

DIGITAL HIGHLIGHTING AND ANNOTATING AS FORMATIVE ASSESSMENT:
PATHS TO TECHNOLOGY-ENABLED READING COMPREHENSION

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Jonathan D. Cohen, B.A., M.A.

May, 2014

© Copyright by
Jonathan D. Cohen
All Rights Reserved
May, 2014

EXECUTIVE SUMMARY

Glen L. Bull

Reading comprehension is an essential skill for students to develop in order to become productive members of society. It is a skill that can be learned, and a number of interventions and instruction types have been developed over the years to help students do so. One limitation of many of these interventions is that they require students to be interrupted during the reading process, whether to participate in the intervention itself, or so that the instructor/researcher can deliver a formative assessment. Such interruptions into the reading process can have a deleterious impact on the comprehension it is meant to address.

The current research reports on the results of an investigation into the relationship between the types and patterns of annotations students make in a digital text and the students' comprehension of the text. To capture the students' annotations, a digital e-reader was developed that allows students the ability to both highlight and leave textual comments on the text. These annotations are minimally intrusive activities, unlike many other reading comprehension interventions.

A sample of seventh grade students, $N = 250$, read a short story using the e-reader, then took a posttest measuring their comprehension of the story. An neural network analysis was employed to examine the nature of the relationships between the highlights

and annotations and comprehension of the text. Results indicated that a pattern is evident, particularly when the type of highlights are examined. Further, data suggested that the presence of certain types of inferences are more closely associated with stronger comprehension.

DEDICATION

To Cecily and Nathan, Lynne and Mark Cohen, and Karen and Dana Czapanskiy, whose love and support made this experience possible—my deepest, sincerest, and most loving gratitude to you all.

ACKNOWLEDGEMENTS

A project such as this is the culmination of years of effort—work that was necessarily extensive, but also necessarily rewarding, due to the number of people who have given generously of their time and experience to support it.

I wish to begin by thanking the members of my dissertation committee. Bill Ferster nurtured the growth of this project, from a half-formed concept to a fully realized implementation. This project could not have happened without his patient explanations and insightful suggestions. I am similarly indebted to Mike McKenna, whose expertise in both reading comprehension and instructional technology were hugely influential in my thinking. Finally, I am appreciative of the invaluable methodological support Patrick Meyer provided throughout the planning and analysis stages.

I also wish to use this space to thank my doctoral advisor and dissertation committee chair, Glen Bull, for his mentorship over the past four years. I value Glen's wisdom and experience tremendously. I particularly value his concept of a graduate education as a complete academic experience—when I think of my learning at Curry, I picture the lab conference table first and foremost, before I picture classrooms in Ruffner Hall. I will always be grateful for Glen's support of my growth as a student and as a professional.

This section would not be complete without special mention of Laura Smolkin. I am deeply grateful for the quality of support Laura has given me over the past years. Consistently, our weekly meetings were the highlight of my week, and her influence over my thinking about this topic in general and this study in particular is unmistakable. Quite simply, without her consistent encouragement, support, and feedback, this project would not have been possible.

And, finally, I want to acknowledge my earliest and most influential mentor, Gary Laws, Director of the Royster Middle School at Norfolk Academy. I have met countless numbers of educators since first beginning work with Gary in 1998, and I have read and experienced the thinking of even more than that, but the wisdom Gary shared with me upon first getting an interactive whiteboard installed in my classroom remains, simply, the best I've ever received: "Remember, it's not about the things, it's about the stuff. And it's not even really about the stuff. It's about the kids."

TABLE OF CONTENTS

EXECUTIVE SUMMARY	vi
DEDICATION	viii
ACKNOWLEDGEMENTS	ix
TABLE OF CONTENTS	xi
LIST OF TABLES	xiv
LIST OF FIGURES	xv
CHAPTER 1 INTRODUCTION	1
Comprehension Instruction	5
Technology and Reading Comprehension	10
The Study	11
CHAPTER 2 LITERATURE REVIEW	16
Reading Comprehension Defined	17
Construction-Integration Model of Reading Comprehension.....	19
Inferences	21
Activities During Reading	22
The Relationship Between Reading Activities and Working Memory, Inferencing, and Comprehension	23
Annotations and Highlighting During Reading	27
Summary of Activities During Reading Literature.....	29
Reading Comprehension and Technology in Secondary Schools	29
Digital Comprehension Supports	30
E-readers in the Classroom	33
Assessment of Comprehension of Specific Texts.....	35
Multiple Choice Assessments	37
Oral Retellings	40
Cloze/Maze	40
Sentence Verification Technique	42
Summary of Assessment Literature	43
Interactive Digital Design Principles	46
Neural Networks	48
CHAPTER 3 METHODOLOGY	50
Research Design.....	50
Methods.....	51

Tool Development	51
Researcher experience.	52
Pilot testing.	52
Tool design.....	54
Test Development	55
Multiple choice test development.	56
Sentence verification technique.	60
Pilot testing.	61
Site Selection and Participants.....	61
Site Selection.	61
Participants.....	63
Text Selection	64
Accessibility.....	64
Ability to engage.....	66
Ability to generate multiple inferences.....	67
Intervention Procedure.....	68
Fidelity of Implementation	70
Access, Role, and Ethics	70
Data Collection and Analysis.....	71
Collection.....	71
Analysis – neural network.	72
Analysis – multiple regression.....	73
CHAPTER 4 RESULTS OF THE STUDY	74
Posttest Development and Refinement	74
Data Collection Results.....	76
Log Data.....	76
Descriptive Statistics.....	77
Data Exploration	80
Model Training Phase	80
Model Testing Phase	84
Regression Analysis.....	86
Summary	87
CHAPTER 5 DISCUSSION	90
Review of the Research Questions	91
Research Question 1	91
Research Questions 2 and 3	92
Research Question 4	93
Additional Finding – Elaborative Inferences and Predictions	93

Limitations	96
Implications.....	98
Suggestions for Future Research	100
Summary	102
REFERENCES	104
APPENDIX A: TABLES AND FIGURES	118
APPENDIX B: TEST BLUEPRINT	126
APPENDIX C: ITEM WRITING GUIDELINES	127
APPENDIX D: INFERENTIAL ANALYSIS OF “SLOWER THAN THE REST”	128

LIST OF TABLES

TABLE	Page
1. Typical Lexiles reader measures, by grade	125
2. Best Net Search Summary	83
3. Comparison of Posttest Scores Between Students Who Highlighted Multiple Instances of Each Type of Inference	85
4. Predictors of Comprehension Posttest Scores.....	87
5. Comparison of Posttest Scores Between Students Who Predominantly Made Elaborative Inferences and Those Who Did Not	96

LIST OF FIGURES

	FIGURE	Page
1.	The ICON interface.....	119
2.	RAND Reading Study Group comprehension paradigm.....	120
3.	E-reader.....	121
4.	Passage dependency.....	59
5.	Sample SVT assessment	122
6.	Self-reported ethnicities of participants	123
7.	Frequency of highlights per page.....	124

CHAPTER 1

INTRODUCTION

One constant among the many changes in American education is the critical and continuing importance of teaching reading in our schools. Various and interweaving pressures, including the vast amount media students are exposed to and consume on a daily basis, have resulted in higher literacy demands on students than at any other time in recent history (Dalton & Strangman, 2006; International Reading Association, 2012). Simply put, participation in modern society demands not only fluent readers, but also skilled readers, so it is vital that students in schools learn “the language of books” (Oakhill & Cain, 2003, p. 53). In fact, students’ ability to comprehend text is of central importance to their ability to obtain an education (National Institute of Child Health and Human Development [NICHD], 2000).

International data reveals that American students perform better than average on tests of reading skills. On the Progress in International Reading Literacy Study (PIRLS) assessment, administered most recently in 2011 to fourth graders from 49 participating countries, American students had higher-than-average achievement scores, and placed among the ten highest scoring participating countries (Thompson et al., 2012). Scores on the Program for International Student Assessment (PISA) test, which is given to fifteen-year-old students in 33 countries, indicated that American students’ scores were not

significantly different than the international average (Fleischman, Hopstock, Pelczar, & Shelley, 2010). So, while the news is good that American students are, at worst, average readers when compared with the rest of the world, the fact remains that literacy skills in American students still fall short of what many believe is required by modern society (Reardon, Valentino, & Shores, 2012).

National testing data reflect a lack of broad scale progress over the past decades in supporting students' development of reading comprehension (McNamara, O'Reilly, Best, & Ozuru, 2006). The issues appear as early as fourth grade, where just 35% of students tested at or above the proficient level, leaving 65% of students at the basic or below basic level, on the 2013 National Assessment of Educational Progress (NAEP; National Center for Education Statistics (NCES), 2013). Though this does represent the highest level of students reading at the proficient level since 1992, it is only a 2% improvement since 2002. There is a degree of urgency here, as research has shown that a lack of success in reading comprehension in early grades leads to a persistent and increasing gap between struggling readers and non-struggling readers (Wharton-McDonald & Swiger, 2008). The presence of this "Matthew Effect" (Stanovich, 1986)—its name inspired by the biblical Gospel of Matthew, which states that to those who have, more will be given, and from those who have little or nothing, all will be taken—is evident. Following the trend from prior test administrations, reading comprehension scores on the 2011 NAEP plateau at the same level for eighth graders as fourth graders. Predictably, the issues with reading

comprehension persist into high school and college: Biancarosa and Snow (2004) reported that as many as 70% of secondary students require some form of reading remediation, and Kamil (2003) found that many students are unable to read the types of challenging texts that lead to academic success in high school.

It is significant that the NAEP assessment happens during the fourth-grade year. Prior to fourth grade, the focus of reading and literacy instruction is teaching students the mechanics of reading—essentially the process of decoding and fluency. Around the fourth grade, though, reading instruction has traditionally flipped from learning to read to reading to learn (Chall & Jacobs, 2003; Durkin, 1979), though the Common Core State Standards aim to engage early elementary students in the process of reading to learn (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010a). Reading becomes less about fluency and more about extracting and constructing meaning from texts in service of content learning in the various disciplines. Inherently, this type of “reading to learn” is more challenging for students. One of the main factors influencing comprehension is prior knowledge (Willingham, 2006). The more prior knowledge a reader has, the better the chance of that reader comprehending the text he or she is reading. However, much of the post-“learning to read” texts students are given to enable students to acquire new information. By definition, these are topics for which the reader has limited prior knowledge. Therefore, the reading is inherently more challenging. Further increasing the challenge level

involved in this type of reading is the presence of often unfamiliar vocabulary (Wharton-McDonald & Swiger, 2008), linked to the unknown concepts.

Perhaps it is not surprising, then, that the decline in attitudes toward reading continues in the middle grades. In a survey of 4,491 American students spread through 23 states and the District of Columbia, McKenna, Conradi, Lawrence, Jang, and Meyer (2012) found that attitudes toward academic digital, recreational digital, and recreational print texts declined from grade 6 to grade 8. Though both McKenna et al. and Petscher (2010) point out that the body of research on reading attitudes in middle school is problematic—attitude studies tend to focus on elementary grades, few established measures have been widely adopted for use with middle schoolers, and other measures are of low reliability—the broad trends all point to a reduction of positive affect towards reading around the time students begin to “read to learn.”

The decline in positive attitude toward reading is significant because research shows that reading attitude is a moderator or mediator of reading achievement (McKenna, Kear, & Ellsworth, 1995; Petscher, 2010; Sainsbury & Schagen, 2004). Petscher (2010) conducted a meta-analysis of the relationship between student attitudes toward reading and reading achievement, including 32 studies, containing an aggregate sample size of 224,615 and 118 effect sizes. He found a moderate relationship ($Z_r = .32$) for the entire corpus, though the relationship was weaker for middle school students ($Z_r = .24$). (Petscher’s meta-analysis did not specifically attribute the possible cause of the

weaker relationship between attitude and achievement in middle school students to any factors.) So, while attitude toward reading does not correlate directly with student achievement, it is a factor that impacts the slide in achievement that corresponds with the middle school years.

Comprehension Instruction

Given the unsatisfactory progress made in improving students' reading comprehension over the years, it is fair to ask whether it is even possible for schools and teachers to increase students' reading comprehension? According to Duke and Pearson (2002):

The answer is a resounding yes. A large volume of work indicates that we can help students acquire the strategies and processes used by good readers—and that this improves their overall comprehension of text, both the texts used to teach the strategies and texts they read on their own in the future. (p. 206)

However, it has not always been the case that teachers have taught reading comprehension in class. Following up on an assumption made in a National Institute of Education request for proposal that reading comprehension was being taught in schools, Durkin (1979) and two assistants spent 4,469 minutes observing reading instruction in 24 different fourth grade classrooms. Of these 4,469 minutes of reading class time they observed, they found only 28 minutes of comprehension instruction. The bulk of the time was spent instead on activities such as assigning work, checking assigned work, and providing information. The better comprehending students were able to use class time to

practice their comprehension skills, while the poorer comprehenders did not receive any of the additional comprehension instruction we now know would be beneficial to their achievement. Ness (2009) replicated the Durkin study in secondary classrooms, and found a similar result: Over 2,400 minutes of classroom observations yielded just 82 minutes of reading comprehension instruction, or, approximately 3% of the instructional time.

In the past two decades, though, reading comprehension instruction has become a more prominent part of instruction in the upper elementary and middle grades, though perhaps not to the extent that the research suggests is necessary to impact students' achievement levels (Wharton-McDonald & Swiger, 2008). The focus of the most recent push has been comprehension strategy instruction. Strategy instruction “centers on the direct teaching of specific procedures, such as summarizing, making inferences, and generating questions, and using them in working with text” (McKeown, Beck, & Blake, 2009, p. 218). It depends on students becoming aware of their own cognitive processes during the reading process, and selecting and employing the most effective strategy to aid in their comprehension.

The most prominent influence on the implementation of strategy instruction in classrooms is the report of the National Reading Panel (NRP) (NICHHD, 2000). Convened at the request of the United States Congress, the NRP was charged with “assess[ing] the status of research-based knowledge, including the effectiveness of

various approaches to teaching children to read” (NICHD, 2000, p. 1). In response, the NRP conducted a review of the literature, and found that “The past 2 decades of research appear to support the enthusiastic advocacy of instruction of reading strategies” (NICHD, 2000, pp. 4–46). The NRP synthesized existing research into 16 categories of instruction, from which they identified seven as having enough support in the literature to warrant advocacy: comprehension monitoring, cooperative learning, graphic and semantic organizers including story maps, question answering, question generation, and summarization. Pearson and Hiebert (2010) described this report as “amazingly influential in shaping policy and practice at both the federal level... and the state level” (p. 287).

An example of the type of strategy instruction advocated by the NRP is reciprocal teaching. Developed by Palincsar and Brown (1984), reciprocal teaching involves students guiding each other through some of the featured strategies of the NRP, including collaborative learning, predicting, clarifying, question generation/answering, and summarization. Teachers begin the process by modeling the use of specific techniques, then, gradually, students in small groups take responsibility over the implementation of the various techniques, using roles assigned to them by the teacher. It is, at its core, a student-centered pedagogy that relies on the interactions between learners as the main engine of comprehension development. Rosenshine and Meister (1994) conducted a meta-analysis of the effectiveness of reciprocal teaching, and found effect sizes of .32

(when measured by standardized measures of comprehension) and .88 (when using experimenter-developed measures) for 16 studies of reciprocal teaching, indicating its general effectiveness.

There is an argument to be made, though, that the amount of strategy instruction in today's classrooms is an overcorrection from the dearth of instruction reported by Durkin (1979). Two meta-analyses (Rosenshine, Meister, & Chapman, 1996; Rosenshine & Meister, 1994) of two comprehension strategies, reciprocal teaching and question generation, found moderate effect sizes for the strategy use, which align with expected results. However, Rosenshine and his colleagues also found that in each case, extended dosages of strategy instruction had no effect on results. For example, in the question generation meta-analysis (Rosenshine et al., 1996), time of implementations ranged from two hours to 12 hours, but no pattern related to duration of implementation was discovered.

McKeown, Beck, and Blake (2009) reported on an alternative to strategy instruction as a primary mode of reading comprehension instruction: a content approach. Content approaches are based specifically on text-processing models of reading, such as Kintsch's construction-integration model (e.g., Graesser, Singer, & Trabasso, 1994; Kintsch, 1998, 2004). In broad strokes, these models view comprehension as an automated process of identifying new textual information and integrating it with prior knowledge gained either from the text itself or from the reader's own experiences. Some

scholars (e.g., Dole, Duffy, Roehler, & Pearson, 1991; Gersten, Fuchs, Williams, & Baker, 2001) suggest that focusing students' attention on deriving meaning from specific texts rather than on selecting and deploying specific protocols for comprehending text is a more effective way to support reading comprehension.

An example of a content approach is questioning the author (QtA) (Beck, McKeown, Sandora, Kucan, & Worthy, 1996). QtA involves having students consider the meaning of a text through queries designed to encourage students to consider and grapple with what an author is trying to say. QtA is “a deceptively simple approach with a minimum of apparatus” (Beck et al., 1996, p. 387), and it is based in the idea that inviting students to explore the mechanisms of a text helps to develop not only comprehension of the particular text, but also of comprehension skills in general.

Like reciprocal teaching, QtA relies on in-class collaboration to support students' comprehension. Educators and researchers have, quite correctly, recognized the potential of collaboration to support comprehension, but less attention has been paid to ways to support comprehension while students are working independently. While anecdotal evidence seems to indicate that students are doing less academic reading independently for homework, the traditional model of teachers assigning out-of-class reading to be later discussed and debriefed in class still exists in secondary education. In order to explore ways to provide the support that some students need while reading independently, researchers are increasingly turning to digital technologies as a potential solution.

Technology and Reading Comprehension

The general body of evidence on using technology to improve students' reading comprehension suggests that technology, generally speaking, tends to have a small-to-moderate, positive effect on students' comprehension (e.g., Cheung & Slavin, 2012; Kamil & Chou, 2008; Moran, Ferdig, Pearson, Wardrop, & Blomeyer, 2008), about which more is discussed in the literature review, below. Many of these technology-based interventions are conceived as digital strategy coaches that attempt to replicate the interactions that students might expect in a collaborative classroom situation (Dalton & Strangman, 2006). Others are supported text reading environments, in which texts are altered or supplemented in an effort to lower or eliminate barriers to access and increase readers' ability to extract meaning (Anderson-Inman & Horney, 1997, 2007; Dalton, Proctor, Uccelli, Mo, & Snow, 2011; Dalton & Proctor, 2008; Dalton & Strangman, 2006).

Like their classroom counterparts, many digital reading comprehension interventions can be intrusive to the reading process. Take, for example, the Improving Comprehension Online (ICON) system (Dalton et al., 2011). The ICON system is "a universally designed Web-based scaffolded text environment designed to improve fifth-grade monolingual English and bilingual students' reading achievement" (Dalton et al., 2011, p. 68). Its design was guided by the principles of Universal Design for Learning (UDL), which "considers diverse learners from the earliest stages of design, with the

expectation that the inclusion of a fuller array of learning supports will better serve all learners, including those with identified needs” (Dalton et al., 2011, p. 71). In an effort both to satisfy the dictates of the UDL framework and to provide adequate support for all levels of readers, ICON includes a number of elements, including bilingual text-to-speech read-aloud functionality, bilingual translation of instructional supports, digital reading coaches, an editable response journal, a multimedia glossary, highlighting of unclear referents, and illustrative graphics (see Figure 1). Readers are intended to access the supporting material as needed to increase their levels of comprehension. It can be argued, though, that systems like ICON function as a Swiss Army knife, when all that certain readers may need is a simple pair of tweezers. As discussed in the literature review, below, such intrusions into the reading process can strain readers’ working memory capacities to the point that comprehension suffers.

