comprehension monitoring.

2.1 Building the Knowledge Base

The goal for this ITS is to support comprehension on grade appropriate texts through explicit strategy instruction for comprehension monitoring, which is a metacognitive process (Jacobs & Paris, 1987). To build a strong knowledge base, it is therefore important to first understand the relevant research in metacognition.

Educational psychologist, John Flavell, originally coined the term metacognition in 1979 to mean "cognition about cognitive phenomenena" (Flavell, 1979, p. 906.) More simply, metacognition is "thinking about thinking" (Jacobs and Paris, 1987, p. 255; Martinez, 2006, p. 696). This definition can be misleading, as metacognition also includes the active monitoring and control, or management, of one's thought (Martinez, 2006; Kuhn & Dean, 2004). Thus, discussions about metacognition generally focus on two categories of mental activities: knowledge of cognition and monitoring of cognition (Baker & Brown, 1980; Cross & Paris, 1988; Flavell, 1979).

2.1.1 Knowledge about Cognition

Flavell (1979) argued that metacognitive knowledge consists of knowledge of oneself, knowledge of the task and knowledge of the strategies. Further review of the literature surrounding metacognitive theory supports dividing metacognitive knowledge into three groups; however, it is generally broken down into: (1) Declarative knowledge, (2) Procedural knowledge and (3) Conditional knowledge (Schraw and Moshman, 1995; Jacobs & Paris, 1987).

While these approaches are related, they are not the same. Flavell (1979), for instance, emphasizes a person's understanding of his or herself as part of the metacognitive process. While these could be considered declarative statements such as *I am good at math*, or *Reading long texts is hard for me*, other researchers broaden their view of declarative knowledge to include facts we know about the world (Anderson, 1983; Jacobs & Paris, 1987; Sahdra & Thagard, 2003; Sun, Merrill, & Peterson, 2001). Some refer to declarative knowledge as *knowing that*, as in, one may know that slowing down can help them better understand text, or one may know that sentence structure can provide clues to unknown words (Jacobs & Paris, 1987; Sun, Merrill, & Peterson, 2001). The knowledge base for this prototype will take this broader view that declarative knowledge include more generalizable world knowledge.

Declarative Knowledge and Comprehension Monitoring

In order to design the knowledge base for an ITS that supports comprehension monitoring, it is important to consider the declarative, or world knowledge, that students must know in order to effectively monitor their comprehension. In a classroom context, this can be seen as the content of a lesson—the facts about reading or language that are always true or at least generalizable to a broad context. Determining the appropriate learning goals to build the knowledge base is no easy task given that reading is considered an ill-defined domain (Jacovina & McNamara, 2017). It is also challenging because declarative, or world, knowledge plays a role in the rest of the metacognitive process (Flavell, 1979; Schraw, 1998). The research that informs those learning goals is discussed in this section.

There is a common misconception, especially among struggling readers, that reading is simply saying the words on the page rather than an active process that involves thinking (Schoenbach, Greenleaf, & Murphy, 2012; Association for Supervision and Curriculm Development [ASCD], 2005). Therefore, a first step must be to ensure students know that reading is an active process and that this process begins by paying attention to what they do and do not understand. Additionally, failure to recognize a comprehension breakdown may come, at least in part, from not knowing the standard against which a reader is judging understanding (Markman, 1979). Therefore, another early lesson must clarify this standard for students.

In reading comprehension, this may look like judging whether or not one can paraphrase the text. Paraphrasing a text is considered a strong reading strategy because it helps students understand text better. It is useful because it forces the reader to put text into language that is more easily accessible and also make connections between content. (National Reading Panel, 2000). If the reader cannot put the text into their own words, it likely means there was something they did not understand (Jacobs & Paris, 1987). Thus, students need to know that they likely understand the text when they can paraphrase it and that they likely don't understand if they cannot.

In addition to knowing what comprehension is and the standard used to measure the extent of comprehension, declarative knowledge can also refer to one's awareness about the factors that affect comprehension (Cross & Paris, 1988). While these factors certainly vary from reader to reader, there are some frameworks used to determine text difficulty. For instance, The Common Core State Standards, the national learning standards used in the United States, uses a three-part model ("English Language Arts…", n.d.). This model considers the qualitative components of a text, such as levels of meaning, knowledge demands, and language conventions or clarity. It takes reader and task into consideration. This means that what the reader is asked to do with the text can determine the difficulty of the text. Perhaps the reader must simply read to identify a piece or two of explicitly stated information. This would make the difficulty easier compared to a task where the reader must interpret multiple conclusions based on inferential information. Finally, the model considers the quantitative demands of a text—those demands that can be assessed using a computer. ("English Language Arts…", n.d.). Figure 1 illustrates these three factors that affect text complexity.

Figure 1. Measuring Text Complexity: Three Factors ("English Language Arts…", n.d.)

Although the Common Core model requires human considerations about text, there is some evidence that quantitative measures are enough to accurately predict text difficulty, particularly in lower grades. Nelson, Perfetti, D. Liben and M. Liben (2102) tested several quantitative tools and concluded that these tools could objectively measure text complexity. While there is some variation in what aspect of texts the tools analyze, there is also notable overlap, as can be seen in Table 1.

TABLE 1. Measures of Text Complexity (Nelson et al., 2012)

Tool	Variables Used
• Lexile®	• Word Frequency • Sentence Length
• ATOS	• Word length • Word grade level • Sentence length (with adjustments for extreme sentence length in the ATOS for books formula) • Book length (in ATOS for books formula)
• Degrees of Reading Power®: DRP® Analyzer	• Word length • Word difficulty • Sentence Length • Within-sentence punctuation
• REAP (REAder-specific Practice) Readability Tool	• Word frequency • Word length • Sentence length • Sentence count • Parse tree of sentences and paragraphs • Frequency of node elements
• Source Rater	• Word frequency • Word length • Word meaning features (concreteness, imagability, etc.) • Word syntactic features (tense, part of speech, proper names, negations, nominalizations, etc.) • Word types (academic verbs, academic downtoners, academic word list) • Sentence length • Paragraph length • Within-sentence and between-sentence cohesion measures • Number of clauses (including type and depth) • Text genre: informational, literary, or mixed (computed automatically or manually overridden, if preferred)

• Pearson Reading Maturity Metric	• Pearson Word Maturity Metric • Word length (e.g. syllables per word) • Sentence length • Within-sentence punctuation • Within and between-sentence coherence metrics • Sentence and paragraph complexity (e.g. perplexity) • Order of information
• Coh-Metrics Text Easability Assessor	• Word frequency • Word length • Word meaning features (concreteness, imagability, number of senses, etc.) • Word syntactic features (part of speech, negations, etc.) • Sentence length • Sentence complexity • Paragraph length • Within-sentence and between-sentence cohesion measures

Each of the tests includes some measure of words used in the text. This suggests that text difficulty can be determined at least in part by the vocabulary used. This is unsurprising given that researchers have consistently conclude that improving student vocabulary is one key to improving comprehension (Beck, Perfetti, & McKeown, 1982; Biemiller & Slonim, 2001; Edmonds et al., 2009). Each test also analyzes the characteristics of the sentences used in a text: length, punctuation, and coherence with other sentences. This is because the syntactic frame (syntax) can impact comprehension (Dapretto & Bookheimer, 1999). Finally, most of the tools measure text length, either the length of individual paragraphs or the length of the text as a whole. This speaks to the issue of reader stamina. Students who lack stamina could quickly become fatigued when reading longer texts and would then struggle to actively monitor their comprehension (Hiebert, Wilson, & Trainin, 2010).

Thus, while reading can be a highly individualized process, it is possible to draw some general conclusions about what makes text difficult in order to develop the knowledge base for this prototype. Based on the above analysis, the knowledge base should teach students that difficult vocabulary, challenging syntax, and length of text are all features that may complicate comprehension.

Another aspect of text complexity that is important to consider is the knowledge demands it places on the reader. Research has defined background (or world) knowledge as (1) prior knowledge about the text's topic; (2) prior knowledge that the text is about that topic (context v. no context) and (3) the extent to which the text's vocabulary clarifies the topic (Carrell, 1983). Because the first two knowledge demands are reader-specific rather than text specific, let's take a closer look at their connection to comprehension.

One study, popularly referred to as "The Baseball Study", found that prior knowledge in the content area of the text was the most important factor in predicting a child's ability to recall and summarize a text (Recht & Leslie, 1988). In the study, students were divided into four categories: low-reading skills/low-knowledge of baseball; low-reading skills/high-knowledge of baseball; high-reading skills/low-knowledge of baseball; high-reading skills/high-knowledge of baseball. The study found that students with high-reading skills and high-knowledge of baseball were best able to summarize a text about baseball and recall information from it. It found that the group that had the second highest success was the group that had low-reading ability but high-knowledge of baseball. Students who had high-reading ability, but low-knowledge of baseball were no more likely to accurately summarize or recall the text than low-readers who also had low-knowledge of baseball. The researchers concluded that instructors need to spend at least as much time building content knowledge as skill building. This study also supports the theory that text complexity is determined in part by the extent of reader's knowledge about the topic.

Another study found that prior knowledge that the text is about a particular topic also improves comprehension. In this study, researchers introduced readers to a pre-reading strategy called THIEVES. This strategy aims to help students understand text structure as well as "steal information" prior to reading the text (Al-Faki & Siddiek, 2013, p. 42). The researchers found that activating prior knowledge led to increased comprehension of a text, in this case, for English Language Learners. This further supports the idea that a reader's knowledge of the topic impacts text complexity.

Because a reader's knowledge about a topic impacts the text complixity, the knolwedge base for an ITS that supports comprehension monitoring should also teach students that topics they are more familiar with may be easier to comprehend while topics they know little about may be more difficult to comprehend.

Procedural/Conditional Knowledge and Comprehension Monitoring

When talking about metacognition, declarative knowledge can be simplified to refer to *what* a person knows about a strategy. Procedural knowledge then refers to knowing *how* to execute the strategy, and conditional knowledge refers to knowing *when* to execute the strategy and *why* that is the right choice (Jacobs & Paris, 1987; R. Lorch, E. Lorch, & Klusewitz, 1993; Schraw & Moshman, 1995). These three knowledge domains are related (Flavell, 1979). Figure 2 was designed to provide an example of this relationship

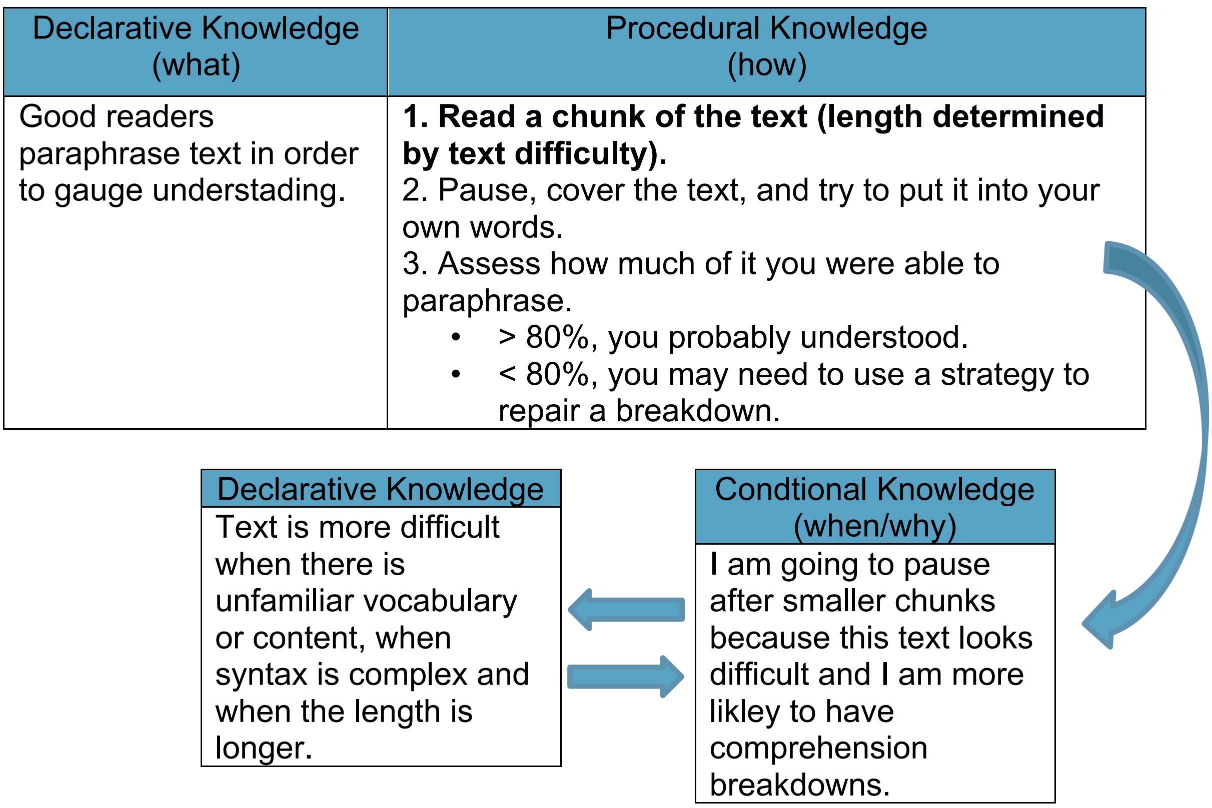

Figure 2. Example of the Relationship Between the Three Knowledge Domains

As illustrated, these knowledge domains can be intertwined (Sun, Merrill, & Peterson, 2001). This is one challenge to building the necessary knowledge base for this ITS. Furthermore, the ill-defined nature of reading means that there are innumerable strategies readers can employ depending on the features of that specific task (Jacovina & McNamara, 2017).