The Study

Research has consistently shown that teachers can improve students’ reading comprehension skills through instruction, whether it be strategy instruction or content-based approaches. Much of the methods highlighted by the research as effective is collaborative in nature, which, in most instructional contexts, suggests in-class comprehension-related activities. However, reading as it is typically practiced in schools remains a solitary behavior. Students are assigned reading to do independently, with the

instruction and analysis happening in a classroom context. It is this time when students are reading independently that is the impetus behind this study.

When students read independently, they are inherently unable to access the benefits that come from the type of collaborative learning that has proven to be effective in improving reading comprehension. Technology offers an opportunity to bring elements of collaborative, face-to-face learning to students working in isolation. It is feasible to create a technological tool that provides real-time analysis and targeted comprehension feedback to students during the reading process that mirrors the type of interactions that happen in collaborative, face-to-face environments. However, to accomplish this the digital interface needs to receive some type of input from the reader that indicates his/her levels of comprehension during the reading process.

Some digital reading comprehension supports solicit this input in the form of multiple choice questions asked a various times during the reading process (c.f., Dalton & Strangman, 2006; e.g., Ray & Belden, 2010). These are problematic for a two reasons. First, asking readers to stop reading and to move into a different cognitive task—i.e., problem solving—can place undue strain on the working memories of struggling readers (Andreassen & Bråten, 2009). Moreover, multiple questions are needed to obtain the type of granularity necessary to provide adequate, personalized support to each reader, and, of course, the more interruptions readers have to deal with during reading, the harder it is for them to form coherent mental comprehension models. Research also shows that

readers read differently when they know that objective, multiple-choice style questions will follow their reading (Rupp, Ferne, & Choi, 2006).

The second reason why embedded multiple choice questions are problematic as a way of assessing comprehension during the reading process is that multiple choice questions introduce distractors into the reading process. Unskilled readers could potentially conflate the distractors with the mental model of the text in formation.

There is a need, then, as we work to further support individuals' reading for a way to assess comprehension during the reading process that does not interrupt the reading process to the point that comprehension suffers. To this end, I conceptualized an e-reader that provides students with the ability to read and annotate a text online. The annotations consist of three different kinds of textual highlights and reader-generated comments tied to portions of the text of their choosing. The three highlight types reflect different types of inferences readers make while reading. Inferencing "is the process of connecting information within the text or within the text and one's knowledge base, and drawing a conclusion that is not explicitly stated in the text" (McNamara & Kendeou, 2011, p. 35). Each of the three highlight colors corresponds to a different type of inference: elaborative inferences in which the reader connects the text to his or her own knowledge or experience, "A-ha!" inferences in which the reader arrives at a moment of insight into the meaning of the text, and predictions in which the reader anticipates later occurrences in the text. Further description of these inferences can be found in chapter 2.

These highlights, as well as any comments associated with the highlights, are recorded by the system.

The present study takes the annotations of 250 seventh-grade students from a wide variety of general comprehension levels and explores the nature of the relationship between their types and patterns of highlights and annotations and their comprehension of the text they read, as measured by a posttest. The relationship is analyzed by a neural network, a digital system modeled on physical neural networks found in the brain, which is used to uncover latent patterns found in complex data sets, such as this one. This research serves as a first step in a larger investigation of whether a computer can be trained to monitor and ultimately assist a reader's comprehension during the reading process.

Specifically, the research will answer the following main question: What is the nature of the relationship between the types and patterns of students' highlighting and their comprehension of the specific text? Additionally, the following sub-questions will be addressed:

- What is the nature of the relationship between the types of inferences students make, and their level of comprehension as measured by both a multiple choice assessment and a sentence verification technique assessment?
- What is the nature of the relationship between the density and frequency of highlighted inferences and students' comprehension of the text?

- What is the nature of the relationship between time spent on the reading task and students' comprehension?
- What kind of patterns emerge from an exploration of the textual annotations of students of varying reading levels?

CHAPTER 2

LITERATURE REVIEW

The goal of this study is to examine the nature of the correlation between students' annotations of a short story, as captured by a Web-based e-reader, and their comprehension of the text. The study was informed by a review of the literature, reported below, of two key domains: reading comprehension and technology. Specifically, I begin this review with a description of both a theoretical framework of reading comprehension, the Construction-Integration (C-I) model, and the centrality of inferencing to comprehension. Then, related literature is explored, including:

- Activities during reading – Asking students to perform an activity during reading will necessarily alter the process of reading. Cognitive load theory (Chandler & Sweller, 1991; Sweller, Van Merriënboer, & Paas, 1998) offers a window to understanding the effect of activities during reading. Highlighting during reading is specifically explored.
- Reading comprehension assessment – Essential to this study is determining what type of assessment can capture a valid and reliable estimate of the students' comprehension of the specific text chosen for this study. Four assessment types are considered in particular, including multiple choice

questioning assessment, cloze/maze assessment, oral retelling, and sentence verification technique.

- Technology reading comprehension interventions with secondary students – The growing body of literature concerning the effects of technology on reading comprehension in secondary students interventions is explored.
- As this study necessitated the development of an interactive digital reading environment, a review of design principles and heuristics for the development of effective and usable digital-human interfaces is included.
- Finally, this study takes advantage of neural network analysis software to assist in the data analysis. Included in this review of literature is a description of the function and utility of digital neural networks.

Reading Comprehension Defined

Reading comprehension is a complex concept, and the varied definitions of it reflect that complexity (e.g., Paris & Hamilton, 2008). Many of the definitions feature a widely agreed-upon component of reading comprehension, which is the construction of meaning from the skeletal blueprint of the text. For example, Calfee (2009) defines comprehension as “the strategic reconstruction of a text toward a particular purpose” (p. xiii). The National Assessment of Educational Progress (NAEP) framework defines reading comprehension as “an active and complex process that involves understanding

written text, developing and interpreting meaning, [and] using meaning as appropriate to type of text, purpose, and situation” (2013a, p. 2). The Rand Reading Group (2002) definition of reading comprehension follows a similar line. It defined reading comprehension as involving two parallel and simultaneous processes: extracting meaning from a text and constructing meaning from a text.

The Rand definition is notable, in part, for its elaboration on the choice of the words “extracting” and “constructing.” The report explains that those words were chosen “to emphasize both the importance and the insufficiency of the text as a determinant of reading comprehension” (2002, p. 11). They point to the presence of “the reader who is doing the comprehending, the text that is to be comprehended, and the activity in which comprehension is a part” as each equal contributors to the process of comprehension. Moreover, they emphasize that reading occurs in a sociocultural context that affects the nature of comprehension (see Figure 2).

In essence, reading comprehension is built upon, but not limited to, the reader’s ability to decode the words on the page. Unskilled readers often struggle with comprehension because all of their cognitive energy is dedicated to the process of decoding, leaving few cognitive resources available for meaning making. As decoding becomes more automated, as it is in skilled readers, then more cognitive resources can be made available for meaning making (Paris & Hamilton, 2008).

This is not to say, though, that perfect fluency leads to perfect comprehension. Consider, for instance, the following passage, taken from the philosopher G. W. F. Hegel's *Phenomenology of Spirit* (1977): "But the formative activity has not only this positive significance that in the pure being-for-self of the servile consciousness acquires an existence; it also has, in contrast with its first moment, the negative significance of *fear*" (p. 118). A skilled reader would likely understand what each word in the passage means individually. However, it is also reasonable to assume that many highly skilled readers could read that passage and have a difficult time forming a mental representation of the meaning of the statement. The reader may have insufficient prior knowledge of Hegelian philosophy, or philosophy in general, to assist in meaning making. In addition, the passage includes references to antecedents not located within the passage. For a variety of reasons, decoding is not adequate to meaning making for most readers in this particular situation. Research bears out the assertion that decoding is a necessary component of comprehension, but not sufficient on its own for comprehension to be achieved (Paris & Hamilton, 2008; Yuill & Oakhill, 1991).

Construction-Integration Model of Reading Comprehension

One theory of reading comprehension particularly relevant here is the Construction-Integration (C-I) Model (Kintsch, 1998, 2004). The C-I model states that "readers construct simultaneously a model of the literal text and an elaborated model of the situation implied by the text" (Paris & Hamilton, 2008, p. 35). The theory describes a

textbase and a *situation model* of a text. The textbase is the literal content of the text. It is dependent on the reader's ability to construct meaning from both the microstructure, or sentence-level local properties, and macrostructure, or global, properties of the text. The microstructure is made up of textual units called *propositions*, which are relational units of nouns, predicates, and modifiers. Readers construct a mental representation of the textbase through assembly of the proposition into networks of nodes.

The *situation model* refers to the construction of the text that consists of the integration of the micro- and macrostructure textbases with the reader's prior knowledge and experience. The textbase and the situation model operate in parallel, so that it is possible for a reader's situation model to differ from the textbase, based on the reader's experience.

Through re-reading and mental revision, these two levels of understanding inform and reinforce each other. The bottom-up portion of this model, the decoding of the text, is (ideally) largely automated, but the top-down portion, the construction of the situation model, is a more conscious process. The ability to construct the situation model is the result of a reader's ability to make inferences both to elements of the textbase as well as to elements of the reader's own knowledge. The integration portion of the model describes how readers reconcile the various inferences with the continuing input from the text. It is clear, then, that according to the C-I model, inferencing is key to the comprehension process.

Inferences

A central component of the meaning-making involved in comprehension occurs through the reader's drawing of inferences. Inferencing "is the process of connecting information within the text or within the text and one's knowledge base, and drawing a conclusion that is not explicitly stated in the text" (McNamara & Kendeou, 2011, p. 35). The centrality of inferencing to reading comprehension is hard to overstate. Willingham (2006) described the necessity for inferencing in comprehension:

The writer cannot specify every last detail or the text would become impossibly long. The writer must make assumptions about what the reader knows. If the level of knowledge that the writer assumes does not match the level of knowledge that the reader actually has, the reader won't comprehend the text. (p. 50)

The writer makes assumptions about what the reader knows, but the reader must instantiate those associations, or else comprehension will not occur. Research consistently confirms this claim (e.g., Duke & Pearson, 2002; Graesser, Singer, & Trabasso, 1994; McNamara & Kendeou, 2011; McNamara & Magliano, 2009; Nation, Cocksey, Taylor, & Bishop, 2010; Oakhill & Cain, 2003; RAND Reading Study Group, 2002; Urquhart, 2002).

Many taxonomies of inference have been proposed (e.g., Gordon & Pearson, 1983), but most follow Oakhill and Cain (2003) in dividing inferences into two categories: coherence and elaborative. Coherence inferences—sometimes called on-line inferences (Graesser et al., 1994), local inferences (Urquhart, 2002) or bridging

inferences (McNamara & Kendeou, 2011)—are strictly necessary to constructing meaning from a text because they establish cohesion between sentences. Elaborative inferences, also known as global inferences (Urquhart, 2002), are, as their name suggests, connections that link the surface level of the text to the broader world. While elaborative inferences are not strictly necessary to understanding text, they do facilitate readers' abilities to fill in the blanks of the text, and to make rich connections between text and experience (Graesser et al., 1994).

In addition to using inferences to draw logical/mechanical connections between sentences and to fill in blanks that, if explicitly enumerated, would make most texts tedious to the point of unreadability, skilled readers use inferences to make predictions (Trabasso & Magliano, 1996 as cited in Lomicka, 1998), reveal textual inconsistencies (Urquhart, 2002), assess motivations, infer actors' goals, and determine global themes and story morals (Graesser et al., 1994). These kind of almost-subconscious analytic acts undertaken by skilled readers stand in contrast to the predominantly superficial reading performed by unskilled readers, whose comprehension is limited by their lack of inferencing (Nation et al., 2010).

Activities During Reading

In an intervention that involves adding a secondary activity to the reading process, it is important to ensure that the introduction of the activity does not significantly detract from the reader's ability to comprehend the text. To understand one process by which

activities undertaken during reading can disrupt comprehension, it is necessary to explore the relationship between working memory, inferencing, and comprehension. This review will begin with a section that explores that relationship, then it will review the literature on the process of annotating and highlighting in reading.

The Relationship Between Reading Activities and Working Memory, Inferencing, and Comprehension

The integration part of Kintsch's C-I Model (Kintsch 1998; Kintsch 2004), of which inferencing is an essential component, takes place in readers' working memory (Andreassen & Bråten, 2009; Zumbach & Mohraz, 2008). Working memory refers to the cognitive resources which people use as they process information during tasks that involve some kind of cognition (Sweller et al., 1998). In reading, working memory "serves as a buffer for the most recently read propositions in a text, enabling their integration to establish coherence, and holds information retrieved from long-term memory to facilitate its integration with the currently active text" (Cain, Oakhill, & Bryant, 2004, p. 31). Or, put differently, working memory is necessary both for the construction of a coherent textbase and for the formation of a situation model via integration of that textbase with the readers' prior knowledge. It would stand to reason, then, that there is a relationship between working memory capacity and reading comprehension ability, and the research bears that out. Daneman and Merikle (1996) performed a meta-analysis, including 77 studies and 6,179 participants, which found that

measures of reading span, a typical measure of working memory, are reliable predictors of comprehension, $r = .41$, confirming findings from earlier empirical studies (e.g., Daneman & Carpenter, 1980, 1983).

Much of the subsequent research has focused on further clarifying the nature of the relationship between working memory and comprehension. Swanson (2003) compared learning disabled and skilled readers' working memory performance across four age groups, ranging from age 7 to age 20, focusing on the phonological, visual-spatial, and semantic aspects of working memory. The study confirmed earlier research that associated comprehension difficulties with working memory deficits, and further indicated that the deficits emerge across both verbal and visual-spatial tasks. This relationship between working memory and comprehension has been observed at students of varying age groups. Engle, Corullo, and Collins (1991) gave first-, third-, and sixth-grade students ($N = 120$) word and reading span tasks and correlated the results of those to a following-directions task, which involved comprehension. They found an increase in the predictive value of the span tasks as the students aged. This study included, crucially, students on both sides of the fourth- and fifth-grade level, at which the major component of comprehension in normally developing readers transitions from decoding and fluency to inferencing and meaning-making (Wharton-McDonald & Swiger, 2008). Further, Seigneuric and Elrich (2005) conducted an examination of working memory and comprehension in a three-year longitudinal study of seven- through ten-year-olds. The

data they presented suggests that both vocabulary and working memory may contribute more to reading comprehension variance as the age of the reader increases.

In summary, adequate working memory capacity is necessary for inference making, and given that inferencing is central to comprehension, working memory capacity is therefore highly correlated with children's reading comprehension levels. It is important to note, though, that working memory is not the sole determinant of reading comprehension levels, but rather one of a number of determinants, including factors such as the extent of students' exposure to vocabulary and their prior knowledge of the content, to name just two (Paris & Hamilton, 2008).

The cognitive processes along which readers construct representations from texts happen in two stages: the online construction, which happens in real time as readers proceed through the text, and offline representations, which readers construct following the reading process (van den Broek & Kendeou, 2011). The online process is a complex one which involves creating inferences between the text the reader is currently engaged with and prior knowledge gained from both the text and the reader's prior experience. This process is dependent on available working memory to proceed (van den Broek & Kendeou, 2011; van den Broek, 2012).

Any additional strain placed on the reader's working memory should necessarily have a deleterious effect on the ultimate representation the reader creates:

the offline representation and the online processes are causally related: The processes that unfold during moment-by-moment reading comprehension provide the basis for the construction of the offline text representation. If the online processes fail, so does the final text representation. (van den Broek & Kendeou, 2011, p. 260)

Therefore, any intrusion into the reading process with some kind of secondary activity must be made with sensitivity toward the potential working memory strain placed on the reader's cognition.

For example, research into the effects of hypertextual reading environments on working memory have found decreases in reading comprehension as compared to comprehension levels found in linear texts (DeStefano & LeFevre, 2007). Zumbach and Mohraz (2008) asked a sample of 60 university students to read a narrative text presented in either a linear or a non-linear, hypertextual environment. Through delivery of a multiple-choice test of content, an essay prompt also based on content, and a pair of measures of cognitive load, they found that the students in the non-linear condition demonstrated significantly reduced scores on tests of content knowledge and significantly higher levels of cognitive load than the students in the linear narrative condition. Other studies that presented such hypertextual intrusions into the reading of elementary school students produced similar results, particularly in those who had low amounts of prior knowledge (e.g., Paolucci, 1998; Shin, Schallert, & Savenye, 1994).

Not all secondary activities affect cognitive load to the extent that comprehension suffers, however. In fact, research suggests that some secondary actions performed during

reading actually enhance comprehension. For example, studies have demonstrated that questioning techniques performed during reading can enhance students' metacognition, thereby increasing their comprehension levels (e.g., McKeown et al., 2009; van den Broek, Tzeng, Risdén, & Trabasso, 2001; Yuill & Oakhill, 1988). van den Broek et al. (2001) found that questioning during reading is effective for both seventh grade and college students, but not for fourth graders, potentially due to the fourth graders' smaller working memory capacities or lesser vocabulary knowledge.

Annotations and Highlighting During Reading

Of particular relevance to this study is whether actively reading a text, though annotating and/or highlighting a text, produces such a strain on readers' working memory that it becomes counterproductive. Piolat, Oliva, and Kellogg (2005) reviewed the literature on note-taking and working memory, and concluded that note-taking is a "high resource-consuming activity" (p. 297), though they later conclude that the notion of note-taking as a strain on working memory is particularly relevant to the process of taking notes while listening to a lecture. Of the demands on working memory while taking notes on text during reading, they conclude that working memory capacity mediates readers' comprehension, because readers cannot "delay or slow their writing time too much in order to still be able to maintain in working memory the intermediate representations that result from comprehension" (p. 297). However, a number of empirical studies (e.g., Lahtinen, Lonka, & Lindblom-Ylänne, 1997; Slotte & Lonka, 1999) indicate that note-

taking during reading can improve students' performance on comprehension measures. A meta-analysis (Hebert, Simpson, & Graham, 2012) that examined note-taking and comprehension as part of a larger study of the impact of writing on reading comprehension found that the results were inconclusive, due to a wide range of both positive and negative reported effect sizes. It should be noted, though, that this research focuses on non-fiction, informational texts. The idea, therefore, that annotating text is detrimental to the comprehension process is still an open question, particularly with regard to fiction.

Research confirms that highlighting and underlining, which for the purpose of this review will be treated synonymously¹, are indeed common behaviors among students of all ages, including middle school students (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). The majority of studies on highlighting focus on whether the act of highlighting can improve students' retention of factual information or their comprehension of texts (Dunlosky et al., 2013). While most studies of highlighting ultimately show that it has no real comprehension benefit beyond simply reading without highlighting or underlining, the research does not show that it impairs comprehension, either (Dunlosky et al., 2013). Highlighting is therefore considered to be a minimal intrusion into the reading process.

¹ In treating highlighting and underlining synonymously, this review follows the lead of Dunlosky et al. (2013).

Summary of Activities During Reading Literature

While the introduction of any non-automatic activity to the reading process increases the demand placed on readers' active memory, the research shows that this added demand does not necessarily lead to diminished comprehension. In particular, the process of highlighting does not seem to interfere with comprehension, though research investigating the relationship between highlighting and working memory would be beneficial to the field.

Reading Comprehension and Technology in Secondary Schools

Though digital technology has existed for decades, it is only recently that a significant amount of scholarship on the effects of technology on reading comprehension has emerged (Kamil & Chou, 2008). The National Reading Panel (NICHHD, 2000) was launched with the aim of producing a meta-analysis of the impact of technology on reading, however a paucity of studies meeting its quality guidelines prevented the panel from completing the meta-analysis (Moran et al., 2008). The last decade, though, has seen enough research that some clarity is emerging on the topic of technology and reading comprehension.

Two recent meta-analyses, Cheung and Slavin (2012) and Moran et al. (2008), found similar overall effect sizes on technology in reading achievement. Cheung and Slavin (2012) conducted a meta-analysis which included 84 studies covering over 60,000 participants in grades K-12. They found a small overall effect size ($ES = 0.16$) for the use

of technology over traditional instructional methods across all grades. When limited to secondary grades, however, the studies presented a larger, positive effect size, $ES = 0.31$. Moran et al. (2008) focused on middle school interventions in their meta-analysis, which covered 20 articles including 89 effect sizes. Like Cheung and Slavin, they also found a moderate effect size for the impact of technology on reading comprehension, $ES = 0.49$. Both of these studies, though, point out that their results need to be interpreted cautiously, given the small number of qualifying studies of technology and reading achievement, their typically small sample sizes, and the preponderance of quasi-experimental designs. In addition, as Moran et al. (2008) point out, within the corpus of studies on technology and reading, there is a “narrow focus on cognitive outcomes (comprehension)” (p. 28). The research is consistent, though, in its findings that technology-based interventions can increase students’ reading comprehension.

Digital Comprehension Supports

Many of the research studies focusing on technology’s impact on reading comprehension are clustered around the topics of vocabulary and comprehension strategies/metacognitive activities (Kamil & Chou, 2008). As bandwidth and computing power continues to increase, researchers are developing and testing more complex and media-heavy reading comprehension intervention software. These types of software support students’ ability to extract meaning from texts, often by lowering or circumventing the barriers to entry into texts often present for less skilled readers. An

example of this type of digital scaffold is the Improving Comprehension Online (ICON) program (Dalton et al., 2011), which its authors describe as “a universally designed web-based scaffolded text environment” (Dalton et al., 2011, p. 68). Texts included in ICON are supplemented with illustrations, text-to-speech capabilities, anaphoric highlighting (i.e., highlighting referents of references), and an on-demand glossary, with the aim of helping students to extract meaning from the text.

Many of the tools that have been developed to strengthen students' comprehension abilities take the form of strategy coaches (Dalton & Strangman, 2006). One of the earliest such technology-based tools that researchers used to improve students reading comprehension skills was the Reading Partner (Salomon, Globerson, & Guterman, 1989). Along with a core text, the Reading Partner offered metacognitive elements, such as self-guiding questions, self-monitoring questions, and specific strategies for enhancing comprehension, like engaging in mental imagery and inferring meaning from titles. Participants were encouraged to write responses to the metacognitive elements. The researchers found that the 25 seventh-grade students who used the Reading Partner showed significant improvement in their reading comprehension and quality of written analysis than students who did not receive Reading Partner, though the researchers admitted that their computer tool was “relatively primitive” (p. 626), leaving room to speculate on the extent to which the presence of added metacognitive instruction, rather than the digital delivery system, precipitated the gains reported.

Another, more recent tool that focuses specifically on inferencing is Interactive Strategy Trainer for Active Reading and Thinking (iSTART; McNamara et al., 2006). As its name states, iSTART is a strategy trainer that aims to teach students five comprehension strategies—one of which is making bridging inferences—that can be applied broadly to different reading situations, though iSTART focuses primarily on science texts. The strategies are used to influence self-explanation, which “involves having a reader type or say aloud what a sentence or portion of a text means to a reader” (McNamara et al., 2006, p. 149). The program itself contains three components: Introduction, Demonstration, and Practice. In Introduction, students watch an animation of a teacher-agent introducing self-explanation to student-agents, and then in Demonstration, a student-agent presents a self-explanation and answers questions from the teacher-agent about how strategies influenced the explanation. Finally, in Practice students create self-explanations and the digital teacher-agent offers feedback on the self-explanations.

McNamara, O'Reilly, Best, and Ozuru (2006) conducted a study comparing students who used iSTART with students who learned self-explanation strategies without a digital tool. Results indicated that the rising eighth- and ninth-grade students who used iSTART benefitted more from the strategy training than students without iSTART. In particular, the tool helped high-strategy knowledge students generate the inferences necessary to improve their ability to answer bridging inference questions. Low-strategy

knowledge students only showed improvement on text-based questions, for which the answers required little-to-no inference-making. However, at present, descriptions of iSTART do not provide detail on how it delivers bridging inference strategy instruction.

What unites ICON, Reading Partner, and iSTART is that these digital products aim to enhance the reading process by introducing a variety of supports, from surface-level lexical support to deeper metacognitive training. Though each study is far from conclusive—in no case was it clear what part of the variance could be attributed to the digital system, versus the increased comprehension instruction each enabled—all showed some positive gains in one or more aspects of comprehension.

E-readers in the Classroom

Though not a robustly developed line of research, of particular interest to this study is the body of research concerning the use of electronic readers, also called e-books or e-readers. E-readers are either pieces of software or hardware which have the primary function of presenting text to be read (Mealer, Morgan, & Williams, 2011). They differ from a Web browser in that texts on e-readers tend to have fewer, if any, hyperlinks and reduced multimedia elements, though it should be noted that this type of uncluttered interface does not apply to e-readers marketed towards preschool readers. E-reader software includes commercial products such as Adobe's Acrobat reader, and e-reader hardware include such dedicated devices as Amazon's Kindle and the Barnes & Noble

Nook. In addition, tablet computers like the Apple iPad and Google's Nexus line are increasingly used as e-readers (Mealer et al., 2011).

Researchers have explored the question of whether comprehension is helped or hindered by reading texts on e-readers. In the infancy of e-readers, early studies tended to explore the usability and readability of digital texts (Margolin & Driscoll, 2013). Many of these studies found that readers read faster and more accurately using paper texts, but that the discrepancy between the two media could be attributed to the

physical novelties and constraints (e.g., backlighting and flickering of electronic text, differences in font and spacing across media, angle of observance and scrolling of electronic text as compared to page turning of traditional text) inherent in the use of—what was then—an emerging technology. (Margolin & Driscoll, 2013, p. 513)

Later studies turned their attention away from reading speed and accuracy and more toward comprehension.

Recent studies have examined the more direct relationship between e-readers with their printed analogues. Green, Perera, Dance, and Myers (2010) studied the difference in undergraduates' textbase-level recall of an informational text read digitally in one condition and on paper in another. They found no significant difference in recall. Similarly, studies on undergraduate populations have found that there is no significant difference in comprehension between readers using e-readers and print texts (Connell, Bayliss, & Farmer, 2012; Kang, Wang, & Lin, 2009; Margolin & Driscoll, 2013; Schugar, Schugar, & Penny, 2011). The same pattern followed for younger students.

Grimshaw, Dungworth, McKnight, and Morris (2007) found no difference in comprehension between e-readers and paper in upper elementary/middle school students.

Assessment of Comprehension of Specific Texts

Reading researchers have developed and studied a multitude of reading comprehension interventions and tested students in a number of ways to determine the latent constructs that impact comprehension skills. However, measuring reading comprehension is not a straight-forward proposition. Though an argument can be made that assessing reading comprehension is actually impossible—Smith (2004) wrote, “Comprehension cannot be measured at all, despite constant educational efforts to do so, because it is not a quantity of anything” (p. 43)—a more reasonable contention would be that, due to its intangible nature, measuring comprehension is difficult (Leslie & Caldwell, 2009; Snow, 2003).

For this reason, likely, researchers often use well validated, widely used tests to measure reading comprehension when possible. However, for a variety of reasons, such measures are not always practical to administer in research settings. One such reason is that these general measures of reading comprehension are not designed, by their very nature, to be sensitive to short-term interventions (Moran et al., 2008). Another reason why such measures may be inappropriate is that they can be arduous to administer, both for the teachers who have to accommodate them and for the students who have to take them. Finally, and of critical importance here, general reading comprehension measures

do not exist to measure comprehension of specific texts beyond the samples used in the tests. Researchers sometimes decide, then, to design their own assessments, including subjectively scored writing assessments, error identification protocols, and researcher-developed multiple choice tests.

There is research to suggest, though, that these researcher-developed assessments tend to produce larger effect sizes than general reading comprehension assessments (Willingham, 2006). For example, in a meta-analysis of reciprocal teaching interventions, Rosenshine and Meister (1994) found an effect size of 0.88 over 10 studies for experimenter-designed assessments, which includes $ES = 0.85$ over 5 studies for summarization tests, and $ES = 1.00$ over 7 studies for short-answer tests. These were all larger than the effect size of 0.32 over 9 studies found for studies that used assessments not specifically developed for the particular interventions. The same pattern repeated in a later study by the same authors on questioning strategies: Experimenter-developed measures resulted in an effect size of 0.86, while scores generated from general measures of reading comprehension resulted in an effect size of 0.36 (Rosenhine et al., 1996). Edmonds et al. (2009) conducted a meta-analysis of reading interventions for older struggling readers and found an effect size of 1.19 for researcher-developed measures and 0.47 for non-experimenter developed assessments.

Since assessments tend to be the bottom-line determinants of the efficacy of reading interventions—one only need to consider the influence the National Reading

Panel report (NICHHD, 2000) on reading instruction as evidence of this point—it is crucial that researchers understand how the choices made regarding assessment type can impact their results. The following sections review four assessment types, multiple choice, oral retellings, cloze/maze, and sentence verification technique. These four assessments are among the most commonly used in research situations which call for objective measurement of comprehension of specific texts.

Multiple Choice Assessments

Multiple choice assessment rose to prominence in the so-called Boom Period resulting from the U.S. military's need to expand rapidly in the wake of World War I (Thorndike & Thorndike-Christ, 2010). They are now arguably the most common method of assessing reading comprehension by both educators and researchers (Andreassen & Bråten, 2009; Magliano, Millis, Ozuru, & McNamara, 2007). Multiple choice assessments consist of two parts: a *stem*, which presents a question or a statement to be completed, and a list of potential answers, usually numbering between three and five (Thorndike & Thorndike-Christ, 2010). The incorrect answers are referred to as *foils* or *distractors*.

McKenna and Stahl (2009) identify three types of questions that can be asked to target different levels of comprehension. The most superficial questions are *literal* questions. These questions ask students to recall individual components of the text, typically consisting of specific facts. These questions assess students' fluency, recall,

and, to a certain extent, their ability to form a mental representation of the text. Answers to literal questions can be found explicitly in the text.

Conversely, *inferential* questions are questions that have answers not explicitly mentioned in the text. Rather, answers to inferential questions must be generated by the reader through the process of connecting elements of the text to other parts of the text or to the reader's own knowledge base. Inferential questioning is not tantamount to asking the reader for his or her opinion, though; answers to inferential questions must be grounded in the text.

The third type of question identified by McKenna and Stahl (2009) is a *critical* question. Critical questions "call upon students to form value judgments" (p. 161) about a particular text, such as judgments about a text's aesthetic worth, bias, or persuasiveness. Unlike literal or inferential questions, critical questions are ill suited for multiple choice assessments, due to their inherently subjective nature.

Magliano, Mills, Ozuru, and McNamara (2007) further subdivide literal questions into two varieties: *local-textbase* and *global-textbase* questions. Local-textbase questions tend to focus on recall of explicitly stated information contained in an area no larger than a sentence or two. Global-textbase questions differ from local-textbase questions in that global-textbase questions increase the target area of a question to larger chunks of the text. These questions require readers to "search and locate the appropriate segments of

the texts, construct a summarization of that segment, and then assess which answer option best matches that summarization” (Magliano et al., 2007, p. 118).

The assumption made in Magliano, Mills, Ozuru, and McNamara’s (2007) distinction between local- and global-textbase questions is that readers will have access to the text upon which they are being assessed during the assessment. Such is the case with widely used, standardized assessments of reading comprehension ability, such as the SAT or the Gates-MacGinitie Reading Tests. However, educators and researchers have numerous considerations to take into account when deciding whether to allow students access to the text during assessment. About tests which do make texts available during assessment, Snow (2003) wrote:

The questions can often be answered without reference to the passage, or by simply finding key words in the passage and identifying matches with one of the multiple-choice responses. Test items of this type do not come close to replicating what goes on when real comprehension occurs, nor to reflecting the full array of skills and capacities we would agree fall under the rubric of comprehension. (p. 193)

In contrast, when readers know they will not have access to the text during the assessment, more onus is placed on the reader to establish a coherent mental representation of the text during the reading process. However, this onus introduces the factor of students’ memory resources to the assessment equation, which can function to complicate the construct validity of the measure (Andreassen & Bråten, 2009).

Oral Retellings

An oral retelling is a type of comprehension assessment that requires the student to recount what happened in a previously read text. It is based on Kintsch's (Kintsch & Dijk, 1978; Kintsch, 1998) conception of a *proposition* as a building block of text, with a proposition being defined as a combination of a predicate and one or more associated arguments (i.e., an object, agent, or goal). The student's performance is gauged by the percentage of propositions the student recalls from a list of propositions assembled from the text by the test administrator, and the general coherence of the student's retelling. This assessment technique has the advantage of being more naturalistic than questioning techniques and the flexible, open-ended nature of retelling can result in a richer picture of the student's comprehension (Leslie & Caldwell, 2009).

However, there are significant drawbacks to the technique as well. Oral retellings are inherently dependent on the student's oral expression abilities, and a deficiency in this area can confound the results of the assessment (Leslie & Caldwell, 2009; McKenna & Stahl, 2009). Further, students' memory abilities are even more closely tied to the outcome of a retelling assessment than other forms of assessment. Finally, oral retellings are difficult to score and time-consuming to administer (Leslie & Caldwell, 2009).

Cloze/Maze

The original cloze procedure (Taylor, 1953) is an assessment in which words are deleted from a text at a regular interval, often every fifth word, regardless of part of

speech or textual context. The testee fills in each blank, and his or her score is calculated by counting the number of exact matches made. Modern cloze assessments are less stringent than the originals; synonyms can serve as correct answers, and only every fifth content words are deleted, preserving functional words (Pearson & Hamm, 2005; Stahl, 2009).

A number of studies have correlated achievement on cloze assessments with other measures of reading comprehension (e.g., Rankin & Culhane, 1967). However, the construct validity of cloze has come into question (Gellert & Elbro, 2012; Pearson & Hamm, 2005). In particular, research seems to indicate that the cloze procedure is effective in assessing comprehension at a sentence level, but less so in assessing more global-level comprehension (Pearson & Hamm, 2005; Stahl, 2009). Shanahan, Kamil, and Tobin (1982) conducted experiments that presented one group of undergraduate students with a traditional cloze assessment and another with the same cloze assessment, but with the sentence order scrambled. No differences due to sentence order was found, lending credence to the idea that cloze is inadequate as a measure of intersentential comprehension.

A variant of the cloze procedure that replaces the fill-in-the-blanks with multiple choices is the maze task (Guthrie, Seifert, Burnham, & Caplan, 1974). Maze tasks have shown sensitivity to reading comprehension in novice readers, and they are attractive to

some researchers due to the ease of scoring involved (McKenna & Stahl, 2008; Stahl, 2009).

Sentence Verification Technique

Sentence verification technique (SVT; Royer, Hastings, & Hook, 1979) is grounded in the idea that comprehension involves the creation of a memory representation of a text that preserves the meaning of the text, but not necessarily the precise language. In SVT, students read a passage and are then presented with a series of four-sentence clusters tied to the passage. Within the four-sentence cluster, one sentence is a verbatim reproduction of a sentence from the text, and another is a paraphrase of the original sentence. A third sentence is a “meaning change” (Royer et al., 1979, p. 356) sentence, in which one word in the original sentence is altered such that the meaning of the sentence is changed fundamentally. The fourth kind of sentence is a distractor, which is consistent with the general meaning of the passage, but is unrelated to any of the sentences in a passage. Students then mark each sentence with *yes* or *no*, depending on whether the sentence reflects the meaning of the passage.

SVT is usually used on short passages of about 12 sentences (Royer, 2001), though SVT has been used with passages of approximately 1,000 words (Royer, Sinatra, Greene, & Tirre, 1989). Students from third grade through undergraduates have been tested with SVT, and the results have been found to be valid and reliable (Kardash, Royer, & Greene, 1988; Marcotte & Hintze, 2009; Royer, Greene, & Sinatra, 1987;

Royer, 2001). Royer (2001) found that SVT assessments that used four passages with 64 test sentences produced reliability scores (Cronbach's alpha) of .70 to .80 and tests that include six passages and 96 test sentences have Cronbach's alphas in the .80 to .90 ranges. It has demonstrated strong correlations with other measures of reading comprehension (Leslie & Caldwell, 2009; Marcotte & Hintze, 2009; Royer, 2001). Despite the positive aspects of SVT—it is proven effective with a broad range of readers and a broad range of texts, and it is easy to create and administer—it is not widely used as a measure of comprehension of specific texts (Leslie & Caldwell, 2009; Pearson & Hamm, 2005). Pearson and Hamm (2005) speculated “that for many educators, it flunks the prima facie test: It just does not have the look or feel of what we mean by ‘comprehension assessment’ “ (p. 38). They also described the underuse of SVT as “unfortunate” (p. 38), a sentiment echoed by Leslie and Caldwell (2009) who declared SVT “worthy of more than a second glance” (p. 418).

Summary of Assessment Literature

Of the four assessment types examined above, each has its own advantages and disadvantages as a tool for measuring students' comprehension of a specific text. Multiple choice assessments are widely used and accepted as an assessment type, and they are easy to administer and score. Thorndike and Thorndike (2010) concluded that multiple choice items can measure “not only the recall of knowledge but also the use of skills of comprehension, interpretation, application, analysis, or synthesis” (p. 284),

which is of particular relevance to a comprehension assessment that aims to measure both textbase comprehension and situation model construction. However, they go on to content that while writing multiple choice items is easy, writing high quality multiple choice items is difficult.

Oral retellings are specifically grounded in comprehension theories (Leslie & Caldwell, 2009) and can provide a more naturalistic assessment environment, as students engaged in retelling are not constrained by structures imposed by other assessment types. However, the results of oral retellings can be confounded by the verbal abilities of the students being assessed, making the results less reliable. Additionally, oral retellings require each student to be assessed individually, which can be impractical for larger-scale studies.

Like multiple choice assessments, the administration and scoring of cloze and maze assessments is easy, and, unlike multiple choice assessments, creating cloze and maze assessments is a simple process. However, as described above, research has shown cloze and maze assessments to be ineffective measures of global-level comprehension. Given the centrality of inferencing to the comprehension of text, a measure that does not assess that aspect of comprehension will introduce threats to construct validity.