To further understand the complexity involved in designing the knowledge base for an ITS that supports comprehension monitoring, consider the fact that readers must be taught what, how, when and why for every comprehension monitoring strategy (Jacobs & Paris, 1987), and that the interrelatedness of each component makes logical scaffolding a challenging task. Consider, for example, just one learning goal in the knowledge base: *Students will be able to monitor their comprehension as they read*. This is a seemingly straightforward goal; however, students need to build declarative, procedural, and conditional knowledge (and not necessarily in a specific order) to achieve it. Figure 3 has been created to illustrate this complex relationship and demonstrate the amount of knowledge needed for just one learning goal (D=Declarative Knowledge, P=Procedural Knowledge, and C=Conditional Knowledge). This underscores the complexity of building the knowledge base for this ITS.

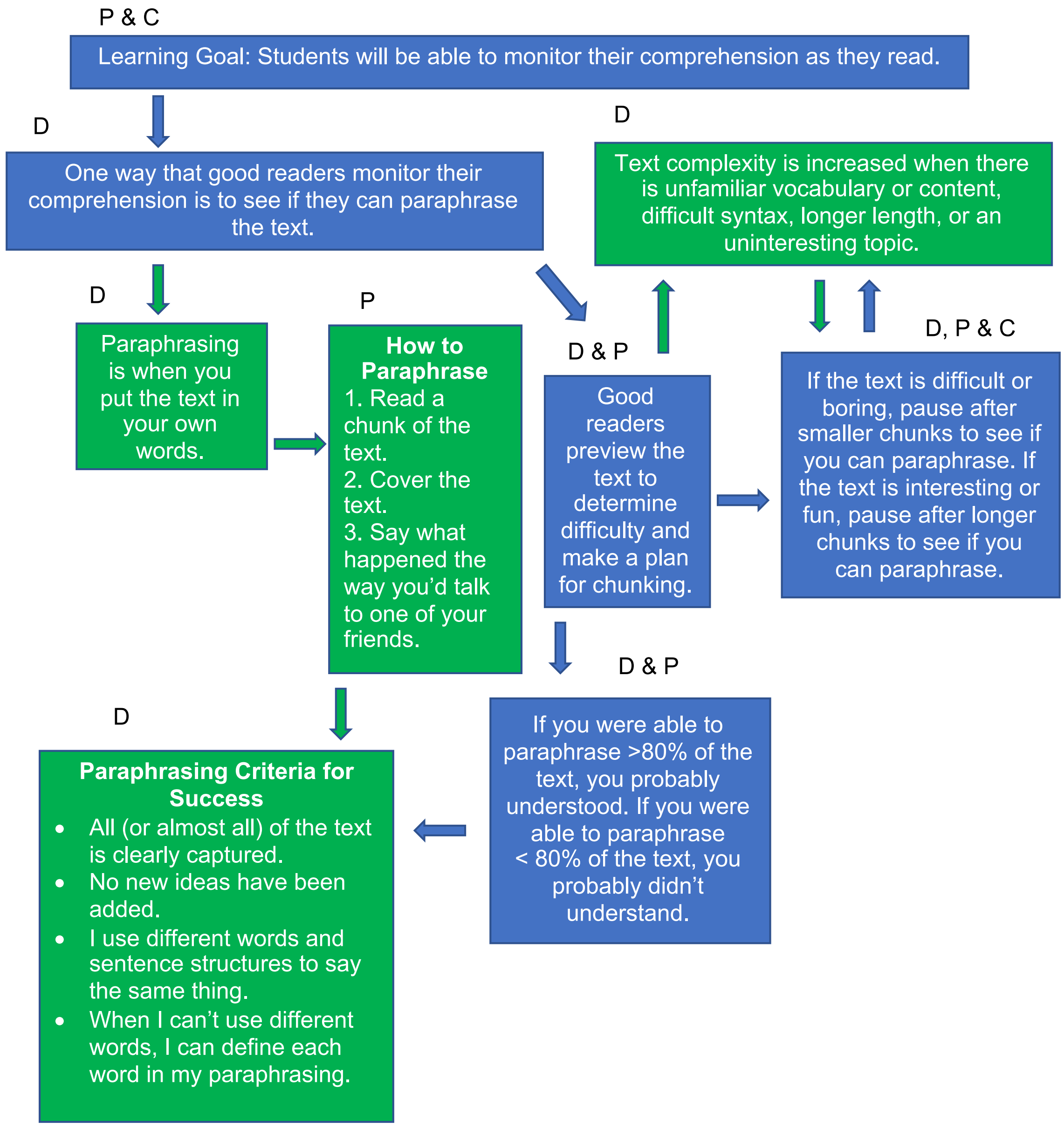

Figure 3. Learning Goal Breakdown into Knowledge Domains

This only captures the first learning goal. The metacognition required to monitor comprehension also includes applying strategies when there is a breakdown and different strategies when there is not one. However, given the time constraints of this study, this will not be explored; knowledge beyond building awareness will not be built into the knowledge base.

Even without the additional aspects of metacognition, there are still countless learning goals to consider, prioritize and scaffold in the knowledge base for this prototype. In order to do this effectively, it is useful consider research in effective curriculum planning. One framework that stands out is the Understanding by Design framework, first developed by Grant Wiggens and Jay McTiegh in 2000.

This framework encourages instructors to first think of the desired student outcomes and then create a backwards map of all of the content and skills necessary to reach this goal (Wiggens & McTiegh, 2011). Use of this framework has been associated with improved scaffolding in lesson planning and transparent and explicit instruction (Bowen, 2017).

2.1.2 Monitoring of Cognition

Flavell (1979) argues that metacognitive monitoring occurs in the action and interactions of metacognitive knowledge, metacognitive experiences, goals (or tasks) and actions (or strategies). A metacognitive experience can be described as having a moment where one notices something about where they are at in relation to their goal and what progress they are likely to make towards it. It is related to metacognitive knowledge because metacognitive knowledge helps one become aware in the first place, and the act of awareness adds to one's metacognitive knowledge bank. When one has a metacognitive experience, they must reflect on the goals/task in front of him or her and make some decisions regarding the task. For instance, they must consider whether or not the goals are currently appropriate. If they are, they may choose to move forward as planned. If a different goal needs to be met first in order to provide support towards the ultimate goal, the current aim may shift. Actions (or strategies) refers to both cognitive and metacognitive strategies. According to Flavell (1979), cognitive strategies are the approaches one might choose in order to make progress, and metacognitive strategies are those one might use to monitor it.

Other researchers describe metacognitive monitoring as a process of planning, monitoring, and (re)evaluation (Jacobs & Paris, 1987; Schraw & Moshman, 1995). While the language may be different, there is significant overlap between this approach and Flavell's. When boiled down, monitoring of cognition involves the reoccurring process of analyzing the goal, selecting appropriate strategies to achieve the goal, applying those strategies and assessing their effectiveness, and then reconsidering the goal in light of this analysis (Flavell, 1979; Jacobs & Paris, 1987; Schraw & Moshman, 1995). To do this well, one must have significant knowledge of cognition as it relates to the task as well as the ability to execute this process with a level of automaticity.

In order to build the knowledge base of this ITS, it useful to consider how these two facets of metacogntion interact in reading comprehension. The following section illustrates these interactions and discusses the implications for this study.

Monitoring of Cognition and Comprehension Monitoring

Researchers have found that good readers tend to be metacognitive readers (Jacobs & Paris, 1987; National Research Council, 1998; Mirandola, Ciriello, Gigli, & Cornoldi, 2018). These readers understand that comprehension is an active process which requires a significant amount of strategizing and reflecting rather than just simple decoding of words (Jacobs & Paris, 1987). This active reading process includes comprehension monitoring and using repair strategies when understanding breaks down. The use of declarative, procedural, and conditional knowledge is integral to this process.

Readers who monitor for comprehension well are effective at judging their own comprehension. They recognize when they can continue moving forward in the text and when they need to pause to adjust goals and apply strategies (Mirandola et al., 2018). According Markman (1979), noticing a breakdown might occur at a superficial level, such as realizing you don't understand the jargon being used or that you've lost track of the subject in the sentence. Comprehension breakdown might also occur at a more complex level. One may, for instance, understand each sentence or component of a text individually, but struggle to make connections between ideas in order to synthesize and understand the deeper meaning (Markman, 1979). A strong reader might recognize that they failed to comprehend at either of these levels and can then activate knowledge and strategies to repair their understanding. Less effective readers may not have the knowledge or strategies for repair. They may also fail to realize that they did not understand and therefore neglect to adjust course in order to gain meaning.

Readers who effectively monitor their own comprehension, and are therefore more likely to be strong readers, are often not aware that they have engaged in this process (Baker, 1979). Indeed, this is the ultimate goal of this ITS, as automated comprehension monitoring frees up a reader's working memory. This supports deeper comprehension as it frees up space for other impactful strategies such as summarizing and predicting (Bryant, Cain, & Oakhill, 2004; U.S. Department of Education, 2010).

In order to support automation, it may be useful to develop a framework for metacognitive monitoring that helps students organize the extensive cognitive knowledge necessary to support comprehension monitoring. This framework may also support teacher (or tutor) clarity around the strategies, which can further aid comprehension monitoring, (Fisher, Frey & Hattie, 2016).

2.1.3 Implications for the Knowledge Domain

The previous analysis of research in the field of metacognition and comprehension monitoring has led to the following conclusions about the knowledge domain for this prototype:

- Students must build specific declarative knowledge:
  - Reading is an active process that involves thinking.
  - Good readers pay attention to how well they are understanding.
  - One way that good readers know they understand is they can put the text in their own words.
  - Vocabulary, syntax, topic familiarity, and length are four text features that can be used to assess difficulty.
- Students must build specific procedural and conditional knowledge such as:
  - how good readers use paraphrasing to assess understanding.
  - when good readers use paraphrasing to assess understanding.
- The Understanding by Design framework can be leveraged to identify, prioritize, and scaffold specific learning goals.
- A framework for metacognition can simplify the relationship between the metacognitive process and cognitive knowledge, two key components for effective comprehension monitoring.

Using these principals, a framework has been developed for building awareness of comprehension. Please see Apendix 1. Metacognitive Toolbox.