Finally, SVT offers an assessment that combines the ease of construction and scoring found in cloze and maze procedures with the construct validity found in good multiple choice assessments. The literature indicates that SVT is a good choice as an

assessment for the present study. However, based on that review of the literature, it was determined that a combination of a traditional multiple choice assessment with SVT would yield the most accurate measure of students' comprehension of the text. This determination was made for a variety of reasons. First, as Fletcher (2006) indicated in an introduction of a special assessment issue of *Scientific Studies of Reading*, "there are one or more latent variables that represent the different components of reading comprehension that are imperfectly indexed by individual measures, but these indices are correlated" (p. 327). Therefore, the use of multiple assessments can provide a more accurate estimation of comprehension (Andreassen & Bråten, 2009; Fletcher, 2006). Second, the extent to which results of SVT assessments have been proven to be valid and reliable made it an attractive option for this study (e.g., Marcotte & Hintze, 2009; Royer, Greene, & Sinatra, 1987; Royer, 2001, 2005). The inclusion of multiple choice questions also should help to mitigate concerns about either floor or ceiling effects, due to the ease with which the difficulty of multiple choice questions can be calibrated given expert assistance. In this case, that expert assistance consisted of both the participating classroom teachers and children's literature/reading experts from a nationally recognized university.

Finally, both multiple choice assessments and SVT are convenient to administer. Both can be administered in a group setting, as opposed to other procedures, like oral retellings, that must be done individually. They are familiar formats, so very little

additional instruction is required for students to be able to complete them. Moreover, as they are objective measures, they can be scored consistently, as opposed to free-response assessments, which must rely on multiple raters to ensure reliability of scores.

Interactive Digital Design Principles

The field of human-computer interaction (HCI) is concerned with “the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them” (Hewett et al., 1996, p. 5). Among the central concerns of HCI is usability. Usability has been defined by a number of sources (Jokela, Iivari, Matero, & Karukka, 2003). Among the most cited definition is the International Organization for Standardization’s, which defined usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (ISO, 1998).

Bevan (1995) provides a more nuanced definition of usability. He argues that usability is comprised of two complementary roles: The first is the usability designed into the product (i.e., whether the product is easy to use and effective), and the second is the “highest level quality objective which should be the overall objective of design” (p. 116) (i.e., whether the product can be used for its intended purpose). Therefore, a piece of software must be usable not only in form, but also in function when applied to a specific task.

Design and usability guidelines for digital-human interfaces are numerous and overlapping. An influential work in the field of design is Don Norman's *The Design of Everyday Things* (2002). In it, Norman outlines six design principles that influenced the design of the e-reader used in this study:

- Visibility – usable and useful functions should be visible to the user;
- Feedback – the system should acknowledge having received an input from the user;
- Constraints – an interface should be designed such that its only functions are those intended by the designer;
- Mapping – the relationship between a control and its effects should be clear, intuitive, and consistent;
- Consistency – similar controls should function in a similar way, with minimal exceptions; and
- Affordance – clues about the function of a control should be embedded in the control itself.

Similar to Norman's principles are another widely-used set of heuristics for the creation of a useable digital interface, created by Nielsen (1995). These ten heuristics were based on a factor analysis that identified nine heuristics assembled from a database of 249 usability problems spanning 11 projects (Nielsen, 1994). The ten heuristics are:

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Aesthetic and minimalist design

- Help users recognize, diagnose, and recover from errors
- Help and documentation

A usable interface—one that allows its users to effectively and efficiently accomplish a task without causing dissatisfaction—will necessarily satisfy many, if not all, of these heuristics.

Neural Networks

An artificial neural network is a digital network “composed of a large number of simple processors (neurons) that are massively interconnected, operate in parallel, and learn from experience (examples)” (Specht, 1991, p. 572). First brought to the attention of researchers by McCulloch and Pitts (1943) and Marvin Minsky, Seymour Papert, and Frank Rosenblatt in the late 1950’s, artificial neural networks (ANN) are inspired by their neurobiological namesakes. The human nervous system is made up of nerve cells, called *neurons*. Neurons are, essentially, biological input/output devices; electrical impulses are received by branches protruding from the neuron (*dendrites*) and are transmitted by different branches (*axons*) from the neuron. Locations where axons interface with dendrites are called synapses. Neurons are massively interconnected in vast networks. Each neuron in the cerebral cortex of humans, for example, is connected to between 10^3 and 10^4 other neurons (Jain, Mao, & Mohiuddin, 1996). The communication between synapses can be regulated by the neurons themselves, so that their ability to transmit to

other neurons can either be inhibited or enhanced, depending on conditions. In this way, neurons can learn from the environment in which they are functioning.

A digital neural network operates similarly. Within a network of artificial neurons, a single digital neuron or group of neurons (called a *node*) can receive inputs, and, by the application of either positive or negative weighting values, can produce an output. Like their biological counterparts, neural networks can “learn” through exposure to new data. The “learning” is embodied by the alteration of the weighting factors used in calculating the neuron’s output and the refiguring of the network architecture. Jain, Mao, and Mohiuddin (1996) described the significance of this ability to learn:

[Artificial neural networks’] ability to automatically learn from examples makes them attractive and exciting. Instead of following a set of rules specified by human experts, ANNs appear to learn underlying rules (like input-output relationships) from the given collection of representative examples. This is one of the major advantages of neural networks over traditional expert systems. (p. 34)

This type of learning is useful in a variety of fields, including forecasting, data compression, and, most relevant to the present study, pattern recognition (Gershenson, 2003).

CHAPTER 3

METHODOLOGY

In order to examine the correlation of certain types or patterns of annotations with students' levels of reading comprehension, seventh-grade students read a short story using a Web-based e-reader. The e-reader has the capacity to capture and save students' highlights and textual annotations. Before reading the story, the researcher instructed students to highlight the text at moments when students make certain types of inferences about the text. Immediately upon concluding their reading, students took a test that measured their comprehension of the short story. The students' patterns of highlighting and annotations and their scores on the posttest were then analyzed using a neural network to determine the nature of the relationship between the patterns of their annotations and their levels of comprehension.

Research Design

This study uses a quasi-experimental posttest-only design. The primary reason for choosing this design is that this study is intended to be a only preliminary investigation of the relationship between the patterns or types of highlighting with students' comprehension of a specific text. As such, the most important data are the results of the use of the e-reader; the review of the literature of highlighting and reading on a digital screen reveals little chance that the process of highlighting and annotation using the e-

reader will produce any measurable effect on the students' comprehension, so therefore a pre-post control-group experimental design is unnecessary. Furthermore, a pretest would necessarily introduce a testing threat to the study's validity. A testing threat is one in which "participants become familiar with the outcome measure and remember the responses for later testing" (Creswell, 2009, p. 164). It would certainly be introduced, as students would be primed to look for answers to questions during their reading of the text following the pretest.

Methods

Tool Development

In preparation for this project, I worked with a programmer to develop an e-reader which possessed the capacity to capture students' annotations as they read the text. The e-reader is a Web-based piece of software, programmed primarily in Javascript, that displays a text to users. Embedded in the e-reader is annotation functionality. Users can choose one of three colors to highlight sections of text by clicking-and-dragging a mouse across the screen. When the user releases the mouse button after a highlight, a box appears in which users can type brief notes based on the text. After typing the note, the box disappears, and an icon indicates that presence of the comment. Users can access the comments by hovering over or clicking on the icon. See Figure 3 for screenshots of the interface.

The design of the e-reader was informed by three sources: my personal experience observing middle school students' annotation of texts in a classroom setting, a pilot test with a small group of student participants, and human-computer interaction usability principles (Nielsen, 1995; Norman, 2002).

Researcher experience. Prior to conducting this research, I spent eight years teaching middle school English. As part of my program of instruction, I would instruct and observe students on the principles of active reading, which encourages students to leave notations in texts as tangible reminders of their thought processes as they read (Faber, Morris, & Lieberman, 2000). Though my experience was limited to analog, paper-based annotations and highlighting, it indicated that students will indeed actively read a text if instructed to do so, and the nature of these annotations and highlights varied considerably. My work with students and my experience studying technology in the classroom suggested that students would be able to translate the ability to read actively to a digital environment, particularly if the digital environment closely simulated the physical world.

Pilot testing. My experience with students and active reading in the classroom informed the design of the e-reader, but more rigorous testing was required before the system could be used with all of the participants in the study. To that end, a pilot test was conducted to evaluate the usability of the software. One class of seventh-grade students ($N = 21$) from Roland Ryans Middle School participated in the tool development pilot. A

more thorough description of the site and participants can be found in Site Selection and Participants, below. These students read the same text used in the main intervention (see Text Selection, below) using the e-reader.

Two data sources informed the conclusions drawn from the pilot study. The first was researcher observation. The researcher observed the students as they used the software, both during a free practice run in which students used the software without attending to the actual text, and during a more formal trial, in which the students read the text used in the main study.

The other data source that informed the evaluation of the pilot test was the data recorded by the e-reader. The e-reader has the capability to record not only the textual highlights and annotations, but also environmental data, including information such as time spent per page, number of clicks per page, and attempts to highlight without first selecting a highlighter (see Data Collection, below, for more on the e-reader's data collection capabilities).

Each data source informed the others during the analysis portion of the pilot test. Observations, including brief, informal conversations with users, revealed that students were able to use the software with no apparent difficulties. Similarly, examination of the log data did not suggest any irregularities that could be associated with difficulty using the software. Such evidence could have taken the form of multiple highlights of the same information, unreasonable amounts of clicking on navigation buttons, or multiple

refreshes of the browser windows. Upon completion of the analysis, the decision was made not to modify the tool in any way.

Tool design. The design of the present e-reader reflects many of Nielsen's (1995) heuristics. The page navigation bar, for example, reflects the idea that the system status should be visible to the user, and that the system should include flexibility and efficiency of use. The navigation bar displays the user's progress through the text by displaying "Page X of X" along with a slider that indicates proportional progress (see Figure 3). Users can move to the next or previous pages through a single mouse click, press of the right or left arrow keys on the keyboard, or by clicking-and-dragging the status bar.

Important to the design of the interface was the principle that there should exist a "Match between system and real world." Therefore, the e-reader reflects the design of a paperback book as closely as possible. The system is designed to be used in full-screen mode in a Web browser, eliminating the various browser and operating system controls and designs that are irrelevant to the reading task (see Figure 3). The main text box is surrounded by a subtle wood-grain pattern. The only other element present on the page other than the main text box is the highlighter control box, which is removed to the side of the interface, leaving the user's main frame of focus to remain on the text.

Nielsen's principle of user control and freedom states that "Users often choose system functions by mistake and will need a clearly marked 'emergency exit' to leave the unwanted state without having to go through an extended dialogue. Support undo and

redo” (Nielsen, 1995, para. 3). The idea that users will behave unpredictably is certainly relevant to middle school users. Therefore the e-reader was designed with an “Undo” functionality. Additionally, after each highlight the function is immediately turned off, which is the equivalent of the metaphorical highlighter cap being replaced after each use. Students have to reactivate, or “uncap,” the highlighter before each use. The reasoning behind this design decision is that forcing the students to select the highlighter color before each highlight will reduce the possibility that a student would highlight predominantly in one color only because that color had been previously selected, and not for any inferential reasons.

Test Development

The most widely used measures of reading comprehension, such as the Gates-MacGinitie Reading Tests, have thoroughly validated and reliable scores, but they are designed to be used to measure students’ general levels of reading comprehension, not their comprehension of a specific text. Therefore, researchers who wish to assess students’ comprehension of specific texts have to create their own assessments. A review of the literature indicated a number of approaches to assessing comprehension of specific texts (see Assessment, above). The determination was made to combine a multiple choice assessment with SVT to assess students’ comprehension in this study.

Another determination was whether to have the text available to students as they took the assessment. In a study of the relationship between aspects of comprehension

(e.g., working memory, word recognition, strategic text processing skills, and reading motivation) and different multiple choice measures, Andreassen and Bråten (2009) concluded that

When the text is not available for rereading and search during testing, it seems reasonable that students' comprehension performance comes to rest at least as much on the quality and accessibility of those representations as on their word recognition skills. (p. 277)

Therefore, the determination was made not to have the text accessible to students as they took the assessment. The point of the assessment is to test the extent to which students formed a coherent and textually accurate mental representation of text, and therefore providing students access to the text during the assessment would confound the results. Particularly with regard to the SVT, much of the assessment would function as a scavenger hunt for the original sentences, as opposed to a test of comprehension.

Multiple choice test development. Construct validity was assured primarily through the process of expert review. An expert in children's literature was consulted in the creation of the multiple choice responses. This expert read the text and then participated in the item development procedure.

The item development procedure consisted of the following steps:

1. Define the construct – In service of content-related validity, a test blueprint was created (Appendix B). The test blueprint is “an explicit plan that guides test construction. The basic components of a test blueprint are the specifications of

cognitive processes and the descriptions of content to be covered by the test” (Thorndike & Thorndike-Christ, 2010, p. 156). The test blueprint was completed in partnership with the children’s literature expert. The test blueprint is a two-column table. One column contains process objectives, which includes textbase and situation model objectives from both local- and global-textual levels. The second column includes specific items from the text, which corresponded to the process objectives. Weighting of each of these areas was performed in consultation with the expert, and it determined the proportion of the different question types.

2. Create item writing guidelines – Item writing guidelines provide a set of criteria that ensure that items are written in such a way that construct irrelevance variance is minimized. Guidelines include issues relating to the page formatting (e.g., “Item choices will be written vertically,” or “Number the items,” etc.), the phrasing of items (e.g., “Avoid double negatives,” or “All directions and question stems should be written at a fifth-grade reading level”), and the items themselves (e.g., “Items should not occur in the same order as the plot of the text occurs”). The item writing guideline can be found in Appendix C.
3. Create items – It was determined that the multiple choice section should include approximately 20 questions. This number was deemed to be enough to ensure reliability of scores, but not so many that items were forced to overlap or to

become unreasonably granular. Therefore, a minimum of 30 questions were initially created. The expert helped to winnow that number to the 22 best questions.

4. Test inspection – The children’s literature expert reviewed the assembled items and accompanying directions to ensure that the item writing guidelines were followed.

To minimize construct irrelevance variance, particular care was taken during the item development to avoid reading independent questions. Reading independent questions are those that can be answered without having read the text being assessed, either from the application of reasoning skills (Andreassen & Bråten, 2009) or by asking about information the student already knew (McKenna & Stahl, 2009). Further, care was also taken to ensure that only questions whose answers can be generated by a reading of the text would be included.

To illustrate the issue of reading dependency, consider the following example, taken from McKenna and Stahl (2009):

Crows

Crows are large black birds that pose a threat to farmers. They belong to the genus *Corvus*.

1. What color are crows?
2. Have you ever had a pet crow?
3. To what genus do crows belong?
4. How long do crows usually live?

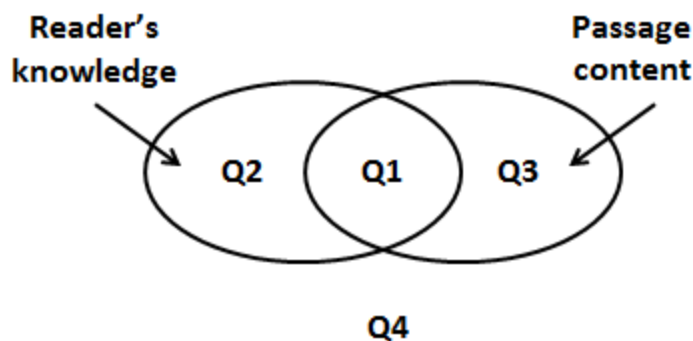


Figure 4. Passage dependency (adapted from McKenna & Stahl, 2009).

The answer to question 1 does come from the passage, but it also reasonable to expect many readers would know that crows are black. Therefore, question 1 is, at best, only partially reading dependent. Question 2's answer is entirely dependent on reader's own knowledge, and thus is completely passage independent. Some readers may know the answer to question 3, but likely for most readers that question is passage dependent, and is therefore the best of the four questions at assessing reading comprehension. The answer to question 4 lies neither in the passage, nor in most readers' knowledge, and would therefore be a poor comprehension question.

Passage dependency is, by definition, relative to the reader. Further, passage dependency is less of an issue in fiction, and more of an issue in the testing of information texts, which relies less on some of the more inferential processes present in fiction. Still, passage dependency is enough of an issue with fictional texts that the inspection of the posttest for reading independent questions was warranted.

Sentence verification technique. There is a prescribed procedure for creating a sentence verification technique (SVT) assessment. When using SVT for longer passages, such as the story used in the study, the first step is to identify “important” (Royer, 2001, p. 35) sentences from the passage. For this study, “important” passages were those that corresponded to the content items contained in the multiple-choice test blueprint, described above.

Once the sentences for use in the study were identified, then four sentences were written for each in accordance with published guidelines: an original sentence, a paraphrased sentence, a meaning change sentence, and a distractor sentence. It was determined that 8 sentences (32 test sentences) would provide sufficient data for determining reliability of scores. A sample of the four sentence type follows:

- Leo was slow in reading, slow in numbers, slow in understanding nearly everything that passed before him in a classroom. (original)
- Leo was a poor reader, bad at math, and he had a hard time understanding most everything that happened in class. (paraphrase)
- Leo was slow in reading, and slow in numbers, but he understood nearly everything that his teacher covered in class. (meaning change)
- Leo was a slower runner than all of his classmates (distractor)

A larger sample of SVT sentences can be found in Figure 5.

Following the example of Kardash, Royer, and Greene (1988), the items were arranged such that the sentences clusters drawn from the first half of the story appeared in the first half of the test. This was done to decrease the role of short-term memory in respondents’ choices.

The final SVT was evaluated by the children's literature expert, with particular attention paid to ensuring that (1) the content of the paraphrased sentences did not differ from the original content, (2) meaning change sentences were altered at a level between a minor, insignificant change, and a change so major that the answer becomes too easy to identify (Royer, 2001), and (3) distractor sentences were "comparable to original sentences in vocabulary, theme, and syntax" (Royer, 2001, p. 34).

Pilot testing. Prior to administration, the multiple choice/SVT posttest was pilot tested with the same students used in the tool design pilot test ($N = 16$). The researcher was present during the pilot testing and recorded flaws observed in the test. Following the pilot test, an item analysis was conducted determine which, if any, items need to be eliminated from the test. Results of the pilot test can be found in chapter 4.

Site Selection and Participants

Site Selection. This study involved each seventh grade student enrolled in a language arts class at two separate middle schools. The decision to use middle school students arose from a desire to focus on the middle grades as a key part of the comprehension development process. The choice of seventh grade students in particular was born from the idea that this type of research is not necessarily profitable with students younger than fifth grade, where the nature of the reading tends more towards textbase-level comprehension, and less toward the development of a situation model (Chall & Jacobs, 2003). The schools were chosen for three reasons. First, the presence of

a preexisting relationship with the school district's central administration provided an ease of access to the schools. Second, the schools were geographically accessible for the study. Finally, both seventh grades have a one-to-one laptop initiative, which eliminated the potential barrier of access to technology for the study.