2.2 Building the Pedagogical Model

In order to design this prototype, it is next important to explore research that informs the ideal learning approach for the pedagogical model. This model will include explicit modeling of skills. The following section will discuss a specific strategy for this explicit modeling, as well as other instructional design considerations to support internalization. It will then investigate pedagogical models currently in use and will synthesize this information to identify the ideal learning approach for this pedagogical model.

2.2.1 Comprehension Instruction

In order to improve comprehension, students need to see models of thinking that support comprehension monitoring (Collins, Brown, & Newman, 1989). One way many educators make their thinking visible is through cognitive modeling, or the think-aloud. The think-aloud has been proven to be an effective way to support reading comprehension (Baumann, Seifert-Kessell, & Jones, 1992; Sönmez & Sulak, 2018; Wade, 1990). In a think-aloud for reading, this may look like the instructor reading the text, and then verbally describing the process he or she goes through to make sense of the text as they read. The best think-alouds include a rationale for *why* teachers think the way they do. The think-alouds for this ITS should include examples of monitoring understanding and noting why breakdowns occurred.

After students see the model, they need multiple, authentic, and varied practice opportunities to help them apply strategies in a variety of contexts (Rupley, Blair, & Nichols, 2009). Research shows that this practice should be purposeful and systematic, or deliberate (Ericcson, Krampe, Tesch-Romer, 1993). In this work, Ericcson et al. (1993) argue that expert performance is the result deliberate practice rather than innate ability. Though the level of importance he ascribed to deliberate practice in determining expertise has since been successfully challenged, researchers still agree that it is part of the equation for success (Hambrick, Oswald, Altmenn, Meinz, Gobet, Campitelli, 2014; Macnamara, Hambrick, & Oswald, 2014).

The need for deliberate practice provides further support for the creation of a framework to support comprehension monitoring. While the specific strategies students apply may be different given the breakdowns and features of text, the process can remain the same. Being able to use the same process in a wide variety of contexts gives students the necessary opportunities for deliberate practice, which increases the likelihood that these processes will become automatic (Coyle, 2009).

When thinking about comprehension instruction, it is also crucial to remember its role as one part of the curriculum rather than the entire curriculum. Educators skeptical of the push for metacognitive instruction worry that students will solely be taught the technical aspects of reading and not the social and emotional aspects (Jacobs & Paris, 1987). Additionally, recent research concludes that students who read just 15 minutes per day demonstrated greater reading ability (Renaissance Learning, 2019). Thus, while the instruction needs to provide explicit modeling and opportunities for practices, it also should (and can) be succinct.

2.2.2 Existing Pedagogical Models

When determining the appropriate learning design, it useful to consider the pedagogical models already in use, particularly those whose design has been informed by research. This section will explore three existing models.

With *The Dynamic Support of Contextual Vocabulary Acquisition for Reading (DSCoVAR)*, an ITS that is currently being developed with funding from the US Department of Education, students are taught strategies they can use when they come across words they don't know. ("Dynamic Support of Contextual", n.d.) While the goal of this ITS is to support Tier 2 vocabulary acquisition, the program can also help students learn effective solving strategies for when comprehension fails due to a vocabulary breakdown, as it explicitly teaches students *how* to use context to make sense of unknown words. This system also provides nuanced feedback to students when their responses are incorrect or only partially correct (Jacovina & Mcnamara, 2017).

The *Intelligent Tutoring System of the Structure Strategy* is designed to help middle readers use text structure to make sense of connections between ideas in expository and argumentative texts. The system provides over 100 lessons on the structure strategy; it appears that these strategies include explicit instruction on *how* to use the strategy. It also provides a tutor that helps them with the reading of the passages, and feedback on students' written main idea statements. This system is adaptive in that it not only provides feedback to students based on their written responses; it also adjusts lessons based on student performance. Research about the impact of this ITS demonstrated its ability to markedly improve students' reading comprehension scores (Meyer, Wijekumar, & Lin, 2011).

Interactive Strategy Trainer for Active Reading and Thinking (iSTART) was designed to increase the impact SERT (Self-Explained Reading Training) could have on student achievement. This strategy is drawn from cognitive science research which concludes that students can use self-explanation to solve problems (Chi, Bassok, Lewis, Reimann, Glaser, 1989). SERT applies this to reading comprehension by arguing that self-explanation can support inferencing, problem solving, mental modeling, and deeper understanding (McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). The program teaches four key strategies. The first strategy is comprehension monitoring, where kids are taught to determine whether or not they understand the text. If students determine they do not understand, they are encouraged to try other strategies to repair the breakdown. One strategy is paraphrasing. This serves two goals: put the material into a more familiar mental representation and assess comprehension. Students are then encouraged to elaborate, where they use their world knowledge and knowledge of the text to make logical elaborations on the text. This encourages students to draw inferences. When students cannot complete this, SERT pushes students to use their world knowledge and knowledge of the text to fix the breakdown. Students are also taught to make predictions about upcoming material as a way to encourage metacognition. And finally, students are taught a strategy called bridging. In this strategy, students are taught that they should connect different parts of the text in order to understand the relationship between ideas.

iSTART was developed because researchers realized that scaling the SERT training was difficult, and they wanted to give more students access to the strategies. The program uses a character to serve as the "teacher" who guides students through the lessons. The lessons begin with an introduction to the strategy and a model of characters using the strategy. It is followed by a demonstration where characters demonstrate self-explanation using several of the strategies. In this level, the program can increase or decrease the level of scaffolding depending on student data. In the final module, the reader can practice the reading strategies he or she has learned. The ITS assesses the self-explanations and provides feedback to the student. This feedback includes information on which strategy to try. Researchers found that this approach had a significant impact on comprehension (McNamara et al., 2007).

What is unclear in iSTART (and the second-generation version, iREADY) is the extent to which it teaches students *how* to use the strategies (procedural knowledge), which Grossman (2013) and her colleagues found to be a crucial component of driving student achievement. McNamara (2017), lead researcher and developer for iSTART, mentions in one research paper that the metacognitive instruction in iSTART is, indeed, implicit, which seems to contradict other research about the need for explicit strategy instruction (McCarthy, Johnson, Likens, Martin, & McNamara, 2017).

2.2.3 Implications for the Pedagogical Model

The previous review of research related to comprehension instruction and existing pedagogical models underscores the need to ensure the pedagogical model:

- Includes a think-aloud to explicitly model the strategy.
- Includes deliberate practice.

Additionally, the pedagogical model designed for this study will differ from the model used in iSTART, the closest competitor, in two ways. First, paraphrasing will not be used as a solving strategy in this prototype. Following Markman's (1979) research, it will be used as a measurement tool against which students can assess their comprehension. Second, all instruction in this prototype will address declarative, procedural, and conditional knowledge; the explicit modeling will include an explanation of how and when to use specific strategies.

2.3 Developing the Student Model

One main difference between an ITS and other software that aims to promote learning is the adaptable nature of the system. This is captured in the design of the student model. The following section describes key features of this adaptability before explaining the challenges to creating this model. It concludes by naming the key features that remain in the minimally adaptive prototype.

One feature of adaptability for an ITS is its multimedia environment, which can create a highly integrated learning experience for students by offering significant levels of choice and control (Cairncross & Mannion, 2001). Intelligent tutoring systems also provide students a more personalized experience through individualized practice and feedback (Heilman, Collins-Thompson, Callan, & Eskenazi, 2006; Kegel & Bus, 2012). John Hattie (2012) points out that with an effect size of .75, providing feedback is one of the most important things an instructor can do to drive student achievement (p. 269). Thus, leveraging technology to support this could be highly impactful. Furthermore, training teachers to deliver high quality feedback can be costly and time consuming, and an ITS can easily reach a large number of students with relatively little teacher intervention (McNamara et al., 2007).

However, designing a model that achieves the above aims, particularly in the domain of reading comprehension, presents numerous challenges. Jacovina and McNamara (2017) have even argued that given the ill-defined nature of reading, it may be impossible to model the same success developers have had with ITSs for math instruction. What is more is that a key advantage of an ITS is its ability to provide feedback to students, but feedback is most effective when it is contextualized (Hattie & Timperley, 2007). Given that effective execution of reading skills is also highly contextual, developers face significant obstacles to providing personalized and contextual feedback (Jacovina & McNamara, 2017). Furthermore, students need to read authentic texts in order to become effective readers (Duke & Pearson, 2009). However, finding authentic texts that provide multiple opportunities for students to practice the targeted skill can be challenging, as developers of the ITS, REAP, found (Heilman et al., 2006).

The challenges inherent in designing a highly adaptive system has led some developers to deemphasize the student model (Sedlmeier, 2001). Given the large effect size Hattie (2012) found for providing feedback, it is clear that this feature is a crucial component of the student model. However, higher levels of teacher clarity in the pedagogical model may make it possible to somewhat decontextualize this feedback.

2.4 The Research Question

Comprehension monitoring is an important skill that leads to deeper and more accurate understanding of text. And yet, explicit instruction in classrooms around comprehension monitoring is limited and consistently of poor quality. There is evidence that intelligent tutoring systems may be employed to support this skill, but the highly complex nature of reading comprehension presents numerous challenges.

The aim of this research is to explore the possibility of using a minimally adaptive intelligent tutoring system to support comprehension monitoring. The objective of this research is to determine to what extent a minimally adaptive tutoring system for comprehension monitoring can impact a students' awareness and metacognition while reading. The purpose of this research is to establish evidence that such a system has potential to impact student comprehension. The research and findings will be used in the development phase for the product.

The main research question is:

Does a minimally adaptive intelligent tutoring system designed to provide explicit strategy instruction for comprehension monitoring have potential to improve student comprehension on grade-appropriate text?

The sub-research question is:

- Does explicit instruction in noticing breakdowns in comprehension caused by unfamiliar vocabulary increase student awareness of these potential breakdowns?

The null hypothesis in this study is that there is no statistically significant difference in outcomes for students who participate in the intervention as compared to the control group ($H0: \mu d = 0$). In other words, students in the intervention will show no difference in their ability to notice a planted error, notice vocabulary breakdowns, or paraphrase grade-appropriate text. The alternative hypothesis is that there is a statistically significant difference ($H1: \mu d \neq 0$) in the students' ability to demonstrate these skills.

3 Research Design and Methodology

3.1 Introduction

The National Science Foundation Directorate for Engineering defines proof-of-concept research as "the realization of a certain method or idea to ascertain its scientific or technological parameters. A proof-of-concept should be understood sufficiently so that potential application areas can be identified and a follow-on working prototype designed" (National Science Foundation, 2014, section II., para. 3). Proof-of-concept research is often used for research that is in the early stages where the aim is to demonstrate the scalability of a concept (Kendig, 2016). It is an appropriate model for this study given that one application of the data is to determine whether or not resources should be allocated to design a complete prototype. This research was been designed as a randomized control study.

3.2 Population of the Study

The sample was chosen from students at an urban charter school in Philadelphia. 94% of the students at this charter school qualify for free or reduced lunch, which is often used as an indicator of relative poverty in the United States (Snyder & MUSU-Gillette, 2015). The school serves over 900 students in grades 6-12. The demographic breakdown of the school is presented in Figure 4.

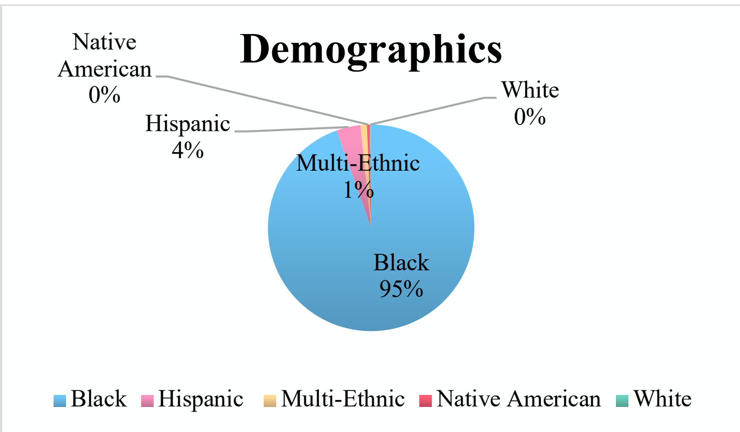

Figure 4. Intervention School Demographics

In the selected 8th grade classroom, 96% of the students identified as Black (28 students) and 4% (2 students) identified as Hispanic. 63% identified as male (19 students) and 37% identified as female (11 students). All qualified for free or reduced lunch.

Reading levels varied in the class, though only three were reading at a Lexile level that fell within the 6th-8th grade band, as determined by the Measure of Academic Progress (MAP) test. This is a computer-based standardized test used in US schools to determine reading levels. Class reading level is presented in Figure 5.

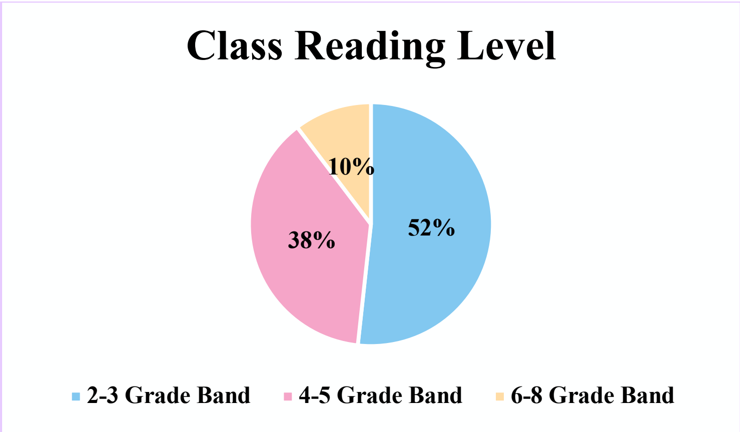

Figure 5. Class Reading Level

3.3 Sampling Procedure

30 students from an 8th grade English Language Arts (ELA) class in this urban charter school in Philadelphia were given the pre-test. The tests were mixed up and 20 were randomly pulled from the pile. The chosen pile was checked to ensure even distribution of students who identify as girls and students who identify as boys. The procedure was repeated until there were ten boys and ten girls in the pile. Then the pile was divided into two groups of ten, again, working ensure an even distribution of girls and boys. Pile one became the intervention group and pile two became the control group. Three of the students chosen were absent on the second day of the intervention: two from the intervention group and one from the control group. In order to maintain like-comparisons, eight students from the nine present in the control group were randomly selected for the data analysis.

3.4 The Sample

This section outlines the demographic data for the Intervention Group and Control Group. It is organized in tabular form in order to provide an easy comparison between the groups as well as between each group and the whole class. Reading levels by grade band for the whole class, intervention group, and control group is presented in Table 2. The gender makeup of the whole class, intervention group, and control group is presented in Table 3. And the racial makeup of the whole class, intervention group, and control group is presented in Table 4.

TABLE 2. Reading Level Breakdown

Reading Levels	**Class**	**Intervention Group**	**Control Group**
2-3 Grade Band	15	4	4
4-5 Grade Band	11	4	4
6-8 Grade Band	3	0	0

TABLE 3. Identified Gender Breakdown

Gender	**Class**	**Intervention Group**	**Control Group**
Male	19	4	4
Female	11	4	4

TABLE 4. Racial Breakdown

Race	**Class**	**Intervention Group**	**Control Group**
Black	28	7	8
Hispanic	2	1	0

3.5 Instrumentation

Numerous tools have been developed to measure one's metacognition while reading. They typically either measure one's knowledge about cognition or their regulation of cognition. While these tools do offer a level of insight into a student's thinking, there are limitations to the insight as metacognition is such a complex process that accurate assessment is difficult (Ozturk, 2017). The following section describes some of the most commonly used tools, what they measure, and the drawbacks to using each one. It concludes by describing the instrument used for this study.

3.5.1 Tools to Measure Metacognition

Questionnaires

To test one's knowledge about cognition, researchers often turn to questionnaires (Baker & Cerro, 2000). The Metacognitive Awareness of Reading Strategies Inventory (MARSI), is one example. This tool was designed to determine the extent to which students used reading strategies while reading academic text (Mokhtari & Reichard, 2002). The researchers constructed this tool with the understanding that making meaning is a deliberate act that can therefore be measured. In this assessment, students were given 30 questions and were asked to score themselves on scale of 1-5 about the extent to which they use that specific global reading strategy, problem solving strategy, or support reading strategy. The researchers tested the final draft of the tool on a pilot group of 443 students in grades 6-12 and found the reliability for the total sample to be .89. They also found correlation between high ratings and reading ability. One limitation of using this tool is that "students' perception of strategy use are a reflection of a moment in time rather than a reflection of their reported strategies across different times, texts, or tasks" (Mokhtari, Dimitrov, & Reichard, 2018, p. 222). Another limitation is that awareness of a strategy does not guarantee that students actually use the strategy. Thus, the researchers caution that use of this tool should be complemented with additional assessments.

Error Detection

Another assessment tool that is often used is the error detection approach. This approach is most often used to assess regulation of cognition, or comprehension monitoring (Baker & Cerro, 2000). In this approach, the student is asked to read a text that contains an error and is monitored to determine whether or not they identify the error. This approach is built upon the theory that students who are actively monitoring their comprehension will notice the problematic text. The issue is that readers don’t treat all breakdowns alike. Baker (1979) found that college students were significantly more likely to notice a problem that interrupted their understanding of main idea than problems with details in a passage. This could suggest that readers prioritize breakdowns that have a more significant impact on their understanding. Failing to call attention to the breakdown may simply mean that it was not considered an important breakdown rather than indicate a lack of metacognition. One additional conclusion Baker drew is that readers will often go to significant lengths to impose sense on a text, which highlights another limitation of this tool. The reader may have noted the inconsistency, but he or she applied solving strategies so quickly that they did not register the breakdown. The research also revealed the fact that though many readers are capable of comprehension monitoring, they may not do so consistently. Additional research notes that this approach is problematic because texts don’t generally contain errors and inconsistencies (Baker & Cerro, 2000).

Verbal Reports

A third approach researchers often take to assessing metacognition is to use verbal reports. This is commonly seen in the form of a think-aloud (Bereiter & Bird, 1985; Wade, 1990). In this assessment, participants describe their thoughts verbally while they read a passage. The researcher will note their thoughts, only interjecting to push the participant to tell them more. The researcher will also take note of the participant’s non-verbal cues like visible signs of frustration or engagement. This approach can be used to measure both metacognitive control and metacognitive knowledge. In spite of this approach’s potential to provide in the moment data, there are many drawbacks to this assessment. First, pausing to verbally describe one’s thinking can disrupt the reading process and place more stress on one’s working memory—a key resource in the reading process. Second, participants, particularly struggling readers, may not feel comfortable describing their thinking, or may not have the language to express themselves accurately. Finally, think-alouds can be difficult to score, and often require two scorers for inter-rater reliability. (Ozturk, 2017).

3.5.2 Instrumentation for this Study

Given the time constraints for this study, questionnaires were not used, as results of this survey should be complemented with another assessment. Similarly, verbal reports were rejected due to the fact that they often require two scorers for inter-rater reliability. This left the Error Detection Paradigm, which was used in this study to assess metacognition.

As shown, this instrument is not without limitations. In their 1982 study to determine what conditions were likely to facilitate error detection, Winograd and Johnston dissected many of these issues. Still, they concluded that the error detection paradigm is best used to assess the reader's ability to explicitly notice and name their breakdown (Winograd & Johnston, 1982). Baker (1979) also argued that students are more capable of noticing errors when explicitly instructed to do so. While she mentioned this as a limitation of the tool, this fact makes it appropriate for this study given that part of the strategy instruction involves making students aware that they should notice breakdowns while they read.

The specific instrument used for collecting data has been designed by the researcher and is based on the instrument described in Winograd and Johnston's 1982 study using the error detection paradigm. The version used in this study has not been validated.

The instrument consisted of two authentic paragraphs for the pre-test and two different paragraphs for the post-test. The decision to use authentic paragraphs was based on Rupley, Blair and Nichol's 2009 study that argued for the use of authentic texts in reading instruction. One paragraph was a fictional passage and one was a non-fiction passage. Because the subjects in this experiment were 8th graders and the research question specifically asks about grade appropriate texts, each paragraph fell towards the upper end of the 6-8th grade band as defined by the Common Core State Standards. Each paragraph was 5-6 sentences long and included a made-up word chosen at random from a free, online fake-word generator. This word was placed about 60 words into the paragraph in a researcher-written simple-sentence that logically connected to the previous sentence. The paragraphs and placement of the error were kept short in order to test for the ability to notice breakdowns rather than the ability to notice breakdowns over a long period of text. The latter would be testing stamina as much as skill. Baker's (1979) research supported this, as she noted that errors placed higher in a paragraph were noticed more often than errors placed lower in the paragraph.

Baker (1979) also noted that failure to report error detection did not necessarily mean the subjects failed to notice it; they simply fixed the error and moved on. In order to control for this possibility, the instrument included a direction to paraphrase the text. If students could accurately paraphrase the text in spite of the error, it would suggest that their response to the planted error was similar to the subjects' response in Baker's (1979) study. Paraphrasing was also included as a measure of comprehension. Student paraphrasing was assessed using a non-validated, researcher created rubric that rates a student on a scale of 1-4 for their ability to paraphrase. See below for the ratings:

1. Paraphrasing is inconsistent/inaccurate
2. Paraphrasing is more consistent/accurate; around half the text is captured either in their own words or by noticing a breakdown; key pieces may be inaccurate
3. A significant portion of the text is captured in their own words or by noticing a breakdown; a significant portion of the understanding is accurate
4. Most to all of the text is captured in their own words or by noticing a breakdown; most to all of the understanding is accurate.

Students were tested as an entire class. They were directed to read each passage, identify anything that was confusing, and then paraphrase the passage. These directions were all written as step-by-step directions under each passage. These tests were conducted on paper and handed in to the researcher.

3.6 Intervention Design

The materials for the intervention included three lessons designed to take no more than 15 minutes. The first two lessons included a short video (2-4 minutes) that named the objectives for the lesson and provided an explicit model of the targeted comprehension skill. The objectives for the first two lessons were the same for all students and focused on building awareness around their understanding. To build their declarative knowledge, students were taught that good readers continually assess their understanding of text, that paraphrasing is one way to do this, and that text is often more difficult when it is boring or includes difficult vocabulary. To build their procedural and conditional knowledge, students were then shown *how* and *when* to use paraphrasing to assess their understanding. After the video, students completed practice using Google Forms. This practice included a review of the introduced declarative knowledge as well as practice using the skills. Where possible, the automated feedback feature on Google Forms was used to respond to student answers. This simulated adaptive technology.

The third lesson also included a brief video that named the objectives and provided an explicit model of the targeted comprehension skill. The objectives for this lesson, however, were more targeted. The researcher analyzed the student work from the first two lessons and used the data to determine the learning outcome for each student. Depending on their data, students were either given additional support with noticing breakdowns or instruction to improve their paraphrasing.

All passages used were authentic texts, and their Lexile levels fell within the 6-8 grade band. The texts alternated between fiction and non-fiction. Students completed these lessons on school issued Chromebooks using Google Classroom.

3.7 Procedure

The intervention was run over the course of two consecutive school days. Students in both the control group and the intervention group were given 30 minutes to complete the pre-test in their first period ELA class. Directions for how to complete the test were given to all students as a group. These directions were repeated when students had individual questions. The word “paraphrase” was also defined for students when asked. This procedure was completed at the end of day two for the post-test.

For manageability, the intervention group was split into two groups of five students. Each group completed a total of three lessons over the course of the two days. The first two lessons were completed on day one, and the third lesson was completed on day two. The intervention was run in a nearby empty classroom to minimize distractions to both the intervention group and the control group.