Both of the middle schools are located in the same school system, which serves a largely rural county in a Southern state. One of the middle schools, Roland Ryans Middle School², serves the students from the main population center of the county, primarily. The enrollment is 618 students, serving grades six through eight. 53.9% of its students identify as white, while 19.1% identify as black, and 13.8% identify as Hispanic. 14.2% of the students are classified as having limited English proficiency, and 46.3% are designated disadvantaged, indicating the students receive free or reduced price lunches. The school has roughly the same number of gifted students (12.5%) as students with disabilities (14.6%), with *disabilities* being characterized as “those identified for special education services, from speech pathology and learning disabilities to severe and profound disabilities.” The three-year pass rates for the state standardized test of reading are 89% for 2010-11, 91% for 2011-12, and 71% for 2012-13 (“Fact Sheet - Roland Ryans Middle School,” 2013).

The other middle school, Johnson Middle School, serves the more rural areas of the county. It is much smaller than Roland Ryans Middle School, with a total enrollment

² School names are pseudonyms.

of 356 students spread between grades six through eight. The school is made up predominantly of students identified as white (74.4%), with 12.1% identified as black, and 6.2% as Hispanic. 40.2% of the student population is considered disadvantaged. Johnson Middle has the same percentage of disabled and gifted students as Roland Ryans Middle. The three-year pass rates for the state standardized test of reading are 85% for 2010-11, 87% for 2011-12, and 68% for 2012-13 (“Fact Sheet - Johnson Middle School,” 2013).

Participants. Two hundred sixty-one students participated in this study, 148 students from Roland Ryans Middle and 113 students from Johnson Middle. Demographic data was collected via an electronic survey administered upon completion of the intervention procedure. The decision was made to collect these data separately from the main procedure of the study in order to reduce the potential for the manifestation of stereotype threat (Steele & Aronson, 1995), which refers to a risk that an individual will subconsciously confirm a negative stereotype held about a group to which the individual self-identifies. Two hundred forty-eight students (95.0%) elected to complete the demographic survey.

The average age of the participants was 12.49 ($SD = 0.55$), and the sample was almost evenly divided between males (49.9%) and females (50.1%). A large majority of the participants speak English at home (90.7%), with Spanish being the second most common language spoken by the participants at home (4.84%). The remainder of the

participants spoke other languages at home. Of the 248 respondents to the survey, 59 (23.8%) of the students declined to self-identify their ethnicity. The self-identified ethnicities of those who did choose to respond are located in Figure 6.

Text Selection

The text used in the study is a short story entitled “Slower Than the Rest,” which appears in the collection *Every Living Thing* (Rylant, 1985). The selection of the text was made in consultation with the participating teachers and a children’s literature expert at a university located in a Southern state. It was determined that the text selected would have to meet three criteria: accessibility, ability to engage readers, and ability to cause multiple inferences.

Accessibility. The first criterion was that the text be accessible for the student participants. In practical terms, an accessible story would be one that is brief enough to be read in one sitting by the majority of the student participants, and one that is gauged to be leveled appropriately by multiple measures of reading levels. At 1,093 words, “Slower Than the Rest” appeared to qualify as sufficiently brief, and that intuition was supported by the participating teachers.

For the sake of reading comprehension, researchers need to match the readability of a text with the reading level of its readers (Fry, 2002). Or, put plainly, “You can’t learn much from books you can’t read” (Allington, 2002, p. 16). A number of measures of reading level have been created to enable the quantitative matching of texts to different

levels of readers. Because there is variation among the most common measures of readability, a range of readability measures were used to determine the suitability of “Slower Than the Rest” for this study.

The Flesch-Kincaid Reading Ease scale (Kincaid, Fishburne, Rogers, & Chissom, 1975) calculates the reading level of a text through analysis of word length and sentence length. The resulting score ranges from 0 to 100, with higher numbers indicating easier readability. “Slower Than the Rest” scored a 84.3 on the Flesch-Kincaid Reading Ease scale, which is higher than the cutoff Flesch set for “easy” texts. That score can be converted to find a Flesch-Kincaid Grade Level score, which corresponds with the American system of grade levels. The 84.3 Flesch-Kincaid Reading Ease score of “Slower Than the Rest” corresponds to a Flesch-Kincaid Grade Level score of 4.9, indicating the story is appropriate for students at a late-fourth/early-fifth grade reading level.

The Lexile Framework (Lennon & Burdick, 2004) functions similarly to the Flesch-Kincaid scores in that it scores the difficulty of a text by using a proprietary algorithm that uses word length and sentence length to calculate a score. Scores range from 200L for beginner texts to 1700L for advanced texts. Lexiles are used to match readers with texts in schools, and they are the readability measure of choice in the Common Core State Standards (CCSS; National Governors Association Center for Best

Practices & Council of Chief State School Officers, 2010b). Lexiles do not judge the absolute readability of a text, however:

When reader and text are appropriately matched, a reader can enjoy a comprehension rate of about 75 percent. The 75-percent comprehension level corresponds to that balance of skill and difficulty that allows reading be a positive but adequately challenging experience, and encourages the reader to grow in proficiency and motivation. (Lennon & Burdick, 2004, p. 3)

Using Lexile Analyzer (www.lexile.com/analyzer), Lexile's official Web-based Lexile calculator, "Slower Than the Rest" scored 930L. According to the CCSS conversion chart, the text is appropriate for grades 4-5 (see Table 1). A pre-intervention interview with three of the participating teachers suggested that texts falling into a fifth-grade reading level would be most appropriate for the students involved. Given other, similar measures of readability agreed with the Flesch-Kincaid scores and the Lexile score, the story was deemed to be the appropriate reading level for the student participants.

Ability to engage. The second of the criteria employed in selecting the text used in this study is capacity to engage readers. As there are no direct ways to measure how engaging a text is for readers, indirect methods were used to judge the capacity for "Slower Than the Rest" to engage readers. One such indirect measure was expert opinion. Both a children's literature expert and the lead teacher at one of the schools indicated anecdotally that similar works by Rylant had been well received by students. Additionally, Rylant has won multiple prestigious awards for her other works, including a Newbery Honor, a Newbery Medal, Caldecott Honors, and multiple American Library

Association “Best Book of the Year for Young Adults” awards. Based on these factors, it was concluded that the story satisfied the engagement criteria.

Ability to generate multiple inferences. Finally, the third criteria for selecting a story was that the story needs to possess the capacity to cause multiple inferences to be created in its comprehension. As the students were asked to highlight the text at moments when they were drawing inferences about the text, it was crucial to find a text that satisfied this criteria.

An inferential analysis was conducted of the text. Specifically, the inferential analysis examined the text for places in which students will be required to draw one of two types of inferences in order to comprehend the text and subtext of the story, coherence and elaborative inferences. Coherence inferences are the “gap filling” inferences, in which separate elements of the text are joined to make meaning. Elaborative inferences include those which involve the reader bearing his or her own knowledge and experience to the text, whether it be for interpretation or for prediction. A sample of the inferential analysis follows (the full inferential analysis can be found in Appendix D):

c = coherence-type inference
e = elaborative-type inference

Leo said much more // (e – about forest fires). Mostly he talked about Charlie, explained what turtles were like, the things they enjoyed, and what talents they possessed. He talked about Charlie the turtle and Charlie the friend, and what he said and how he said it made everyone in the class love turtles and hate forest

fires. Leo's teacher had tears in her eyes // (c – Leo's teacher was made emotional by Leo's presentation) (e – this is unusual/unexpected behavior from Leo).

Intervention Procedure

The intervention began by training students on the use of the e-reader. Using a sample selection from a different Rylant text, students practiced navigating through the pages of the interface, selecting and using the various highlighting tools, and leaving comments. The mechanics of highlighting were familiar to all of the students; as part of a one-to-one technology program, each student has had his or her own laptop for the whole school year, and the intervention followed the winter break, so the students had sufficient experience with this basic function.

The concept of highlighting moments of inference was then explained. Students were instructed to use the blue highlight color, labelled “Reminds me of...,” at any moments that evoked something from his or her prior experience, whether that experience be from the student's own life, or from something the student read earlier in the text. The students were instructed to use the green highlight color, labelled “A-ha!,” when a passage or word in the text caused the student to have a moment of insight into the text, or when the student perceived he or she was reading a moment of particular importance to the text, even if the student could not perceive the nature of that importance. Finally, they were instructed to use the red color, labelled “Prediction,” at any moment in the text in which they find themselves making a prediction about the text.

Students then read “Slower Than the Rest” using the e-reader in class. Students were given however much time they needed to complete the reading, with the expectation (provided by the classroom teachers) that none of the students will require longer than the 80-85 minutes of class time allotted for language arts class to complete the reading.

Upon completion of the reading, all students took a pen-and-paper version of the posttest. The posttests were collected and scored by the researcher. In this case, a pen-and-paper test was preferable to a digitally delivered test. The test did not take advantage of any of the affordances provided by digital tests—for example, this test was not designed to be an adaptive test, nor did it include any multimedia elements. Indeed, the only advantage of delivering this test electronically would have been ease of scoring. However, the potential for the digital format introducing construct irrelevant variance exists, which ultimately outweighed the benefits of efficient scoring.

The decision to administer the test immediately following the reading was based on the idea that differences in memory capacity among students would manifest itself more strongly the more time elapsed between completion of the reading and the administration of the assessment. Ideally, students will rely more on the situation model formed during the reading rather than memory if the test is given immediately following the reading.

Fidelity of Implementation

Fidelity of implementation was ensured by having the research scheduled so that the researcher was present for the interventions in each of the 13 classes. A weather event forced one class at Roland Ryans to be rescheduled so that the researcher could not introduce the procedure. However, the teacher in that class had been present for three other administrations of the intervention, and had been a student of the researcher in a teacher training program. Given that level of familiarity, the decision was made to allow that teacher to introduce and supervise one implementation of the intervention. The researcher was delivering the intervention at another class at Roland Ryans at the same time, so the researcher was available, if needed.

Access, Role, and Ethics

I gained access to the classrooms used in this study through the following steps:

1. I briefed the local school board's Director of Educational Technology and Professional Development, who then put me in touch with the principal of Johnson Middle and the assistant principal of Roland Ryans Middle School. Subsequent face-to-face meetings with both administrators resulted in initial approval to conduct the research in those schools.
2. I communicated the scope and aims of the project directly with the teachers at Roland Ryans Middle and indirectly through the principal of Johnson Middle. Teachers all agreed to participate, and a general time frame for the

intervention, January 2014, was agreed upon. Additionally, all teachers agreed to the proposed text to be used in the study.

3. My university's Institutional Review Board approved the research. The approval process included the drafting of a parental notification letter that will be distributed to the student participants. Students will have the opportunity to opt out of the research.

I was an active participant in the research only to the extent that I introduced the software to the students. In addition, I was present in each school during the days of the student reading and posttest, in case a need for technical support or other types of questions arose.

Care was taken to ensure the confidentiality of all participants involved in the study. Students were identified by the e-reader software by a code. The key to the code was kept by the researcher on a computer enabled with biometric (fingerprint scanning) security. Hard copies of the posttest were kept in a locked filing cabinet. Teachers will not be given access to identifiable results.

Data Collection and Analysis

Collection. Data came from three sources: logs and nature of clicks made using the e-reader, the students' comments on the text recorded by the e-reader, and the posttests. Data collected by the e-reader included:

- Login times
- Duration of time spent per page
- Text highlights, both words highlighted and color used
- Comments left by readers
- Timestamp data (i.e., exact date and time of each recorded activity)

From these data, other variables were calculated, including:

- Length of time spent with the e-reader open
- Number of highlighted passages
- Highlights per page
- Average highlight length
- Average words per comment

While virtually any interaction between the reader and the software can be captured, the literature indicated that the most useful data would arise from the sources listed above.

Posttest data indicated each individual student's comprehension of the text. The posttests were hand scored, using the answer key developed at the time of item development by the researcher.

Analysis – neural network. A software product called *NeuralTools* (<http://www.palisade.com/neuraltools/>) was used for this study to calculate the relationship between the patterns of the students' highlighting and their comprehension of the text as measured by the posttest.

The process of neural network data analysis was a two-stage process. In the first stage, the model training stage, the highlights and scores of 225 randomly selected students from the population were analyzed using *NeuralTools*. The software is designed to weight each of the independent variables until a fit is achieved.

The highlights and scores of the remaining 25 students were held out of the initial model training stage, and were used subsequently in the second stage, the model testing stage. Holding these scores out of the initial training stage helps to reduce internal validity threats posed by the test set's contributions to the testing model. This phase is intended to test the predictive power of the model generated in the model training stage. The absolute values of the differences between the predicted scores and the actual scores, called residuals, were used to judge the quality of the fit generated by the neural network.

Analysis – multiple regression. Multiple regression is a statistical tool used to predict outcome variables from several predictor variables (Field, 2009). In this study, the neural network analysis suggested which of the independent variables were most influential in the various networks created. The influential variables then became the predictors in the multiple regression, with the students' posttest scores serving as the outcome variable. The purpose of the multiple regression analysis was to confirm the neural network's findings regarding the influence of the variables on posttest scores.

CHAPTER 4

RESULTS OF THE STUDY

The goal of this study is to examine the relationship between the types of inferences students make, as indicated by their highlighting and annotation of a text, and their level of comprehension of that text. This chapter reports the results of the investigation, divided into the following main sections:

1. The posttest development and refinement section reports the results of a pilot administration of the posttest, describes the test refinement process, and presents the results of the main posttest administration.
2. The data collection results section describes the data collected by the highlighting software and presents the descriptive statistics that define the data set.
3. The data exploration section describes the process by which the neural network helped to identify the factors that influenced students' performance on the posttest.

Posttest Development and Refinement

A 54-question posttest was developed, following the procedure outlined in the Test Development section of the Methodology chapter. The test underwent expert review before its pilot administration. A group of seventh grade students ($N = 21$) at one of the participating schools, Roland Ryans Middle, read the short story "Slower Than the Rest" using the e-reader and then took the posttest. Students began the posttest immediately upon the completion of their reading, and the students had the remainder of the class

period to finish the posttest. Due to a weather-related alteration in the school's daily schedule, the pilot class met for 63 minutes, a shorter duration than the rest of the classes involved in the study, which met for 80 – 85 minutes. The truncated class impacted the administration of the pilot, as 5 of the 21 students were unable to finish the task. These scores were eliminated from the final analysis of the pilot test.

The 54-question posttest included 22 multiple choice questions, and 32 sentence verification technique (SVT) sentences, four of each type of sentence from eight source sentences drawn from the text. The scores on the posttest ($M = 43.50$, $SD = 8.05$) indicated a negative skewness (skewness = -1.69). Reliability testing indicated the scores generated by the posttest were reliable, Cronbach's alpha = 0.91, standard error of measurement (SEM) = 2.44. The determination was made to administer the full test during the main study, with only minor cosmetic alterations.

Upon completion of the main study, the posttest scores were analyzed to calculate each participant's final comprehension score. An item analysis was conducted using jMetrik (Meyer, 2014). Individual items were examined for discrimination, with items showing a discrimination of between 0.3 and 0.7 considered acceptable. As a result of the discrimination analysis, 16 items (6 multiple choice and 10 SVT) were eliminated. The scores on the posttest ($N = 250$; $M = 29.04$, $SD = 6.08$) were less negatively skewed than the pilot test (skewness = -1.19), and reliability testing indicated that the scores were reliable, Cronbach's alpha = 0.85, SEM = 2.38. Reliability scores in the 0.8 to 0.9 range

are generally considered to be acceptable reliability scores (Field, 2009). Finally, to alleviate concerns related to the negative skewness, a normalized T scores were calculated.

Data Collection Results

Log Data

An advantage of a Web-based e-reader, such as the one developed for use in this study, is that all of the users' interactions with the software can be logged. Each student's interaction with the software generated a log file, from which data were compiled and analyzed. In particular, the e-reader recorded the following data:

- logPageID – identified which text is opened (the e-reader contained two different texts for the students to access: a practice text and “Slower Than the Rest”)
- getData – indicated that the text was successfully displayed on the user's screen and that the highlighting function was active and functioning properly
- selectHighlight – returned a value of “blue”, “green,” or “red,” depending on which color the user selected
- highlight – returned the text that the user highlighted
- message – returned the text of the annotation associated with each highlight, if present
- nextPage and previousPage – indicated when students used the page navigation buttons
- undoHighlight – indicated that a highlight was deleted by the user

Each event recorded by the software was tagged with the student's unique userID and with a timestamp recording the exact time of each interaction. From these timestamps, statistics like total time on task were calculated.

In addition, a second log was created that captured the details of the highlights and annotations. Each highlight was given a unique identifier. The color of the highlight, the text highlighted, and the text of any annotation was recorded, as described above. In addition, the position of each highlight relative to its page and its position in the short story was recorded, using a character count. For example, a four-letter word highlighted at the beginning of the second page of the story would be counted as having a beginning and ending character of 1 and 4, respectively, relative to page two. In addition, the same highlight would have a beginning and ending character of 958 and 961, respectively, relative to the entire document.

From these raw data, it was possible to calculate a variety of secondary statistics, including:

- total time on task
- total number of highlights
- total comment words
- percentage of highlights including comments

Descriptive statistics were collected based on these data, which are enumerated in the following section.

Descriptive Statistics

In total, the e-reader recorded 1,067 highlights among the 250 participants in this study. During the data cleaning process 139 highlights were purged from the main data set. These omitted highlights consisted of redundant highlights (i.e., a highlight contained

in whole in another, same-colored highlight, without a comment indicating that the multiple highlights were intentional), highlights that consisted of blank space with no associated comments, and highlights that were combined with neighboring highlights (i.e., highlights in which it was clear that the student prematurely concluded a highlight, then continued the process as a second highlight, instead of deleting the truncated highlight). The final data set included 928 highlights spread across the 250 participants ($M = 3.71$, $SD = 3.51$). The mean length of highlights (in words) was 14.53 ($SD = 15.63$).

There were three different colors of highlights, each corresponding to a particular type of inference. The blue-colored elaborative highlights, labelled in the e-reader as “Reminds me of...”, corresponded to text-to-self/self-to-text inferences. Readers were instructed to choose this color highlight when a portion of the text, be it a word, phrase, sentence, or group of sentences, reminded the reader of something that occurred earlier in the text (i.e., a text-to-text connection), or some part of the reader’s prior knowledge or experience. The green-colored highlights, labelled “A-ha!” in the e-reader, were to be used when readers experienced a moment of insight or clarity while reading, or when readers perceived a moment of particular importance, even if the reader was unaware of exactly why the moment was important. The red-colored highlights, labelled “Prediction”, was used as readers found themselves making predictions regarding the text. The “Reminds me of...” highlights were most common ($n = 399$), followed by “Prediction” highlights ($n = 294$) and “A-ha!” highlights ($n = 235$).

Of the 928 highlights, 791 had comments associated with them (per user $M = 3.2$, $SD = 3.1$). These comments contained 7671 words (per comment $M = 30.7$, $SD = 37.1$), and a small number of non-traditional responses, including emoticons and texting abbreviations.

Log data also enabled the calculation of total time on task for each participant. The total time on task was calculated as beginning with the loading of the software, as indicated by the `getData` log entry. However, the design of the software failed to include an “End” button, so the calculation of the termination of the task was done indirectly. As there was no end button, most users clicked on the “Next Page” button at the conclusion of their reading. Therefore, any set of log entries for a particular entry that ended with a “Next Page” click from the last page of the document was concluded to be a sufficient marker to indicate the termination of the exercise. For other users, the time stamp of their last action was used as the termination point. As a result of this lack of precision, total time on task data were considered to be estimates, and any conclusions drawn from the data were interpreted cautiously. The mean time on task for the participants was 12.63 minutes ($SD = 7.6$).