3.8 Data Collection

Students participating in the intervention were pulled from their ELA class in order to ensure that potential improvement in reading scores could not be attributed to additional time spent in reading instruction over the course of the intervention. Subjects belonged to the same cohort of students, meaning they saw the same instructors over the course of the school day. This helped isolate the intervention as the variable rather than varying teacher quality. This is important as teacher quality has been found to have the largest impact on student achievement (RAND Corporation, 2012). The support provided by the researcher was limited to technical support to log onto the program and directions for the pre and post-test. Students who lacked motivation to complete the intervention lessons as prescribed were not redirected or given additional prompts. This protected against external motivation or strong relationships as variables to consider in data analysis.

3.9 Ethical Considerations

According to American Educational Research Code of Ethics (AERA, 2011), education researchers work to avoid harm. This study had the potential to cause harm in several ways.

First, students who were already behind were to be pulled from their class to receive an intervention whose efficacy had yet to be proven. In order to mitigate this harm, the researcher worked with a supervisor at the school to determine which cohort this would impact the least. The class chosen was one whose teacher was in her second year and was still struggling to deliver strong instruction.

Second, the intervention provided minimal scaffolding for texts that are at minimum three levels beyond the subjects' reading level. This could harm the subjects by further frustrating already struggling readers and causing shutdown. This was addressed in two ways. First, in early lessons, the researcher defined success as "noticing there was a breakdown" rather than "getting it right." This allowed all students, regardless of reading level, the opportunity to demonstrate success. Second, the researcher included one specific piece of strength-based feedback at the opening of the third lesson.

Finally, in schools where students have faced significant trauma, like this urban charter school, students are more likely to struggle to regulate their emotional response. Outbursts, sometimes violent, can happen, especially when students feel out of their comfort zone or are redirected by an adult with whom they do not have a relationship (National Scientific Council on the Developing Child, 2005/2014). In order to minimize this risk, the researcher split the intervention group into two groups of five. This allowed her to give students space in the testing room and respond more quickly if needed.

It was also important to note that the students who participated in the study were minors. Before they were allowed to participate in the study, they needed to turn in a permission slip signed by a parent or guardian. The permission slip described the study and provided the researcher and principal's contact information for any parent who had additional questions or concerns.

3.10 Data Preparation

Student pre-tests and post-tests were collected. Each student was assigned a number and names were blacked out to preserve anonymity. Each passage was scored in four categories.

1. Error Detection: Subjects who circled, highlighted or wrote the error on the lines provided were assigned a 1. Subjects who did not acknowledge the error were assigned a 0.

2. Vocabulary Specific: Subjects who made note of *any* vocabulary breakdown, including the made-up word and any other word that they notice causing breakdown, were assigned a 1. Subjects who did not acknowledge a breakdown were assigned a 0.

3. Noticing ANY Breakdown: Subjects who made note of *any* breakdown by circling, highlighting or writing on the lines were assigned a 1. Subjects who did not make note of any breakdown were assigned a 0.

4. Paraphrasing: Each passage was given a paraphrasing score from 1-4.

Students' ability to detect error was determined by the following prompt:

"Was there anything in the passage that was confusing or didn't make sense to you? Use the highlighter to circle or highlight it in the passage or copy on the lines below. If you can, write down why it confused you."

These breakdowns were considered general breakdowns if students included larger chunks of text that were highlighted or copied, such as complete sentences, or if students included the text and made a statement about their general confusion. When students circled, highlighted, or copied specific words, it was considered a vocabulary breakdown. Similarly, when students copied larger sections of text and stated something to the effect of "I don't know what these words mean," it was counted as a vocabulary breakdown.

4 Results

4.1 Introduction

Data from the 16 subjects, 8 in the intervention group and 8 in the control group, is reported in this section. Several comparisons were made for each score category in order to determine if the ITS has the potential to impact student outcomes, and which (if any) students are most likely to experience this impact. These include comparing the intervention group to the control group, readers at the 4^{th}-5^{th} grade Lexile level in the intervention group to the same level readers in the control group, and readers below a 4^{th} grade Lexile level in the intervention group to the same level in the control group. In each comparison the raw scores as well as their statistical significance are reported.

The null hypothesis in this study was that there would be no difference in outcomes for students who participate in the intervention as compared to the control group ($H0: \mu d = 0$). In other words, the intervention neither supported nor hurt comprehension or their ability to notice the planted error, vocabulary-specific breakdowns, or paraphrase text. The alternative hypothesis was that there is a difference ($H1: \mu d \neq 0$). This would mean that the intervention helped or hindered student awareness and/or paraphrasing.

4.2 Data Presentation

Error Detection

Table 5 shows that 3 more students in the intervention group noticed the planted error in the post-test than in the pre-test, which is a 300% increase. It also shows that fewer students in the control group noticed the planted error in the post-test than in the pre-test.

TABLE 5. Whole Group Error Detection Pre-Test vs. Post-Test

	Pre-test		**Post-Test**	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	1	7	4	4
Control	2	6	1	7

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.375, SD=.512) and the control group (M=.125, SD=.35); t(14)=1.13, p=.278. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

Table 6 includes all students from the intervention group. It shows (in bold) that each of the students who did not notice the planted error *did* notice a vocabulary breakdown and that this breakdown occurred before the planted error. None of the students who failed to notice the planted error noticed a breakdown in vocabulary after the planted error.

TABLE 6. When Intervention Students Noticed Vocabulary Breakdowns

Student	**Notice Error?**	**Notice vocabulary breakdown?**	**Before Error?**	**After Error?**
IS1	**No**	**Yes**	**Yes**	**No**
IS2	Yes	*NA*	*NA*	*NA*
IS3	Yes	*NA*	*NA*	*NA*
IS4	**No**	**Yes**	**Yes**	**No**
IS5	Yes	*NA*	*NA*	*NA*
IS6	**No**	**Yes**	**Yes**	**No**
IS7	**No**	**Yes**	**Yes**	**No**
IS8	Yes	*NA*	*NA*	*NA*

Table 7 includes students who read at a 4^{th}-5^{th} grade Lexile level. It shows there was a 200% increase in students in the intervention group who noticed the planted error in the post-test as compared to the pre-test. It also shows that there was no change in the number of students in the control group who noticed the planted error in the post-test as compared to the pre-test.

TABLE 7. 4^{th}-5^{th} Grade Lexile Level Error Detection

	Pre-test		Post-Test	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	1	3	3	1
Control	1	2	1	2

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.5, SD=.577) and the control group (M=0, SD=0); t(5)=.102, p=.203. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

Table 8 includes students who read below a 4^{th} grade Lexile level. It shows that in this group, more students in the intervention group noticed the planted error in the post-test than in the pre-test. Percent increase cannot be calculated because 0 students noticed the error in the pre-test. This table also shows a 100% decrease in the number of students who noticed in the error in the post-test as compared to the pre-test.

TABLE 8. Below 4^{th} Grade Lexile Level Error Detection

	Pre-test		Post-Test	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	0	4	2	2
Control	1	3	0	4

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.5, SD=.577) and the control group (M=-.25, SD=.5); t(6)=.049, p=.097. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

There was not a statistically significant difference in the ability of the intervention group to notice the planted error in the post-test as compared to the pre-test in any of the data configurations. However, there is consistent evidence that students in the intervention group were more likely to notice the planted error. There is also evidence to suggest that students who did not notice the planted error may not have due to issues of stamina rather than lack of metacognitive skills.

Vocabulary Specific

Table 9 shows that all of the students in the intervention group noticed at least one breakdown in vocabulary on the post-test, which is a 300% increase from the pre-test. It also shows that one fewer student in the control group noticed at least one vocabulary breakdown on the post-test, which is a 50% decrease from the pre-test.

TABLE 9. Whole Group Noticing ANY Vocabulary Breakdown

	Pre-test		**Post-Test**	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	2	6	8	0
Control	2	6	1	7

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was significant difference in the change in scores for the intervention group (M=.75, SD=.463) and the control group (M=-.125, SD=.354); t(5)=.0004, p=.0008.

Table 10 includes students who read at a 4^{th}-5^{th} grade Lexile level. It shows there was a 200% increase in students in the intervention group who noticed any vocabulary breakdown in the post-test as compared to the pre-test. It also shows that there was no change in the number of students in the control group who noticed the planted error in the post-test as compared to the pre-test.

TABLE 10. 4^{th}-5^{th} Grade Lexile Level Noticing ANY Vocabulary Breakdown

	Pre-test		**Post-Test**	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	1	3	3	1
Control	1	2	1	2

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was significant difference in the change in scores for the intervention group (M=.75, SD=.5) and the control group (M=0, SD=.0); t(5)=.026, p=.052. There is evidence to accept the alternative hypothesis (H1: μd ≠ 0).

Table 11 includes students who read below a 4^{th} grade Lexile level. It shows there was an increase in students in the intervention group who noticed any vocabulary breakdown in the post-test as compared to the pre-test. The percent increase cannot be calculated because the number of students who noticed the breakdown in the pre-test is 0. The table also shows that there a 100% decrease in the number of students in the control group who noticed any vocabulary breakdown in the post-test as compared to the pre-test.

TABLE 11. Below 4^{th} Grade Lexile Level Noticing ANY Vocabulary Breakdown

	Pre-Test		**Post-Test**	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	0	4	2	2
Control	1	3	0	4

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was significant difference in the change in scores for the intervention group (M=.75, SD=.5) and the control group (M=-.25, SD=.5); t(6)=.015, p=.030. There is evidence to accept the alternative hypothesis (H1: μd ≠ 0).

There was a statistically significant difference between the intervention group and control group as a mixed sample, as well as in the 4^{th}-5^{th} grade Lexile level and below 4^{th} grade Lexile level samples. This suggests that the intervention has the potential to impact student ability to notice vocabulary breakdowns. This is further supported by the intervention group's percent increase in noticing breakdowns on the post-test as compared to the pre-test. It also suggests that students who are closer to reading at a grade-appropriate level are better supported by the intervention.

Noticing ANY Breakdown

Table 12 shows that 3 more students in the intervention noticed any type of breakdown in the post-test as compared to the pre-test, which as an increase of 60%. 1 more student in the control group noticed any type of error in the post-test as compared to the pre-test, which is an increase of 25%.

TABLE 12. Whole Group Noticing ANY Breakdown

	Pre-test		Post-Test	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	5	2	8	0
Control	4	4	5	3

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group ($M=.375$, $SD=.518$) and the control group ($M=.125$, $SD=.354$); $t(14)=.139$, $p=.278$. There is not enough evidence to reject the null hypothesis ($H0: \mu d = 0$).

Table 13 shows that while students were likely to notice breakdowns more generally in the pre-test, they were more specifically noticing or naming the breakdowns as vocabulary breakdowns by the post-test. It demonstrates that half the students in the intervention group had some awareness of being confused before the intervention and that this awareness became more explicit after the intervention.

TABLE 13. Types of Errors Students in the Intervention Group Notice in Pre-Test and Post-Test

Student	Pre-Test	Post Test
IS1	No	General
IS2	General	Vocab
IS3	No	Vocab
IS4	General	Vocab
IS5	Vocab	Vocab
IS6	No	Vocab
IS7	General	Vocab
IS8	Vocab	Vocab

Table 14 includes students who read at a 4th-5th grade Lexile level. It shows that there was a 100% increase in students in the intervention group who noticed any type of breakdown, and a 50% increase in students in the control group who noticed any type of breakdown.

TABLE 14. 4th-5th Grade Lexile Level Noticing ANY Breakdown

	Pre-test		Post-Test	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	2	2	4	0
Control	2	1	3	0

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.5, SD=.577) and the control group (M=.333, SD=.577); t(5)=.361, p=.721. There is not enough evidence to reject the null hypothesis (H0: $\mu d = 0$).

Table 15 includes all students who read below a 4th grade Lexile level. It shows that there was a 33% increase in students in the intervention group who noticed any type of breakdown in the post-test as compared to the pre-test. It also shows that there was no change in the number of students in the control group who noticed any type of breakdown on the post-test as compared to the pre-test.

TABLE 15. Below 4th Grade Lexile Level Noticing ANY Breakdowns

	Pre-test		Post-Test	
Group	Noticed	Did Not Notice	Noticed	Did Not Notice
Intervention	3	2	4	0
Control	2	0	2	0

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.25, SD=.5) and the control group (M=0, SD=0); t(6)=.178, p=.356. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

The data shows that many of the students began the intervention with some ability to generally notice breakdowns and that the intervention pushed students to be more specific in the types of breakdown they were noticing.

Paraphrasing

Table 16 shows that there was a 25% improvement in the mean paraphrasing score for students in the intervention group and a slight decrease (10%) in the mean paraphrasing score for students in the control group.

TABLE 16. Mean Paraphrasing Score for ALL Students

	Pre-test	Post-Test
Group	Mean Score	Mean Score
Intervention	0.75	1
Control	1.25	1.125

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.25, SD=.463) and the control group (M=-.125, SD=.354); t(14)=.045, p=.09. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

Table 17 includes all students who read at a 4th-5th grade Lexile level. It shows that there was a 25% improvement in the mean paraphrasing score for students in the intervention group and a 20% decrease in the mean paraphrasing score for students in the control group.

TABLE 17. Mean Paraphrasing Score for Students Reading at a 4th-5th Grade Lexile Level

	Pre-Test	**Post-Test**
Group	Mean Score	Mean Score
Intervention	0.75	1
Control	1.67	1.33

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.25, SD=.5) and the control group (M=.333, SD=.577); t(5)=.105, p=.211. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

Table 18 includes all students who read below a 4th grade Lexile level. It shows that there was a 25% improvement in the mean paraphrasing score for students in the intervention group and no change in the mean paraphrasing score for students in the control group.

TABLE 18. Mean Paraphrasing Score for Students Reading Below a 4th Grade Lexile Level

	Pre-Test	**Post-Test**
Group	Mean Score	Mean Score
Intervention	0.75	1
Control	1	1

A paired-sample t-test was conducted to compare the mean change in the intervention group and the control group. There was not a significant difference in the change in scores for the intervention group (M=.25, SD=.5) and the control group (M=.0, SD=0); t(7)=.146, p=.292. There is not enough evidence to reject the null hypothesis (H0: μd = 0).

Table 19 includes students from the intervention group whose paraphrasing scores improved. It includes either direct quotes from pre/post-tests or a description of the paraphrasing. In two of the three examples, students moved from no attempt at paraphrasing to an attempt, though the paraphrasing remained inaccurate or unclear. The student who attempted paraphrasing in the pre-test attempted to paraphrase more of the text in the post-test, though again, the paraphrasing remained largely inaccurate or unclear.

TABLE 19. Paraphrasing Evidence Comparison for Intervention Students

Student	Pre-Test	Post-Test
IS1	"There's nothing confusing."; no attempt at paraphrasing.	"I just didn't understand what the heck was going on in the second sentence."; attempted to paraphrase first sentence.
IS4	"I don’t get this because it's talking about all these things that just confused me all the way."; no attempt at paraphrasing.	Noted two vocabulary breakdowns; attempted to paraphrase 2 sentences (out of four).
IS8	Noted one vocabulary breakdown; attempted to paraphrase first sentence.	Noted 5 vocabulary breakdowns; attempted to paraphrase each sentence.

The mean paraphrasing score remained low in both groups, and there was not a statistically significant difference between the change in scores in the intervention group as compared to the control group. However, the fact that there is some improvement after just three lessons does suggest that a minimally adaptive tutoring system that explicitly teaches comprehension monitoring may have potential to support comprehension. It also demonstrates the potential for such a system to increase student awareness of their vocabulary breakdowns.

4.3 Summary

The statistically significant results were mixed. Overall, there was not significant evidence to show that the intervention helped students notice a planted error. On the other hand, there was statistically significant evidence that showed the ITS increased student awareness of vocabulary breakdowns. Awareness of one's understanding is part of the metacognitive process. Therefore, it may be concluded that a minimally adaptive intelligent tutoring system has the potential to improve student comprehension (Jacobs & Paris, 1987; National Research Council,1998; Mirandola et. all, 2018;). Given that the improvement in student paraphrasing scores was not significant, and paraphrasing was used as a measure of comprehension, it can also be concluded that there is not yet enough evidence to determine the extent of this potential.

5 Discussion

Though many of the results lack statistical significance, the findings may still be important (Amrhein, Greenland, & McShane, 2019; Wasserstein & Lazar, 2016). For instance, the increased growth for students in the intervention group supports claims that explicit instruction can support comprehension monitoring (US Department of Education, 2010; Fielding & Pearson, 1994; Duke & Pearson, 2009; Grossman et al., 2013). That this instruction occurred through a minimally adaptive intelligent tutoring system confirms other claims that technology can be used to support student learning (Nesbit, Liu, Ma, & Adesope, 2014; Snow, Jacovina, & McNamara, 2016; Wijekumar & Meyer, 2006).

The lack of statistical significance in the students' ability to note the planted error (and overall lesser ability to note it as compared to their ability to notice vocabulary breakdowns) is interesting but not necessarily incongruent with other research. For instance, Baker (1979) noted a similar pattern in her study on metacognition in college students; subjects in this study missed a surprisingly large number of the planted errors. Baker showed that the errors were more likely missed when they were related to details rather than main idea. (Baker, 1979). In this current study, none of the planted errors related to the main idea of the passage. This could explain why students failed to note them as often as they failed to note other breakdowns in vocabulary; the other breakdowns made it difficult to understand the main idea while the planted errors simply related to details.

Another reason there was a statically significant difference in the intervention students' ability to note vocabulary breakdowns in general but not the planted error may be explained by the placement of the error (Winograd & Johnston, 1982). Every student in the intervention read below grade level, some significantly, and poor readers struggle, at least in part, because of issues with stamina (Hiebert, 2014). All of the students in the intervention noted at least one breakdown in vocabulary understanding that took place before the planted error. This suggests that, similar to other studies, the targeted intervention had an impact on their comprehension monitoring, but it did not increase reading stamina (Edmonds et al., 2009). This is unsurprising given that building reading stamina requires a significant time investment and this intervention was completed in a total of 45 minutes over two days (Children's Literacy Initiative, n.d.; Hiebert, 2014).

Additionally, the difference in the number of noted breakdowns and the location of these breakdowns between the intervention group and the control group is worth considering. In the intervention group, none of the students who acknowledged vocabulary breakdowns did so after the planted error even though all but one of them (88%) noted breakdowns in multiple sentences. This suggests that their failure to note the planted error may have had more to do with a lack of time to read far enough in the text rather than a lack of awareness. On the other hand, of the students in the control group who noticed a breakdown, only one student acknowledged more than one (20%). Of those who only acknowledged one breakdown, 75% noted it in the first sentence (three students) and 25% noted it after the planted error (one student). This suggests that, while the students in the intervention group were making an effort to monitor their comprehension for longer, the students in the control group may have quit quickly or may have engaged in fake reading (Heibert, 2014). This study did not specifically measure student motivation, but the intervention group's apparently greater commitment to comprehension monitoring on a difficult passage suggests that the design may have impacted it.

In their 1997 study, *How Motivation Affects Learning,* Vollmeyer, Rollett, and Rheinberg noted that two motivational factors were influential on the cognitive-motivational process: incompetence fear and confidence mastery. They found that students who had lower levels of anxiety about failing a task, and higher certainty that they would succeed used a more systematic strategy when faced with a challenging task (Vollmeyer, Rollett, & Rheinberg, 1997). The scaffolding in the design for this study may have influenced these motivational factors.

To understand this, it is important to first consider the students' probable experience both in this study and in learning up to this point. Research shows that students are likely to experience frustration when less than 93% of a reading task is easily accessible (Burns & Dean, 2005; Hattie, 2012). Given that all of the students in both the intervention and control group were at least three grade levels behind, one can conclude that these passages were at their frustration level. Research also indicates a likelihood that learning goals and experiences for these students have typically focused on comprehension strategies exclusive of comprehension monitoring. That is, strategies that focus on summarizing and predicting (Grossman et al., 2013; Magnusson, Roe, Blickstad-Balas, 2018). Given that these students were reading at frustration level, one may conclude that these learning goals were often unattainable, as they could not comprehend enough of the text to make accurate summaries or predictions (as evidenced by their overall low paraphrasing scores).

Because the learning goals up to this point had likely focused on comprehension rather than comprehension monitoring, the control students may have understood the success criteria for the pre/post-test to be similar. That is, the point of the test was to prove that they understood. Because the text was almost certainly at their frustration level, there was a reasonable chance these students had low-confidence and high-anxiety about their ability to successfully complete the task. As a result, they lacked the motivation to persist (as suggested by their data related to stamina).

The students in the intervention, on the other hand, may have had a different experience. The first comprehension monitoring strategy presented in this study was that *Good readers monitor their comprehension as they read*. The success criteria for this goal was to simply notice a breakdown and admit that they did not understand the text. The lesson then moved on to explicitly teach students what features of a text are likely to make it difficult, such as the presence of unfamiliar vocabulary. For all but one student, this was the extent of the intervention. Because success was not related to comprehending a frustration level text, students could be confident in their ability to attain this goal, thus leading to higher levels of motivation, as evidenced by their persistence on the post-test (Vollmeyer, Rollett & Rheinberg, 1997; Hattie, 2012).

Explicit instruction and scaffolding were not the only elements of instructional design that may have led to the difference in outcomes between the intervention group and the control group. Because the intervention also incorporated feedback and a responsive lesson, both of which have been shown to drive student learning, it can be concluded that these aspects of design also had an impact on the results (Hattie, 2012).

What is perhaps more interesting is that neither the feedback nor the responsiveness was extensive; indeed, the study incorporated minimal adaptability. The texts, for instance, were chosen ahead of time and in were no way responsive to student interest, reading level, or skill-level. That the students still demonstrated growth under these conditions is unexpected in light of research that argues for the importance of such conditions (Guthrie, Wigfield, & You, 2012; Sanden, 2014; Wigfield, Gladstone, Turci, 2016).

Furthermore, the bulk of the feedback was a formulaic response to students' answers on a Google Form. While the feedback overall represented a best guess as to the misconception the answer choice illuminated, it was largely generic and did not address the students' degree of proficiency towards the goal, nor did the feedback provide highly explicit guidance on how to make progress towards the goal. Even the slightly more targeted feedback available in the third lesson was similarly limited in range. Research, on the other hand, shows that the most impactful feedback is more nuanced and explicit than the feedback offered here, so the extent of the growth in the intervention group is notable in this regard (Dean, Hubbell, Pitler, & Stone, 2012; Hattie, 2012; Kegel & Bus, 2012).

This data suggests that some of the stated challenges in developing an intelligent tutoring system for comprehension monitoring, such as consistently finding just right texts or providing highly personalized feedback, may be addressed simply by limiting its adaptability (Heilmen et al., 2006; Jacovina & McNamara, 2017). The data may also emphasize the importance of effective instructional design and provide evidence that the approach used in this study achieves instructional clarity around metacognitive strategies, two components crucial to improvements in comprehension (Grossman et al., 2013; Hattie, 2012).

5.1 Limitations

This study employed convenience sampling and a small sample size. The homogenous nature and small sample size make it difficult to generalize the study to a larger, more diverse sample of students across the United States and the world (Bornstein, Jager, & Putnik, 2013; Tipton, Hallberg, Hedges, & Chan, 2016).

Comprehension is a difficult construct to understand, let alone assess, and as such, there are many issues of measurement in studies that aim to capture comprehension (Edmonds et al., 2009; Snow, 2003). Furthermore, the instrument used in this study was researcher-developed and unvalidated by other studies. Thus, the conclusions drawn from the data must be made with a level of uncertainty (Covacevich, 2014; Edmonds et al., 2009).