The frequency of the highlights declined as the story progressed with a slight increase occurring on the final page of the story. See Figure 7.

Data Exploration

Following the processing of the raw log data, exploratory investigations of the relationships between the various independent variables and the dependent variable, the comprehension posttest score, was conducted. The primary vehicle for this exploration was NeuralTools, a neural network generation tool that functions as a Microsoft Excel plugin.

Model Training Phase

The configuration, or shape, of a neural network is referred to as its topology. Topologies are dependent on a number of factors, including the nature of the variables (i.e., continuous/categorical), the number of independent variables, and the statistical structure of the data set. The topology consists of nodes, which, like their biological namesakes, are places where inputs come together and are weighted based on their relevance, and outputs are generated. While NeuralTools allows users the ability to manually choose various topologies, it also offers a Best Net Search, in which the software tries multiple iterations of all the different major topology types. It selects what it judges to be the best topology for the data based on the error it calculates between the actual dependent variable values and the values it predicts for the set.

A Best Net Search was conducted using a data set containing 250 rows, one for each participant, and 80 columns, one for each of the 72 sentences in the story, one for each of the three types of highlights, and one each for total time on task, total number of

comments, and total words written in comments. The other two columns were an unused column for the user identification numbers, and one for the dependent variable, which was the normalized posttest scores. It tested the data set using a Generalized Regression Neural Net (GRN) and five Multi-Layer Feedforward Network (MLF). A GRN contains two hidden layers of neurons and one node per independent variable. An MLF can contain one or two hidden layers as well, but it differs from the GRN in that there can be up to six neurons per hidden layer. An MLF is a more powerful topology, in that it can detect patterns in more complex data.

The Best Net search performed approximately 41.5 million trials over a period of 12 hours. The best performing topology was an MLF with three nodes, with a root mean square (RMS) error of 10.28 between predicted and actual values. By comparison, the GRN RMS error was 10.89. See Table 2, below.

As NeuralTools builds its various neural networks, it generates a relative variable impact value for each independent variable. The relative impact value reflects the influence each independent variable has on the predictions made by the network, as expressed by a percent value. The sum of all of the relative variable impact values is always 100%. Each network creates its own relative variable impact values, and it creates variable impact values only on the training data and not on the testing data; therefore, different networks generated by the same data set can produce differing variable impact analyses. So, relative variable impact value factors are not capable of supporting any firm

conclusions about the data. However, impact variables are useful in suggesting which variables in a data set are the most influential. It is this function that contributed information to this study.

The relative variable impact values generated by the neural network's Best Net Search yielded few usable results. The analysis revealed that largest influencer of the dependent variable was sentence 18 of the text. Only five users (2%) highlighted that sentence, and it only influenced the final predictions by a factor of 2.17%. The inability of the neural network to differentiate between the impact of highlights in individual sentences is likely due to the presence of so many null values (95.84%) in the sentences section of the data set.

Therefore, additional exploratory network trainings were conducted, using varying topologies and combinations of independent variables, including types of different highlights, number of comments, number of words included in comments, and time on task, but not whether individual sentences were highlighted. Trials indicated that using total number of highlights by type and total time on task consistently showed the greatest relative impact values. Thus, a second Best Net search was performed using these four independent variables: one each for the three types of highlights, and one including time on task. The Best Net search performed approximately 1.9 million trials over a period of one hour. The best performing topology was an MLF with six nodes,

with a RMS error of 6.63 between predicted and actual values. Table 2 summarizes the two Best Net Searches.

Table 2. Best Net Search Summary.

	Best Net Search 1	Best Net Search 2
Training cases	225	225
Testing cases	25	25
Independent variables	78	4
Topology	MLFN	MLFN
Nodes	3	6
Bad predictions	4	0
Testing RMS error	10.28	6.63
Testing mean absolute error	7.534	5.57
Testing std. dev. abs. error	6.990	3.59

The relative variable impact values generated by this Best Net Search yielded more usable results than the first Best Net Search. The analysis suggested that Prediction highlights and “A-ha!” highlights were the most influential variables, but that total time on task and the elaborative (“Reminds me of...”) highlights were also influential to a

lesser extent. This conclusion was confirmed by other various trial runs, and it informed the regression analysis, below.

Model Testing Phase

The model testing phase evaluates the trained model's ability to predict dependent variables, by removing a specified number of cases from the training set, then using the omitted cases to predict the values of the dependent variables. These predictions are then compared with the actual dependent variables. Each prediction is judged "Good" or "Bad" based on its residual, or, the difference between the predicted value and its actual value.

Twenty-five (10%) of the cases were randomly selected by the NeuralTools software and withheld from the training sets of both networks trained in the Best Net Searches. The neural net then predicted the posttest scores of these withheld cases. In the first Best Net Search, 21 of the 25 predictions (84%) were considered to be "Good" predictions by the software. The mean absolute residual value of these predictions (i.e., the absolute difference between the predicted score and the actual score) was 7.46 ($SD = 5.68$). In the second Best Net Search, all 25 of the predictions (100%) were considered to be "Good" predictions. The mean absolute residual value of these predictions was 5.57 ($SD = 3.59$).

In order to further investigate the nature of the relationship between the types of inferences generated during reading (as indicated by highlights) and students' level of

comprehension of the text, an independent samples t-test was conducted to determine if students who made two or more Prediction or two or more “A-ha!” inferences scored significantly higher on the posttest than those who did not. Posttest scores for students who made two or more predictions were significantly greater than those who made fewer than two predictions. Similarly, posttest scores for students who recorded two or more “A-ha!” moments were significantly greater than those who recorded fewer than two “A-ha!” moments. In contrast, posttest scores for students who made two or more elaborative inferences (“Reminds me of...”) did not reach a statistically significant difference than those who made fewer than two elaborative inferences. See Table 3, below, for results.

Table 3

Comparison of Posttest Scores Between Students Who Highlighted Multiple Instances of Each Type of Inference.

	Two or more instances		Fewer than 2 instances		95% CI for Mean Difference	<i>p</i>	<i>t</i>	<i>df</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Reminds me of...	50.62	8.97	49.50	10.46	-1.38, 3.62	.38	.881	248	.11
A-ha!	53.04	9.25	49.09	9.89	1.02, 6.88	.01	2.65	248	.41
Prediction	41.96	5.57	38.27	8.18	2.93, 8.05	< .001	4.22	248	.53

Note. *M* = Mean. *SD* = Standard Deviation. CI = Confidence interval. *M* and *SD* of comprehension posttest. *d* = Cohen’s *d*.

Regression Analysis

A multiple regression analysis was conducted to provide convergent validity for the results suggested by the neural network. The factors included in the regression analysis were those suggested to be influential by the neural network analysis: number of elaborative inference highlights per user, number of “A-ha!” moments highlighted by user, number of predictions highlighted by user, and total time on task.

Tests for multicollinearity indicated that a low level of multicollinearity was present ($VIF = 1.23$ for elaborative highlights, 1.40 for “A-ha!” highlights, 1.41 for prediction highlights, and 1.02 for time on task). Prediction highlighting was the first variable entered, followed by time on task, then “A-ha!” highlights, then elaborative highlights, as suggested by the neural net exploration. An examination of a scatterplot of residuals and predicted values indicated that assumptions of the homoscedasticity of residuals were not violated.

Results of the regression analysis confirmed that prediction highlights were influential. Time on task was minimally influential. Neither the “A-ha!” highlights nor the elaborative highlights were influential. The resulting model containing all four of the factors significantly predicted the posttest scores. See Table 4, below. However, the R^2 value indicates that 88% of the variance in the posttest scores are not explained by this model.

Table 4

Predictors of Comprehension Posttest Scores

	Comprehension posttest scores				
	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	<i>p</i>
	B	Std. Error	β		
Reminds me of...	0.24	0.47	0.04	0.50	0.62
A-ha!	0.03	0.45	0.005	0.07	0.95
Prediction	1.86	0.57	0.25	3.34	.001
Time on task	-0.005	0.001	-0.24	-3.81	< .001
R^2	0.12				
<i>F</i>	7.50				

The multiple regression analysis confirmed the neural network models. This was determined by using the neural network and the regression model to predict the posttest scores of the same 25 students. There was a strong correlation between the two sets of predicted scores, $r(23) = 0.71$, $p < .001$. The root mean square (RMS) error of the regression score predictions (6.79) was similar to the (RMS) error of the neural network model (6.63). This indicates that the model created by the neural network functions similarly to the regression model.

Summary

The goal of this study is to explore whether patterns of students' highlighting and annotations of a text can be associated with their comprehension of the text. Following

the methodology described in chapter 3, this chapter presented the results of the exploration.

Two neural networks were used to examine the relationship between independent variables collected during the intervention. The variables included in the Best Net Searches were the number of highlights per sentence per reader, the types of these highlights, the number of comments and words contained in these highlights per reader, and each reader's total time on task. The dependent variable was normalized score on a comprehension posttest.

Two neural networks were trained using both a three-node and six-node MLF. The resulting topologies were able to make predictions based on the independent variables. In both cases, 25 (10%) of the cases were randomly selected by the NeuralTools software and withheld from the training set. The neural net then predicted the posttest scores of these withheld cases. Both topologies resulted in a large number of "Good" predictions, with the topology resulting from the data set that omitted the individual sentence highlights making 100% "Good" predictions. The number of "Good" predictions in both topologies suggests there are indeed patterns present that can describe the highlighting and annotations of different comprehension levels. These findings were confirmed by a multiple regression analysis, which predicted posttest scores with similar accuracy using the same factors as the neural network. However, the residuals of the predictions indicate that further refinement of the methodology for this type of

comprehension assessment is necessary. Further, the R^2 value indicates that the variables included in the regression model do not fully explain the difference in the variance between scores.

CHAPTER 5

DISCUSSION

Reading comprehension is a difficult construct to measure (Leslie & Caldwell, 2009; Smith, 2004; Snow, 2003), yet, like any other essential skill taught in schools, assessment is key to its successful instruction. This study aimed to test the feasibility of a novel form of reading comprehension formative assessment. An e-reader was developed that facilitated student highlighting and annotation of texts. Students read a short story using the e-reader, highlighting moments of different types of inference, each indicated by a different highlight color. Students were afforded the ability to leaving comments associated with their highlights. The students then took a posttest, measuring their comprehension of the text. The students' highlights and annotations were then explored with the aim of determining whether patterns exist within the highlights and annotations that correlate with the students' levels of comprehension.

This chapter begins by discussing the result of the exploration of the relationship between highlights, annotations, and comprehension and reviews the research questions outlined in chapter one. Further, it describes findings related to inference-making during the reading process. The chapter then describes the limitations of the study, explores the implications of this work on both the fields of reading and instructional technology, and, finally, it proposes future research.

Review of the Research Questions

This study was framed by one main research questions, with four sub-questions. The intent of these questions was to generate the knowledge needed to determine the feasibility of this system as a method for the formative assessment of reading comprehension. The main research question asked about the nature of the relationship between the types and patterns of students' highlighting and their comprehension of the specific text. The four sub-questions provided the beginnings of an answer to this question.

Research Question 1

What is the nature of the relationship between the types of inferences students make and their level of comprehension as measured by both a multiple-choice assessment and a sentence verification technique assessment?

This question sought to determine whether a relationship existed between students' inferences, as indicated by the highlights and annotations, and their level of comprehension, and, if so, to explore the nature of that relationship. Implicit in this question is another, which asks about the extent to which seventh grade students would be able to indicate the moments when they made inferences by using the e-reader. The completeness of the data (i.e., the fact that the all of the students read the text, highlighted in multiple colors, and left comments) indicates that this is a viable research method

moving forward. However, further study is needed to determine the accuracy with which this methodology can capture the extent of student inferencing.

The presence of a pattern was suggested by the ability of a neural network to predict the posttest scores of a group of students with a reasonable degree of accuracy. The ability of the neural network to make predictions was verified by a subsequent multiple regression analysis, which produced similar predictions as the neural network. Follow-up analyses of the neural network and the multiple regression model indicated that the presence of student predictions were among the strongest influences on the students' posttest scores. Students who made two or more predictions scored statistically significantly higher on the posttest than students who did not.

Research Questions 2 and 3

What is the nature of the relationship between the density and frequency of highlighted inferences and students' comprehension of the text? What is the nature of the relationship between time spent on the reading task and students' comprehension?

Using the same procedure as research question 1, neural networks were trained using data describing the density and frequency of highlighted inferences, as indicated by highlights per page and total number of highlights, as well as time on task, measured by total seconds elapsed. After training on this data, the neural network indicated that density and frequency of highlights did not influence the students' posttest scores. Time

on task did influence the students' performance on the posttest, but it was not a heavy influence.

Research Question 4

What kind of patterns emerge from an exploration of the textual annotations of students of varying reading levels?

The rationale behind this question was to see if students who have already been identified as struggling, on grade level, or above grade level proceeded to highlight or comment on the text in identifiably different ways. This study failed to answer this research question. In order to provide a reasonable answer to this question, results of prior diagnostic testing would need to be available for each participant. The privacy policies of the participating school district prevented the collection of this data. Each parent or guardian would have to release these scores to the researcher, which would be a time-consuming procedure, unlikely to result in a complete data set. Administering a diagnostic test as part of the research proved untenable as well. Ultimately, the decision was made not to pursue these data in this study. Future research will be designed such that diagnostic reading test scores will be available to the researcher.

Additional Finding – Elaborative Inferences and Predictions

The idea of encouraging elaborative inferences, variously referred to in the practitioner literature as self-to-text or text-to-self connections, is well established. For example, it figures prominently in the practitioner-targeted book *Mosaic of Thought*

(Keene & Zimmerman, 1997), which has been described as “landmark” (Bluestein, 2010, p. 597) and was named by an unscientific poll conducted by *The Reading Teacher* journal as one of eight works that has most significantly influenced the practice of teaching reading (Barone & Mallette, 2012). In *Mosaic of Thought*, the practice of making text-to-self connections was highlighted in a chapter-long case study of a second-grade classroom. The teacher, Debbie, was held up as a model instructor because of an eight-week-long instructional unit, in which she taught her students to activate their schema before, during, and after the reading process. Keene and Zimmerman (1997) report that the result of this strategy instruction was that “[t]he eight weeks devoted to schema had left the children with a kind of independence, a kind of power in their thinking that engaged them in reading more thoughtfully, critically, and enthusiastically” (p. 70).

The case study concludes with the book’s authors’ further reflections on text-to-self connections:

The [Public Education Coalition] staff developers have heard dozens of stories from teachers and other friends who became obsessed with making schematic connections once they consciously learned to do so.

I have come to believe that this hyperawareness may be a necessary and inevitable stage in coming to know oneself as a reader. (1997, p. 71)

Mosaic of Thought is but one of many texts that advocates that teaching children to become consciously aware of the connections between the text and themselves. However,

as Fisher and Frey (2012) observed, “The challenge is in not becoming so focused on background knowledge and prior experiences such that we end up spending little time on the textual information. Activation alone, although important, doesn’t expand knowledge” (p. 179). While Fisher and Frey do not cite any research to support their claim, the supposition they make appears reasonable, and is worthy of further investigation. The present study could provide some evidence to support their claim.

The present study can provide context for an investigation of this relationship between text-to-self connections and comprehension of specific texts. One of the three highlight types available to students was the “Reminds me of...” highlights, which students used when a part of the text reminded them of either a moment or situation from their own past experience, or from the text they were reading. This was the most frequently used of the three highlight types, with 43% of all highlights being of this type. A statistical investigation concluded that students whose inferences were predominantly of this type, with “predominantly” defined as 75% or more of the student’s total highlights, scored statistically significantly lower ($M = 47.43$, $SD = 8.50$) on the posttest than other students ($M = 51.36$, $SD = 10.20$; $t[208] = 2.34$, $p = .02$). See Table 5, below. In addition, as reported in chapter 4, making multiple elaborative inferences did not correspond to any significant difference in posttest scores, whereas making multiple inferences of the other two types did correspond to statistically significantly higher posttest scores.

Table 5

Comparison of Posttest Scores Between Students Who Predominantly Made Elaborative Inferences and Those Who Did Not.

Fewer than 75% highlights are “Reminds me of...”		More than 75% highlights are “Reminds me of...”		95% CI for Mean Difference	<i>p</i>	<i>t</i>	<i>df</i>	<i>d</i>
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
51.36	10.2	47.43	8.50	-7.25, -0.62	0.02	-2.34	208	.42

Note. *M* = Mean. *SD* = Standard Deviation. CI = Confidence interval. *M* and *SD* of comprehension posttest. *d* = Cohen’s *d*.

These results do not argue for any type of causality; simply instructing students to make two predictions, for example, is unlikely to result in higher levels of comprehension. Rather, the data suggest that the ability to make these other types of inferences (i.e., predictions and the more analytic, “A-ha!” types of inferences) is associated with stronger comprehension, whereas text-to-self inferences are not. Certainly, the current study has limitations, described below, but this finding does suggest that a closer investigation of the impact of prominent text-to-self strategy instruction on comprehension is warranted.

Limitations

The results of this study are limited by three predominant issues: a small sample size, a single population, and issues with the text.

The study commenced with 261 total students, from which 11 students' records were subsequently eliminated due to various factors. Though a final *N* of 250 is within the range of many studies of classroom interventions involving reading comprehension, for the purposes of training a neural network, more cases will result in a more accurate model (Jain et al., 1996). The findings of the neural network analyses, therefore, are necessarily constrained, and must only be interpreted more generally.

Additionally, though the participants in this study are drawn from two separate schools, the two schools are in the same school district. Therefore, the external validity of the study is threatened by a potential interaction of setting and treatment (Creswell, 2009). To mitigate this threat, future studies will need to be repeated with students drawn from a different population.

Finally, issues exist with the text chosen for this study. As described in the "Text Selection" section of chapter 3, the story "Slower Than the Rest" was chosen because it was accessible, engaging, and it allowed students to generate multiple inferences. However, the elevated posttest scores revealed that the text was below reading level for a large portion of the students. Moreover, the text did not contain enough thematic elements which would have allowed for a sufficiently thorough and rigorous posttest. Discussions with the participating teachers suggested that the text used in the study needed to be written at a fifth-grade reading level. Future iterations of this study will have to commence with a text more closely calibrated to the students' grade level. Further, the

findings related to the different type of inferences need to be replicated with additional texts, to facilitate the generalizability of the results.

Implications

The present study suggests a number of implications for both reading comprehension research and for reading comprehension instruction.

One cluster of implications result from the novel data collection methodology used in this study. This research demonstrates that students are indeed capable of indicating their inference-making processes through highlighting and annotating texts. The extent to which the highlights and annotations represent the totality of the students' inference making is unclear, and will need further investigation in later studies. However, it is clear from an examination of the highlights and the comments associated with them that students are able to make their inferencing process tangible, at least in part.

One implication of this finding is that teachers may be able to deliver formative assessments in reading comprehension in a less intrusive way than more typical formative assessments, such as multiple choice questions, free response assessments, or think-alouds. This is significant, in that these more typical formative assessments have the potential to distract the reader from reading, thereby impacting the construct they are meant to measure. Particularly when used with out-of-class reading, teachers may be able to gather formative assessment information without having to sacrifice class time in the delivery of an in-class assessment.