Time was a third limitation to this study. In order to minimize the disruption to the students, the intervention was run during the school's "reteach week" on the two consecutive days students did not have incentive trips or celebrations planned. As a result, the intervention only consisted of three lessons. This severely limited the amount of explicit instruction the students could engage with; moving too fast could hinder their ability to learn the new information and begin to build habits around the new skills (National Research Council, 2000; Lally, Van Jaarsveld, Potts, & Wardle, 2010). This may account for the lack of evidence that the intervention improved students' ability to comprehend the grade appropriate text. Due to the short duration, the student outcomes in the study were limited to building awareness in order to help students notice breakdowns in general and breakdowns with vocabulary in particular. Noticing breakdowns is just the first step of the metacognitive process. In order to improve comprehension, students must also learn to make logical connections to the text in order to solve for their breakdown.

6 Recommendations

There is enough evidence to warrant further investigation into the potential a minimally adaptive intelligent tutoring system has to support comprehension monitoring. Specifically, it may be worth further developing the student model beyond noticing one's comprehension to provide explicit modeling and deliberate practice for solving breakdowns as they occur. Additionally, because this study did not explore the research behind building an effective user interface, it is recommended that resources be allocated to this task before moving forward with product design.

It is important to bear in mind that comprehension monitoring is just one of the eight comprehension strategies that have been empirically proven to improve student comprehension. It has also been shown that multiple-strategy instruction has greater potential to impact comprehension than isolated strategy instruction and practice. (National Reading Panel, 2000). Thus, to be most effective, the skills learned in this ITS should become an integrated part of the literacy curriculum. To do that, product designers need to consider how to effectively support teachers' internalization of the strategies taught in this ITS.

One recommendation is to design a teacher-facing model to serve as a representation for the explicit modeling students experience with the ITS. The use of representations is a common practice in teacher education and can be used in this case to deepen teacher content knowledge (Danielson, Shaughnessy, & Jay, 2018). The fact that the lessons in this ITS are designed to be efficient (no longer than 15 minutes each) makes this a good option for teacher development, particularly in the United States where teacher time is severely limited (OECD, 2018).

Offering additional support through webinars, expert coaching or professional learning communities can also help teachers better integrate these comprehension monitoring strategies with other lessons in the literacy curriculum. This will lead to improved student outcomes. (Darling-Hammond, Hyler, & Gardner, 2017; Hattie, 2012). Again, as teacher time is limited, these additional supports should be similarly efficient. The approach should also incorporate best practices for effective teacher professional development as laid out in the 2017 Learning Policy

Institute report (Darling-Hammond, Hyler, & Gardner, 2017).

The hope is that this intelligent tutoring system can provide immediate student support while simultaneously investing in long-term teacher development. This two-pronged approach is necessary because, while technology can improve student outcomes, teacher quality remains the most important variable to student achievement (Hattie, 2012). It is through this combined approach that we can address the urgent literacy crisis facing our world today.

REFERENCES

Al-Faki, I. M., & Siddiek, A. G. (2013). The Role of Background Knowledge in Enhancing Reading Comprehension. *World Journal of English Language*, *3*(4), 42-66. https://doi.org/10.5430/wjel.v3n4p42

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior Knowledge Activation and the Comprehension of Compatible and Incompatible Text. *Reading Research Quarterly*, *20*(4), 420-436. https://doi.org/10.2307/747852

American Educational Research Association. (2011). *Code of Ethics.* Retrieved from https://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics(1).pdf

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305-307. doi:10.1038/d41586-019-00857-9.

Anderson, J.R. (1983). *The Architecture of Cognition.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1996). *Cognitive tutors: Lessons learned*. (ARI Research Note 96-66). Alexandria, VA: Publication of the U.S. Army Research Institute for the Behavioral and Social Sciences.

Association for Supervision and Curriculum Development (ASCD). (2005). *Literacy Strategies for Grades 4-12: Reinforcing the Threads of Reading.* Alexandria, Virginia: K. Tankersley. Retrieved from http://www.ascd.org/publications/books/104428/chapters/The-Struggling-Reader.aspx

Baker, L. (1979). Comprehension Monitoring: Identifying and Coping with Text Confusions. *Journal of Reading Behavior*, *XI*(4) 366-374. doi:10.1080/10862967909547342

Baker, L., & Brown, A. L. (1980). *Metacognitive skills and reading*. (Technical report no. 188). Urbana, Ill.: Center for the Study of Reading.

Baker, L., & Cerro, L. (2000). Assessing metacognition in children and adults. In G. Schraw & J.C. Impara (Eds.), *Issues in the Measurement of Metacognition*, (pp. 99–145). Lincoln, NE: Buros Institute of Mental Measurements.

Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Literacy Research*, *24(2)*, 143–172. https://doi.org/10.1080/10862969209547770

Beck, I. L., Perfetti, C. A., & McKeown, M. G. 1982. Effects of long-term vocabulary instruction on lexical access and reading comprehension.

Journal of Educational Psychology, *74(4)*, 506–521. https://doi.org/10.1037/0022-0663.74.4.506

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2(2)*, 131–156.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations. *Journal of Educational Psychology*, *(September),* 498–520.

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review*, *33(4),* 357–370. https://doi.org/10.1016/j.dr.2013.08.003

Bowen, R.S. (2017). *Understanding by Design*. [Vanderbilt Universty Center for Teaching]. Retreived from https://cft.vanderbuilt.edu/understanding-by-design/

Bryant, P., Cain, K., & Oakhill, J. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96(1),* 31-42. doi: 10.1037/0022-0663.96.1.31

Burns, M.K., & Dean, V.J. (2005). Effect of acquisition rates on off-task behavior with children identified as having learning disabilities. *Learning Disability Quarterly, 28(4),* 273-281. Retreived from https://files.eric.ed.gov/fulltext/EJ725679.pdf

Cairncross, S., & Mannion, M. (2001). Interactive multimedia and learning: Realizing the benefits. *Innovations in Education and Teaching International*, *38(2),* 156–164. https://doi.org/10.1080/14703290110035428

Carrell, Patricia L. (1983). Three components of background knowledge in reading comprehension. *Language Learning: A Journal of Research in Language Studies, 33(2)*, 183-203.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13(2)*, 145-182. https://doi.org/10.1207/s15516709cog1302_1

Children's Literacy Initiative. (n.d.). Important Factors for Improving your Students' Reading Motivation [web page]. Retrieved from https://cli.org/2017/06/14/important-factors-improving-students-reading-motivation/

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.) *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, (pp. 453-494). Hillsdale, NJ: Lawrence Earlbaum Associates.

Common Core Standards. (n.d.). English Language Arts Literacy in History / Social Studies, Science, and Technical Subjects Key Elements of the Standards: Appendix A Research Supporting Key Elements of the Standards Glossary of Terms. United States: Department of Education. Retreived from http://www.corestandards.org/assets/Appendix_A.pdf

Common Core Standards. (n.d.). English Language Arts Literacy in History / Social Studies, Science, and Technical Subjects Key Elements of the Standards: Standard 10: Range, Quality, & Complexity. United States: Department of Education. Retrieved from http://www.corestandards.org/ELA-Literacy/standard-10-range-quality-complexity/measuring-text-complexity-three-factors/

Coyle, D. (2009). *The Talent Code: Greatness Isn't Born. It's Grown. Here's How.* New York, New York: Bantam Dell.

Covacevich, C. (2014). How to select an instrument for assessing student learning. *Inter-American Development Bank: Education Division (SCL/EDU)*. (Technical Note, No. IDBN-TN-738). Retrieved from https://pdfs.semanticscholar.org/bd1d/760edd87c1717648ed0b502bbad7fff1cd0a.pdf

Cross, D., & Paris, S. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology, 80(2),* 131-142. http://doi.apa.org/get-doi.cfm?doi=10.1037/0022-0663.80.2.131

Danielson, K. A., Shaughnessy, M., & Jay, P.M. (2018). Use of representations in teacher education. In Pam Grossman (Ed.) *teaching core practices in teacher eduaction,* (pp. 15-55). Cambridge, MA: Harvard Education Press.

Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, *24.* https://doi.org/10.1016/S0896-6273(00)80855-7

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute.

Dean, C.B., Hubbell, E.R., Pitler, H., & Stone, (2012). *Classroom Instruction that Works* (2nd ed.). Denver, Co: McREL.

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension. *Journal of Education*, *189(1–2),* 107–122. https://doi.org/10.1177/0022057409189001-208

Dynamic support of contextual vocabulary acquisition for reading: DSCoVAR): An intelligent tutoring system. (n.d.) [project description]. Retrieved December 3, 2019 from https://dscovar.org#Demos

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A Synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, *79(1),* 262–300. https://doi.org/10.3102/0034654308325998

Ericcson, K. Anders, Krampe, R.T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100(3),* 363–406.

Fielding, L. G. & Pearson, P. D. (1994). Reading comprehension: What works. *Educational Leadership, 51(5),* 62-69. https://www.researchgate.net/publication/254343574

Fisher, D., Frey, N, Hattie, J. (2016). *Visible Learning for Literaacy: Implementing the PracticesThat Work Best to Accelerate Student Learning*. United Kingdom: Corwin: A SAGE Company.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34(10),* 906-911. https://jwilson.coe.uga.edu/EMAT7050/Students/Wilson/Flavell%20(1979).pdf

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school english language arts and teachers' value-added scores. *American Journal of Education*, *119*(3), 445–470. https://doi.org/10.1086/669901

Guthrie, J.T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In Christenson, S., Reschly, A., Wylie, C. (Eds.), *Handbook of Research on Student Engagement*, (pp. 601-634). New York, Ny: Springer.

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*, *45(1),* 34–45. https://doi.org/10.1016/j.intell.2013.04.001

Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact on Learning*. Abington, Oxon: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77(1),* 81–112. https://doi.org/10.3102/003465430298487

Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi, M. (2006). Classroom success of an intelligent tutoring system for lexical practice and reading comprehension. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2*, 829–832.

Hiebert, E. (2014). The forgotten proficiency: Stamina in silent reading. *Text Project Article Series.* Santa Cruz, CA: TextProject, Inc. Retrieved from http://www.textproject.org/assets/library/papers/Hiebert-2014-The-forgot-ten-reading-proficiency.pdf

Hiebert, E., Wilson, K., & Trainin, G. (2010). Are students really reading in independent reading contexts? An examination of comprehension-based silent reading rate. In Hiebert, E.H. & Reutzel, D.R. (Eds.) *Revisiting Silent Reading: New Directions for Teachers and Researchers*, (pp. 151–167). Santa Cruz, CA: TextProject, Inc. https://doi.org/10.1598/0833.09

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational Psychologist*, *22*, 255–278. https://doi.org/10.1080/00461520.1987.9653052

Jacovina, M. E., & Mcnamara, D. S. (2017). Intelligent tutoring systems for literacy: Existing technologies and continuing challenges. *Intelligent Tutoring Systems: Structure, Applications and Challenges, 153-174.* Nova Science Publishers, Inc. Retrieved from https://files.eric.ed.gov/fulltext/ED577131.pdf

Kegel, C. A. T., & Bus, A. G. (2012). Online tutoring as a pivotal quality of web-based early literacy programs. *Journal of Educational Psychology*, *104(1),* 182–192. https://doi.org/10.1037/a0025849

Kendig, C.E. (2016). What is *proof of concept* research and how does it generate epistemic and ethical categories for future scientific practice? *Science and Engineering Ethics, 22(3),* 735-753. https://doi.org/10.1007/s11948-015-9654-0

Killion, J. (2013). *Establishing Time for Professional Learning.* Oxford, OH: Learning Forward. Retrieved from https://www.learningforward.org

Kuhn, D., Dean, D. Jr. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into Practice*, 43(4), 268-273. https://doi.org/10.1207/s15430421tip4304_4

Lally, P., Van Jaarsveld, C.H.M., Potts, H.W.W., & Wardle, J. (2010). How habits are formed: Modelling habit formation in the real world. *Euorpean Journal of Social Psychology, 40*, 998-1009. https://doi.org/10.1002/ejsp.674

Lorch, Robert F. Jr., Lorch, Elizabeth P., & Klusewitz, M.A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology*, *85*(2), 239–252.