Beyond the classroom, this data collection method has potential to impact reading comprehension research. Research on the inference-making process that occurs during reading typically takes place after the reading process has concluded, either in the form of a selected response posttest (e.g., Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 1999) or a prompted written response (e.g., Kintsch & Kintsch, 2005). Some assessments of inferencing can occur during the reading, but only by interrupting the reading and soliciting evidence of the students' thinking, usually through think aloud protocols (e.g., Chi, De Leeuw, Chiu, & Lavancher, 1994). In fact, Leslie and Caldwell (2009) found that the majority of think aloud assessments asked the participants to think out loud following each sentence read. It is possible that a methodology of identifying students' inference making as it occurs and without substantively interrupting the reading procedure could produce higher quality evidence.

One further implication that does not have to do with the data collection methodology relates to elaborative inferences and reading comprehension. As described above, the data indicate that students who predominantly make elaborative inferences show weaker comprehension of the text. Should this trend be replicated in additional studies, then it will be incumbent upon researchers to determine whether the reliance on elaborative inferences is symptomatic of broader comprehension issues, or whether this reliance stems from instructional factors.

Suggestions for Future Research

The goal of this study was to determine the nature of the relationship between patterns of students' highlights and annotations of a text and their comprehension of the text. In order to collect the data necessary to conduct this test, an e-reader was developed that captures students' highlights and annotations. Data indicate that patterns do indeed exist, and that the presence of predictive inferences is a particularly relevant part of the pattern. However, the current study did not uncover enough detail to generalize that conclusion. Therefore, the first step in continuing this line of research would be to replicate the basic structure of the current study, but with modifications that address the limitations of the study, as enumerated above. These limitations include increasing and diversifying the student population and using multiple texts that are more age appropriate.

Additionally, the research design may need to be modified further to include exploration of whether the highlighting and annotation process increases cognitive load to the point that it harms the students' comprehension. One potential way to measure this is to set up an experimental design in which one condition highlights and annotates the text, while the other does not. Each group's posttest scores could be compared, using scores on a general measure of reading comprehension administered prior to the intervention as a covariate.

If further testing continues to support the idea that a relationship exists between the pattern of highlighting and students' comprehension of a specific text, then three potential avenues of study could proceed. The first concerns the transformation of the e-reader into an intelligent reading support system. The e-reader could be programmed to monitor an individual student's patterns of highlighting and annotation. If the software recognizes that the pattern indicates that a low level of comprehension is occurring, then the software can intervene and provide support for the student in real time. These supports could take the form of metacognitive prompts that have been shown to be effective in supporting comprehension. If the software recognizes that the pattern matches that of strong comprehenders, then the software would not intervene. In this way, the e-reader would differ from other digital reading comprehension support software in that it would provide a level of support tailored the individual students' needs.

The second potential avenue of study that could proceed from this study involves how teachers could use data collected by the e-reader to support students. Teachers would need training in the use of this data, and a number of instructional responses based on the data could be selected for use and for subsequent study.

Finally, the current research was conceived to address the comprehension of fictional texts. However, because of its emphasis in the Common Core State Standard, a current trend in K-12 schools is towards informational texts. Informational texts are found in all of the content areas, unlike fiction which is predominantly the domain of

English/language arts classes. Students also can read informational texts differently than they do fiction (Duke & Roberts, 2010). Therefore new studies, perhaps using different types of highlighting types, would need to be performed on informational texts.

Summary

This study has demonstrated that the inferences students generate during reading fiction can be captured using a Web-based e-reader. Subsequent analysis indicated that posttest comprehension scores can be predicted using models generated from the data collected by the e-reader. Additionally, the data suggest that the presence of prediction inferences are more closely associated with higher levels of comprehension than other types of inferences. Conversely, the presence of elaborative inferences with minimal evidence of other types of inferencing is associated with reduced levels of comprehension.

The study's results are limited in three main ways. A larger sample size would allow for greater precision in the model, a more diverse group of participants would help to minimize interaction between treatment and location, and, finally, a more grade-level appropriate text would facilitate greater precision in the measurement.

The study has implications both in the methodology of reading research and also potentially on the practice of reading instruction. The methodology used in this study is a new way of identifying students' inferencing during their reading with minimal interference. It also suggests a new way that teachers can formatively assess students'

comprehension. The results of this investigation also suggest that the relationship between reading instructional strategies, elaborative inferencing, and comprehension needs further investigation. Future research can expand on this relationship. In addition, future research can explore the development of responsive comprehension supports, based on the students' use of the e-reader, and on teacher use of the data collected by the e-reader.

REFERENCES

- Allington, R. (2002). You can't learn much from books you can't read. *Educational Leadership*, 60(3), 16–19.
- Anderson-Inman, L., & Horney, M. (1997). Electronic books for secondary students. *Journal of Adolescent & Adult Literacy*, 40, 486–491.
- Anderson-Inman, L., & Horney, M. (2007). Supported eText: Assistive technology through text transformations. *Reading Research Quarterly*, 42(1), 153–160.
- Andreassen, R., & Bråten, I. (2009). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading*, 33(3), 263–283. doi:10.1111/j.1467-9817.2009.01413.x
- Barone, D. M., & Mallette, M. H. (2012). Literacy research that informs practice. *The Reading Teacher*, 66(1), 5–7. doi:10.1002/TRTR.01101
- Beck, I. L., McKeown, M., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, 96, 385–414.
- Bevan, N. (1995). Measuring usability as quality of use. *Software Quality Journal*, 4, 115–150.
- Biancarosa, G., & Snow, C. E. (2004). *Reading next—a vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York*. Washington, DC.
- Bluestein, N. A. (2010). Unlocking text features for determining importance in expository text: A strategy for struggling readers. *The Reading Teacher*, 63(7), 597–600. doi:10.1598/RT.63.7.7
- Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: what do tests of reading comprehension measure? *The British Journal of Educational Psychology*, 75(Pt 2), 189–201. doi:10.1348/000709904X22674

- Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*, 489–503.
- Cain, K., Oakhill, J., & Bryant, P. (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
- Calfee, R. (2009). Foreword. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. xi – xv). New York, NY: Routledge.
- Chall, J. S., & Jacobs, V. A. (2003). Poor children's fourth-grade slump. *American Educator, 27*(1), 14-17.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.
- Cheung, A. C. K., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*. doi:10.1016/j.edurev.2012.05.002
- Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477. doi:10.1207/s15516709cog1803_3
- Connell, C., Bayliss, L., & Farmer, W. (2012). Effects of eBook readers and tablet computers on reading comprehension. *International Journal of Instructional Media, 39*(2), 131–141.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Los Angeles, CA: Sage Publications.
- Dalton, B., & Proctor, C. P. (2008). The changing landscape of text and comprehension in the age of new literacies. In J. Coiri, M. Knobel, C. Lankshear, & D. Leu (Eds.), *Handbook of research on new literacies* (pp. 297–324). Mahwah, NJ: Lawrence Erlbaum.
- Dalton, B., Proctor, C. P., Uccelli, P., Mo, E., & Snow, C. E. (2011). Designing for diversity: The role of reading strategies and interactive vocabulary in a digital

- reading environment for fifth-grade monolingual english and bilingual students. *Journal of Literacy Research*, 43(1), 68–100. doi:10.1177/1086296X10397872
- Dalton, B., & Strangman, N. (2006). Improving struggling readers' comprehension through scaffolded hypertexts and other computer-based literacy programs. In M. C. McKenna, L. D. Labbo, R. D. Kieffer, & D. Reinking (Eds.), *International handbook of literacy and technology* (Vol. 2, pp. 75–92). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Daneman, M., & Carpenter, P. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4), 561–584.
- Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433.
- DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, 23(3), 1616–1641. doi:10.1016/j.chb.2005.08.012
- Dole, J., Duffy, G., Roehler, L., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, 61(2), 239–264.
- Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What research has to say about reading instruction* (pp. 205–242). Newark, DE: International Reading Association.
- Duke, N. K., & Roberts, K. L. (2010). The genre-specific nature of reading comprehension. In D. Wyse, R. Andrews, & J. Hoffman (Eds.), *The Routledge international handbook of English, language and literacy teaching* (pp. 74–86). New York, NY: Routledge.
- Dunlosky, J., Rawson, K. a., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: promising

- directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. doi:10.1177/1529100612453266
- Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14(4), 481–533.
- 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. doi:10.3102/0034654308325998
- Engle, R., Carullo, J., & Collins, K. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, 84(5), 253–262.
- Faber, J., Morris, J., & Lieberman, M. (2000). The effect of note taking on ninth grade students' comprehension. *Reading Psychology*, 21(3), 257–270.
- Fact Sheet - Johnson Middle School. (2013). Retrieved November 20, 2013, from <http://www2.k12albemarle.org/school/WMS/about/Pages/fact-sheet.aspx>
- Fact Sheet - Roland Ryans Middle School. (2013). Retrieved November 20, 2013, from <http://www2.k12albemarle.org/school/JMS/about/Pages/fact-sheet.aspx>
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage Publications.
- Fisher, D., & Frey, N. (2012). Close reading in elementary schools. *The Reading Teacher*, 66(3), 179–188. doi:10.1002/TRTR.01117
- Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights from PISA 2009: Performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context (NCES 2011-004)*. Washington, DC.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323–330. doi:10.1207/s1532799xssr1003_7
- Fry, E. (2002). Readability versus leveling. *The Reading Teacher*, 56(3), 286–291.

- Gellert, A. S., & Elbro, C. (2012). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment, 31*(1), 16–28. doi:10.1177/0734282912451971
- Gershenson, C. (2003). Artificial Neural Networks for Beginners.
- Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research, 71*(2), 279–320.
- Gordon, C. J., & Pearson, P. D. (1983). *The effects of instruction in metacomprehension and inferencing on children's comprehension abilities (Tech. Rep. No. 277)*. Urbana, IL.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–95.
- Green, T. D., Perera, R. A., Dance, L. A., & Myers, E. A. (2010). Impact of presentation mode on recall of written text and numerical information: Hard copy versus electronic. *North American Journal of Psychology, 12*(2), 233–242.
- Grimshaw, S., Dungworth, N., McKnight, C., & Morris, A. (2007). Electronic books: Children's reading and comprehension. *British Journal of Educational Technology, 38*(4), 583–599. doi:10.1111/j.1467-8535.2006.00640.x
- Guthrie, J., Seifert, M., Burnham, N., & Caplan, R. (1974). The maze technique to assess, monitor reading comprehension. *The Reading Teacher, 28*(2), 161–168.
- Hebert, M., Simpson, A., & Graham, S. (2012). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing, 26*(1), 111–138. doi:10.1007/s11145-012-9386-3
- Hegel, G. W. F. (1977). *Phenomenology of spirit*. Oxford, UK: Oxford University Press.
- Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., ... Verplank, W. (1996). ACM SIGCHI curricula for human-computer interaction. Retrieved November 20, 2013, from [http:// old.sigchi.org/sigchi/cdg/](http://old.sigchi.org/sigchi/cdg/)

- International Reading Association. (2012). *Adolescent literacy (Position Statement)*. Newark, DE.
- ISO. (1998). ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31–44.
- Jokela, T., Iivari, N., Matero, J., & Karukka, M. (2003). The standard of user-centered design and the standard definition of usability: Analyzing ISO 13407 against ISO 9241-11. In *Proceedings of the Latin American conference on Human-Computer Interaction* (pp. 53–60). ACM.
- Kamil, M. L. (2003). *Adolescents and literacy: Reading for the 21st century*. Washington, DC: Alliance for Excellent Education.
- Kamil, M. L., & Chou, H. K. (2008). Comprehension and Computer Technology. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 289–304). New York, NY: Routledge.
- Kang, Y.-Y., Wang, M.-J. J., & Lin, R. (2009). Usability evaluation of E-books. *Displays*, 30(2), 49–52. doi:10.1016/j.displa.2008.12.002
- Kardash, C. A., Royer, J. M., & Greene, B. A. (1988). Effects of schemata on both encoding and retrieval of information from prose. *Journal of Educational Psychology*, 80(3), 324–329. doi:10.1037//0022-0663.80.3.324
- Keene, E. O., & Zimmerman, S. (1997). *Mosaic of thought*. Portsmouth, NH: Heinemann.
- Kincaid, J., Fishburne, R., Rogers, R., & Chissom, B. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel (No. RBR-8-75). *NAVAL TECHNICAL TRAINING COMMAND MILLINGTON TN RESEARCH BRANCH*. Retrieved from <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA006655>

- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.
- Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270–1329). Newark, DE: International Reading Association.
- Kintsch, W., & Dijk, T. Van. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.
- Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71–92). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lahtinen, V., Lonka, K., & Lindbiom-Ylänne, S. (1997). Spontaneous study strategies and the quality of knowledge construction. *British Journal of Educational Psychology*, 67, 13–24.
- Lennon, C., & Burdick, H. (2004). The lexile framework as an approach for reading measurement and success. Retrieved from https://lexile.com/m/uploads/whitepapers/Lexile-Reading-Measurement-and-Success-0504_MetaMetricsWhitepaper.pdf
- Leslie, L., & Caldwell, J. (2009). Formal and informal measures of reading comprehension. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension*. New York, NY: Routledge.
- Lomicka, L. (1998). “To gloss or not to gloss”: An investigation of reading comprehension online. *Language Learning & Technology*, 1(2), 41–50.
- Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. (2007). A multidimensional framework to evaluate reading assessment tools. In D. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–136). New York, NY: Erlbaum.
- Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology*, 47(5), 315–35. doi:10.1016/j.jsp.2009.04.003

- Margolin, S., & Driscoll, C. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied Cognitive Psychology*, 27(4), 512–519. doi:10.1002/acp.2930
- McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: Results of a US survey. *Reading Research Quarterly*, 47(3), 283–306. doi:10.1002/RRQ.021
- McKenna, M. C., Kear, D., & Ellsworth, R. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly*, 30(4), 934–956.
- McKenna, M. C., & Stahl, K. D. (2009). *Assessment for reading instruction* (2nd ed.). New York, NY: Guilford Press.
- McKeown, M., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44(3), 218–253.
- McNamara, D., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education*, 4(1), 33–46.
- McNamara, D., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In B. Ross (Ed.), *Psychology of learning and motivation, Vol. 51* (Vol. 51, pp. 297–384). Burlington: Academic Press. doi:10.1016/S0079-7421(09)51009-2
- McNamara, D., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34(2), 147–171. doi:10.2190/1RU5-HDTJ-A5C8-JVWE
- Mealer, C., Morgan, P., & Williams, C. (2011). *eReaders in Higher Education. University of Alabama*. Tuscaloosa, Alabama. Retrieved from <http://www.bama.ua.edu/~ckmealer/MealerEport/EReaders.pdf>
- Meyer, J. P. (2014). jMetrik version 3.1. Retrieved from www.itemanalysis.com

- Moran, J., Ferdig, R., Pearson, P. D., Wardrop, J., & Blomeyer, R. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research*, 40(1), 6–58. doi:10.1080/10862960802070483
- Nation, K., Cocksey, J., Taylor, J. S. H., & Bishop, D. V. M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 51(9), 1031–9. doi:10.1111/j.1469-7610.2010.02254.x
- National Center for Education Statistics (NCES). (2013a). *Reading framework for the 2013 National Assessment of Educational Progress*. Washington, DC.
- National Center for Education Statistics (NCES). (2013b). *The Nation's Report Card: Reading 2013*. Washington, DC.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010a). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010b). *English Language Arts and Literacy in History / Social Studies, Science, and Technical Subjects: Appendix A - Research Supporting Key Elements of the Standards*.
- Ness, M. (2009). Reading comprehension strategies in secondary content area classrooms: teacher use of and attitudes towards reading comprehension instruction. *Reading Horizons*, 49(2), 143–166.
- NICHD. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC. Retrieved from <http://www.nationalreadingpanel.org>
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Conference Companion on Human Factors in Computing Systems - CHI '94*, 152–158. doi:10.1145/259963.260333

- Nielsen, J. (1995). 10 Usability Heuristics for User Interface Design. Retrieved November 20, 2013, from <http://www.nngroup.com/articles/ten-usability-heuristics/>
- Norman, D. (2002). *The design of everyday things*. New York, NY: Basic Books.
- Oakhill, J. V., & Cain, K. (2003). The development of comprehension skills. In *Handbook of children's literacy* (pp. 155–180). Dordrecht, The Netherlands: KluwerAcademic Publishers.
- Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175.
- Paolucci, R. (1998). The effects of cognitive style and knowledge structure on performance using a hypermedia learning system. *Journal of Educational Multimedia & Hypermedia*, 7(2-3).
- Paris, S. G., & Hamilton, E. E. (2008). The development of children's reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 32–53). New York, NY: Routledge.
- Pearson, P. D., & Hamm, D. (2005). The assessment of reading comprehension: A review of practices—past, present, and future. In S. G. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–69). Mahwah, NJ: Erlbaum.
- Pearson, P. D., & Hiebert, E. H. (2010). National reports in literacy: Building a scientific base for practice and policy. *Educational Researcher*, 39(4), 286–294. doi:10.3102/0013189X10370205
- Petscher, Y. (2010). A meta-analysis of the relationship between student attitudes towards reading and achievement in reading. *Journal of Research in Reading*, 33(4), 335–355. doi:10.1111/j.1467-9817.2009.01418.x
- Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, 19(3), 291–312. doi:10.1002/acp.1086
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Pittsburgh, PA.

- Rankin, E., & Culhane, J. (1967). Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 13(3), 193–198.
- Ray, R., & Belden, N. (2010). Teaching college level content and reading comprehension skills simultaneously via an artificially intelligent adaptive computerized instructional system. *The Psychological Record*, 57, 201–218.
- Reardon, S. F., Valentino, R. A., & Shores, K. A. (2012). Patterns of literacy among U.S. students. *The Future of Children*, 22(2), 17–37. doi:10.1353/foc.2012.0015
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530.
- Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, 66(2), 181–221.
- Royer, J. M. (2001). Developing reading and listening comprehension tests based on the Sentence Verification Technique (SVT). *Journal of Adolescent & Adult Literacy*, 45(1), 30–41.
- Royer, J. M. (2005). *Uses for the sentence verification technique for measuring language comprehension*. Unpublished manuscript, University of Massachusetts, Amherst, Amherst, MA.
- Royer, J. M., Greene, B., & Sinatra, G. (1987). The sentence verification technique: A practical procedure for testing comprehension. *Journal of Reading*, 30(5), 414–422.
- Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Literacy Research*, 11(4), 355–363. doi:10.1080/10862967909547341
- Royer, J. M., Sinatra, G., Greene, B., & Tirre, W. (1989). Assessment of on-line comprehension of computer-presented text. *The Journal of Educational Research*, 82(6), 348–355.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441–474. doi:10.1191/0265532206lt337oa

- Rylant, C. (1985). *Every living thing*. New York, NY: Aladdin Paperbacks.
- Sainsbury, M., & Schagen, I. (2004). Attitudes to reading at ages nine and eleven. *Journal of Research in Reading*, 27(4), 373–386. doi:10.1111/j.1467-9817.2004.00240.x
- Salomon, G., Globerson, T., & Guterman, E. (1989). The computer as a zone of proximal development: Internalizing reading-related metacognitions from a reading partner. *Journal of Educational Psychology*, 81(4), 620–627. doi:10.1037/0022-0663
- Schugar, J., Schugar, H., & Penny, C. (2011). A Nook or a book? Comparing college students' reading comprehension levels, critical reading, and study skills. *International Journal of Technology in Teaching and Learning*, 7(2), 174–192.
- Seigneuric, A., & Ehrlich, M.-F. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing*, 18(7-9), 617–656. doi:10.1007/s11145-005-2038-0
- Shanahan, T., Kamil, M., & Tobin, A. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, 17(2), 229–255.
- Shin, E., Schallert, D., & Savenye, W. (1994). Effects of learner control, advisement, and prior knowledge on young students' learning in a hypertext environment. *Educational Technology Research and Development*, 42(1), 33–46.
- Slotte, V., & Lonka, K. (1999). Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology*, 24(1), 1–20.
- Smith, F. (2004). *Understanding reading: A psycholinguistic analysis of reading and learning to read* (6th ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Snow, C. E. (2003). Assessment of reading comprehension: Researchers and practitioners helping themselves and each other. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 192–206). New York, NY: Guilford Press.
- Specht, D. F. (1991). A general regression neural network. *Transactions on Neural Networks*, 2(6), 568–576.