Macnamara, B., Hambrick, D.Z., & Oswalkd, F.L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychological Science Online,* 1-11. DOI: 10.1177//0956797614535810

Magnusson, C. G., Roe, A., & Blikstad-Balas, M. (2019). To what extent and how are reading comprehension strategies part of language arts instruction? A study of lower secondary classrooms. *Reading Research Quarterly*, *54*(2), 187–212. https://doi.org/10.1002/rrq.231

Markman, E. M. (1979). Realizing that you don't understand: elementary school children's awareness of inconsistencies. *Child Development*, 50(3), 643-655. https://doi.org/10.2307/1128929

Martinez, M.E. (2006). What is metacognition? *Phi Delta Kappan, 87(9),* 696-699. https://doi.org/10.1177/003172170608700916

McCarthy, K. S., Johnson, A. M., Likens, A. D., Martin, Z., & McNamara, D. S. (2017). Metacognitive prompt overdose: Positive and negative effects of prompts in iSTART. *Grantee Submission*, (pp. 404–405). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED577125&site=ehost-live

McNamara, D.S., O'Reilly, T., Rowe, M., Boonthum, Ch., Levinstein, I. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D.S. McNamara (Ed.) *Reading Comprehension Strategies: Theories, Interventions, and Technologies,* (pp. 397-420). New York and London: Lawrence Erlbaum Associates, Inc. Retrieved from https://books.google.com/books?id=Xj5HEdLGg-YC&lpg=PP1&pg=PR4#v=onepage&q=McNamara&f=false

Meyer, B. J. F., Wijekumar, K. K., & Lin, Y. C. (2011). Individualizing a web-based structure strategy intervention for fifth Graders' comprehension of nonfiction. *Journal of Educational Psychology*, *103*(1), 140–168. https://doi.org/10.1037/a0021606

Mirandola, C., Ciriello, A., Gigli, M., & Cornoldi, C. (2018). Metacognitive monitoring of text comprehension: An investigation on postdictive judgments in typically developing children and children with reading comprehension difficulties. *Frontiers in Psychology*, *9*, 1-9. https://doi.org/10.3389/fpsyg.2018.02253

Mokhtari, K., Dimitrov, D. M., & Reichard, C. A. (2018). Revising the metacognitive awareness of reading strategies inventory (MARSI) and testing for factorial invariance. *Studies in Second Language Learning and Teaching*, *8*(2 Special Issue), 219–246. https://doi.org/10.14746/ssllt.2018.8.2.3

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies, *Journal of Educational Psychology 94(2),* 249-259. https://doi.org/10.1037/0022-0663.94.2.249

National Scientific Council on the Developing Child. (2005/2014). Excessive Stress Disrupts the Architecture of the Developing Brain: Working Paper 3. Updated Edition. http://www.developingchild.harvard.edu

National Reading Panel. (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction-Reports of the Subgroups*, *1-449.* United States: National Reading Panel.

National Research Council. (1998). *Preventing Reading Difficulties in Young Children,* (pp. 16-446). US Department of Education office of Educational Research and Improvement. https://files.eric.ed.gov/fulltext/ED416465.pd

National Research Council. (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition.* Washington, DC: The National Academies Press. https://doi.org/10.17226/9853

National Science Foundation. (2014). Program Solicitation: *Accelerating innovation research-technology translation.* Directorate for engineering, industrial innovation and partnerships. NSF. https://www.nsf.gov/pubs/2014/nsf14569/nsf14569.htm

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of Text Difficulty: Testing their Predictive Value for Grade Levels and Student Performance*. The Gates Foundation. Retrieved from http://www.ccsso.org/Documents/2012/Measures ofText Difficulty_final.2012.pdf

Nesbit, J.C., Liu, Q., Ma, W., & Adesope, O. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106(4),* 901-918. http://dx.doi.org/10.1037/a0037123.

OECD. (2018). Indicator D4: How much time do teachers spend teaching? in *Education at a Glance 2018: OECD Indicators,* OECD Publishing, Paris. https://doi.org/10.1787/eag-2018-31-en

Ozturk, N. (2017). Assessing metacognition: Theory and practices. *International Journal of Assessment Tools in Education*, *4(2),* 134–148. https://doi.org/10.21449/ijate.298299

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*(1), 16–20. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1988-24805-001&site=ehost-live

Renaissance Learning. (2019). *What Kids Are Reading: World's Largest Annual Study of K-12 Reading Habits: 2019 Edition*. Wisconsin Rapids, WI: Renaissanec Learning, Inc. Retrieved from www.renaissance.com

Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of direct/ explicit teaching. *Reading and Writing Quarterly*, *25*(2–3), 125–138. https://doi.org/10.1080/10573560802683523

Sanden, S. (2014). Text complexity: A new way forward or two steps back? *English Journal,* 25(2), 7-14. Retrieved from http://www.ncte.org/library/NCTEFiles/Resources/Journals/TP/0252-may2014/TP0252Text.pdf

Sani, S.M. & Aris, T.N.M. (2014). *Modeling Student/Tutor Components of Intelligent Tutoring Systems: A Review,* 98-102. Paper. Proc. of the Intl.

Conf. on Advances in Computer Science and Electronics Engineering. 3.8.2013. Kuala Lumpur, Malaysia.

Schoenbach, R., Greenleaf, C., & Murphy, L. (2012). *Reading for understanding: How Reading Apprenticeship Improves Learning in Secondary and College Classrooms.* San Francisco, CA: WestEd.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113-125. Retrieved from http://wiki.biologyscholars.org/@api/deki/files/87/=schraw1998-meta.pdf

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*(4), 351–371. https://doi.org/10.1007/BF02212307

Sedlmeier, P. (2001). Intelligent tutoring systems. *International Encyclopedia of the Social & Behavioral Sciences*, 7674-7679. https://doi.org/10.1016/B0-08-043076-7/01618-1

Snow, C.E. (2003). Assessment of reading comprehension. In A.P. Sweet, & C.E. Snow (Eds.) *Rethinking reading comprehension*, (pp. 192-206). New York: Guilford.

Snow, E.L., Jacovina, M.E., Tanner, Jackson, G.T., & McNamara, D.S. (2016). iSTART-2: A reading comprehension and strategy instruction tutor. In S. A. Crossley & D.S. McNamara (Eds.) *Adaptive Educational Technologies for Literacy Instruction* (pp. 104-121). New York, New York: Taylor & Francis.

Snyder, T. & Musu-Gillette, L. (2015). Free or reduced lunch: a proxy for poverty? *National Center for Education Statistics (NCES) Blog.* Retrieved December 3, 2019 from https://nces.ed.gov/blogs/nces/post/free-or-reduced-price-lunch-a-proxy-for-poverty

Sönmez, Y., & Sulak, S. E. (2018). The effect of the thinking-aloud strategy on the reading comprehension skills of 4th grade primary school students. *Universal Journal of Educational Research*, *6*(1), 168–172. https://doi.org/10.13189/ujer.2018.060116

RAND Corporation. *Teachers Matter: Understanding Teachers' Impact on Student Achievement*. Santa Monica, CA: RAND Corporation, 2012. https://www.rand.org/pubs/corporate_pubs/CP693z1-2012-09.html.

Sahdra, B. & Thagard, P. (2003). Procedural knowledge in molecular biology. *Philosophical Psychology* 16(4), 477-498.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science 25,* 203-244.

The Education Commission of the States. (2005). *Prisoners of Time.* The Education Commission of the States Education Reform Reprint Series. (Reprinted from 1994 Report of the National Education Commission on Time and Learning.) Retrieved from https://www.ecs.org/clearing-house/64/52/6452.pdf

Tipton, E., Hallberg, K., Hedges, L.V., & Chan, W. (2016). Implications of small samples for generalization: Adjustments and rules of thumb. *Journal of Research on Educational Effectiveness, 41(5), 472-505.*

UNESCO. (2005). Education for all: literacy for life; EFA global monitoring report 2006. Paris, France: Global Education Monitoring Report Team.

U.S. Army Research Laboratory. (2015). *Design Recommendations for Intelligent tutoring Systems, Volume 3: Authoring Tools and Expert Modeling Techniques.* Orlando, Florida: R. Sottilare, A. Graesser, X. Hu, & K. Brawner (Eds.).

U.S. Department of Education (2010). *A review of the current research on comprehension instruction.* The National Reading Technical Assistance Center, RMC Research Corporation. Washington, D.C.: S. Butler, K. Urrutia, A. Buenger, & M. Hunt. Retreived from https://www2.ed.gov/programs/readingfirst/support/compfinal.pdf

Vollmeyer, R., Rollett, W., & Rheinberg, F. (1997). How motivation affects learning. In M.G. Shaffo & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society,* (pp. 796-801). Hillsdale, New Jersey: Erlbaum.

Wade, S. (1990). Using think alouds to assess comprehension. *Reading Teacher*, *43*(7), 442–451.

Wasserstein, R.L. & Lazar, N.A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133, https//:doi.org/10.1080/00031305.2016.1154108.

Wigfield, A., Gladstone, J., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child Development Perspective, 10(3),* 190-195. doi:10.1111/cdep.12184

Wiggens, G., & McTighe, J. (2012). *The Understanding by Design Guide to Creating High-Quality Units*. Alexandria, VA: Association for Supervision & Curriculum Develepent.

Wijekumar, K., & Meyer, B.J.F. (2006). Design and pilot of a web-based intelligent tutoring system to improve reading comprehension in middle school students. *International Journal of Technology in Teaching and Learning,* 2(1), 36-49.

Winograd, P., & Johnston, P. (1982). Comprehension monitoring and the error detection paradigm. *Journal of Reading Behavior,* 14(1), 61-76. https://journals.sagepub.com/doi/pdf/10.1080/10862968209547435

World Literacy Foundation. *The Economic & Social Cost of Illiteracy.* (2018). United Kingdom: World Literacy Summit.

APPENDICES

Appendix 1. Metacognitive Toolbox 1(2)

This resource was initially created by the researcher in 2016 for use in an instructional guide for Mastery Charter Schools in Philadelphia, Pennsylvania in the United States. It has been adapted for this study.

Goal: Students are aware of when they understand the text and when they stop understanding. When they have a breakdown, they can ask a specific question about it, and use a toolbox of strategies to solve the breakdown to make more meaning. Students realize that these tools simply ask them to use what they do know in order to figure out what they are confused by. Over time, students get better at metacognition and comprehension of longer and longer texts.

Figure 1. Metacognitive Process

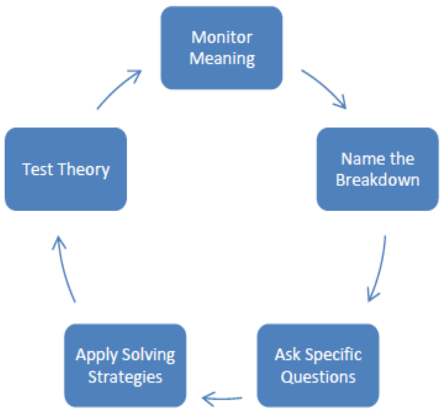

Table 1: Metacognitive Toolbox-Building Awareness

Elements of Metacognition	Goals: Students…
Awareness: • Monitor Meaning • Name Breakdowns • Ask Specific Questions	1. Think about the meaning of the texts they are reading and pay attention to what their brain is thinking. 2. Articulate every chunk of text they do "get". 3. Articulate specific questions about what they "don't get." 4. Identify the breakdown at the root of their question.
Building Awareness	
Goals:	1. Students think about the meaning of the texts they are reading and pay attention to what their brain is thinking. 2. Students slow down their brains enough to - see that they DO understand some things. - ask questions about the things they don't understand - identify the breakdown at the root of their question - pay attention to EVERY part of the passage.
Metacognitive Toolbox	1. As you read, pay attention to what you are thinking. This means you: • keep track of the things you "Get". You know you "Get" it because you can put it in your own words. • realize if you're not able to put it in your own words. When that happens, finish the thought to see if you can. If you still can't, ask a specific question about what you "Don't Get", use the question to identify what type of breakdown is preventing you from understanding (i.e. vocabulary, syntax, text length)