- Stahl, K. D. (2009). Assessing the comprehension of young children. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 428–448). New York, NY: Routledge.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360–407. doi:10.1598/RRQ.21.4.1
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Swanson, H. L. (2003). Age-related differences in learning disabled and skilled readers' working memory. *Journal of Experimental Child Psychology*, 85(1), 1–31. doi:10.1016/S0022-0965(03)00043-2
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415–433.
- Thompson, S., Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from PIRLS 2011: Reading Achievement of U.S. Fourth-Grade Students in an International Context (NCES 2013-010)*. Washington, DC.
- Thorndike, R., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Pearson.
- Urquhart, I. (2002). Beyond the literal: Deferential or inferential reading. *English in Education*, 36(2), 18–30. doi:10.1111/j.1754-8845.2002.tb00758.x
- Van den Broek, P. (2012). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J. Sabatini, E. Albro, & T. P. O'Reilly (Eds.), *Measuring up: Advances in How We Assess Reading Ability* (pp. 39–58). Lanham, MD: Rowman & Littlefield.

- Van den Broek, P., & Kendeou, P. (2011). Preparing for reading comprehension: Fostering text comprehension skills in preschool and early elementary school children. *International Electronic Journal of Elementary Education*, 4(1), 259–268.
- Van den Broek, P., Tzeng, Y., Risen, K., & Trabasso, T. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*, 93(3), 521–529. doi:10.1037//0022-0663.93.3.521
- Wharton-McDonald, R., & Swiger, S. (2008). Developing higher-order comprehension in the middle grades. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 510–530). New York, NY: Routledge.
- Willingham, D. (2006). The usefulness of brief instruction in reading comprehension strategies. *American Educator*, 25(1), 39–50.
- Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, 2, 33–45.
- Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, UK: Cambridge University Press.
- Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, 24(3), 875–887. doi:10.1016/j.chb.2007.02.015

APPENDIX A: TABLES AND FIGURES

Figure 1. The ICON interface (Dalton et al., 2011).

CAST Folktales **How Coyote Stole Fire**

home | glossary | strategy help | resources | worklog | 0 1 2 3 4 5 6 7 | activities | logout

Look at the pictures and highlighted words to help you visualize what is happening. Describe what you see in your mind.

en español

Tek Pat Digs

Type your response below:

I see cold, sad humans. They are hungry because they have no fire. Coyote's golden fur keeps him warm. He is watching the humans.

record save

One afternoon, as Coyote was hunting in the newly fallen snow, he came upon a family huddled in a circle. They were singing a sad song about their lost loved ones, who had died in the ice-cold winter blizzards. The sorrow in their voices prickled the fur on Coyote's neck. Coyote saw how their mournful tears froze on their cheeks, and how their icy breath hung in the air. Coyote, like many of the animals, had no need for the warmth of fire, because he had a thick winter coat. But he felt sorry for the Humans, and decided to help them find fire.

Coyote knew of the Fire Beings, who lived on a nearby mountaintop. The Fire Beings were not known for their hospitality. They kept fire to themselves because they were selfish and afraid to share it with the Humans. But Coyote, known as the Trickster of the animals, knew there must be a way to steal fire from the Fire Beings. So he hurried off to the mountaintop to watch them.

Now is a good time to stop and think about the story

2 of 7

CAST Folktales **How Coyote Stole Fire**

home | glossary | strategy help | resources | worklog | 0 1 2 3 4 5 6 7 | activities | logout

wade in

Wade In to each power word to explore its meaning. Read it, view it, hear it, connect it to make this word your own, and then learn more about the word with a Language Alert. Click on any word to begin!

en español

✓ anxiously ✓ bitter
✓ dense ✓ menacing
✓ relieved

anxiously
Worried about a certain matter.

Example sentence:
Before the test, Maria looked around anxiously.

La palabra en español es:
ansiosamente

Sports fans anxiously watching a game

Connect It Make a personal connection to the power word using your own experience, knowledge and feelings. Type or record your response and save to worklog.

en español

Pat

hear-it

0:02

connect-it

I get really anxious right before I swim in a race.

Figure 2. The RAND Reading Study Group's (2002) paradigm for reading comprehension.

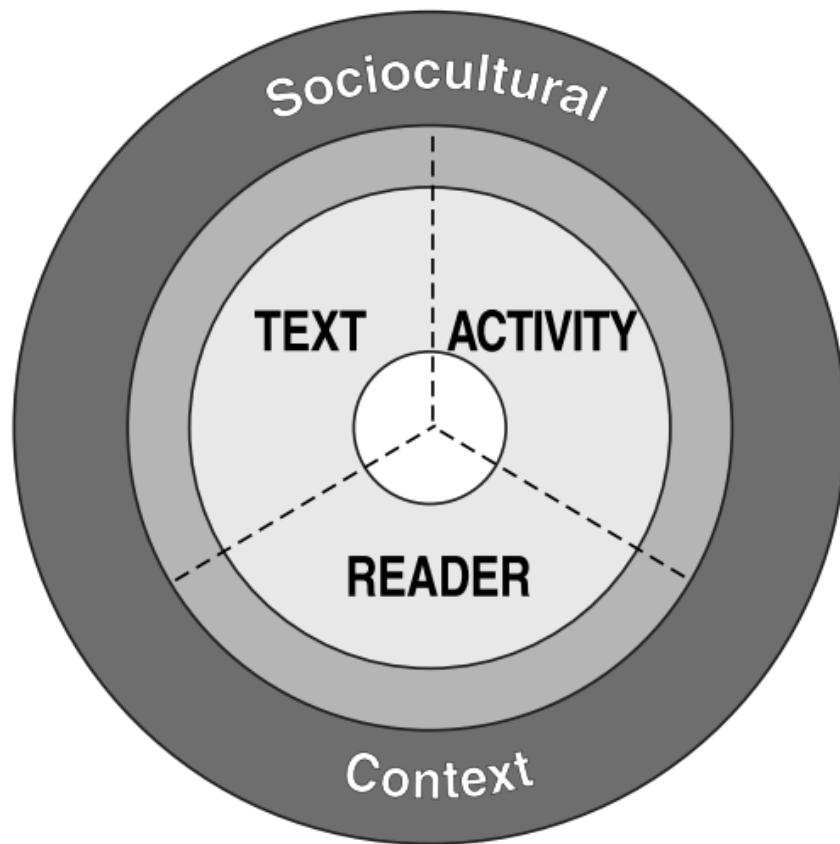


Figure 3. The e-reader developed for use in this study.

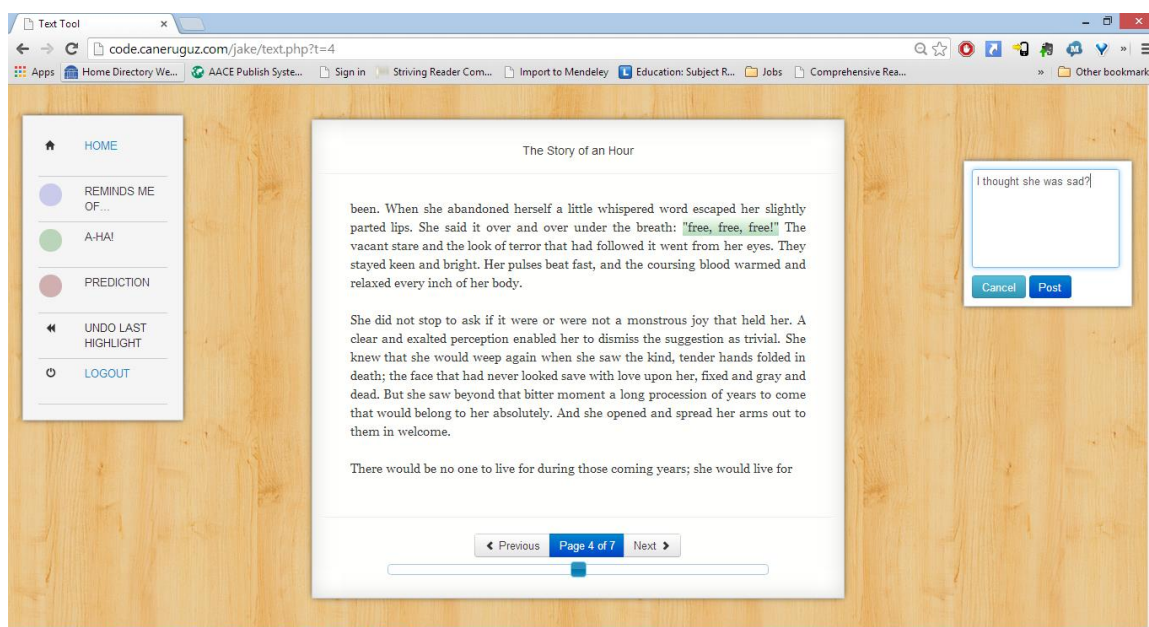
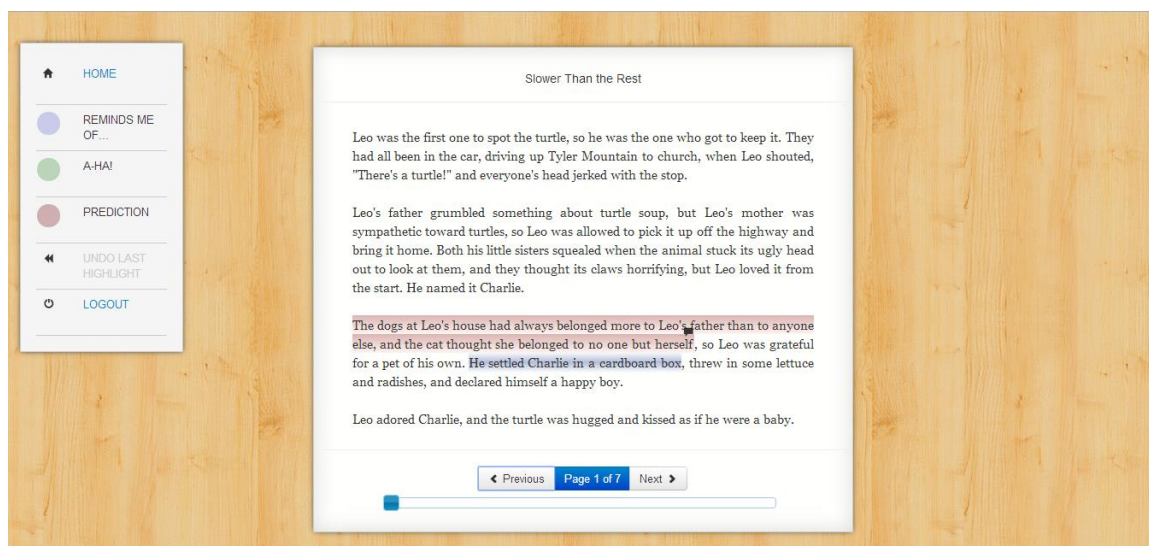


Figure 5. A sample of an SVT assessment.

For each sentence, circle “Yes” if the sentence means the same thing as a sentence in the story. Circle “No” if the sentence has a different meaning than a sentence in the story.

YES	NO	Leo was slow in reading, slow in numbers, slow in understanding nearly everything that passed before him in a classroom. (O)
YES	NO	Charlie had a tendency to bite people when he was angry, which was often. (D)
YES	NO	Charlie was the friendliest turtle anyone had ever seen. (O)
YES	NO	Leo would come home from school and talk to Charlie, but he never let Charlie out of his cage. (MC)
YES	NO	Leo even made a leash and collar so that he could take Charlie out for walks in his neighborhood, but Charlie seemed to hate it. (D)
YES	NO	Leo was a poor reader, bad at math, and he had a hard time understanding most everything that happened in class. (P)
YES	NO	Leo took Charlie on backyard exploring trips, during which he would tell the turtle all about the day’s news. (P)
YES	NO	Charlie was a grumpy turtle. (MC)
YES	NO	Leo was a slower runner than all of his classmates (D)
YES	NO	Leo was slow in reading, and slow in numbers, but he understood nearly everything that his teacher covered in class. (MC)
YES	NO	Every day, Leo came home from school, took Charlie to the backyard to let him explore and told him about the things that had happened in fifth grade. (O)
YES	NO	Charlie was nicest turtle in the history of turtles. (P)

O = original, P = paraphrase, MC = meaning change, D = distractor. Labels will not be present on the actual test.

Figure 6. Self-identified ethnicities of participants.

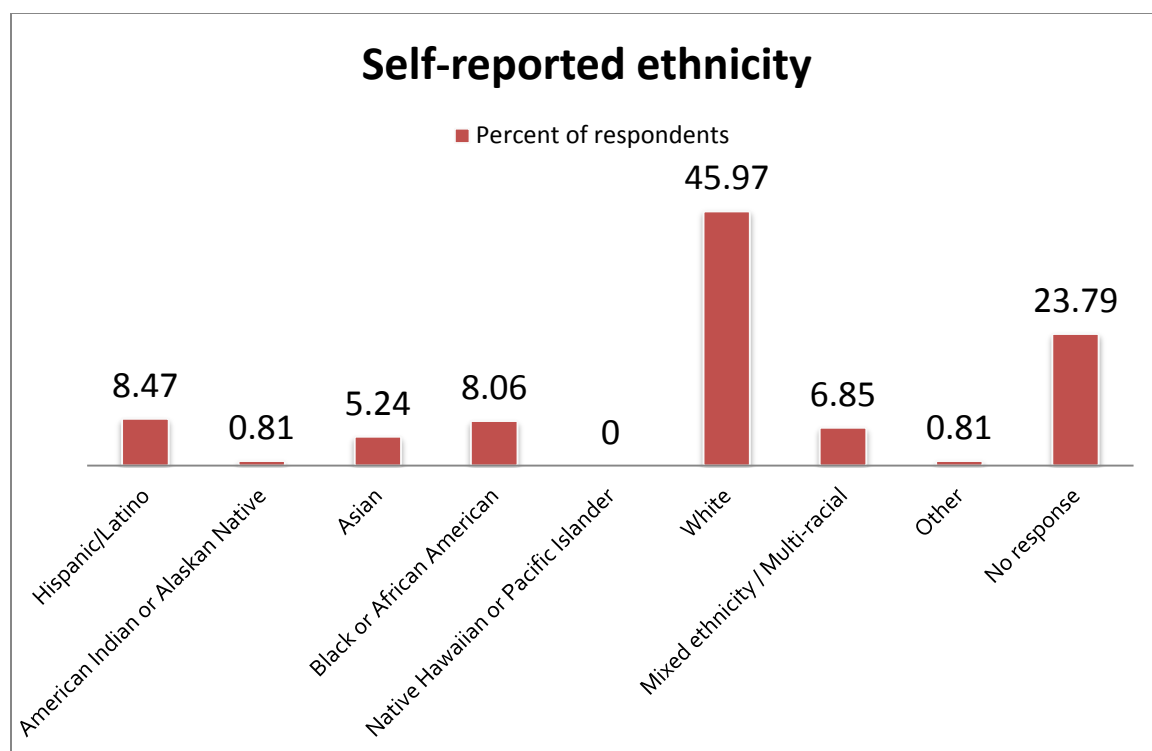


Figure 7. Frequency of highlights per page.

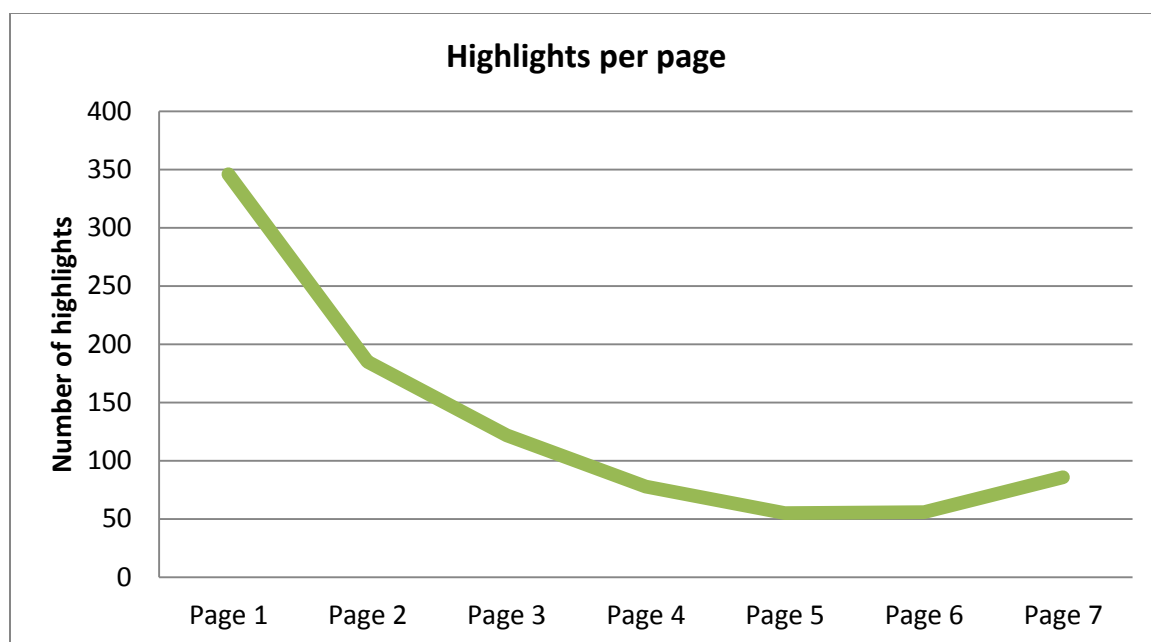


Table 1. Typical Lexiles reader measures, by grade.

Grade Band	Current Lexile Band	"Stretch" Lexile Band*
K-1	N/A	N/A
2-3	450L-725L	420L-820L
4-5	645L-845L	740L-1010L
6-8	860L-1010L	925L-1185L
9-10	960L-1115L	1050L-1335L
11-CCR	1070L-1220L	1185L-1385L

Retrieved from:

<http://www.lexile.com/using-lexile/lexile-measures-and-the-ccssi/text-complexity-grade-bands-and-lexile-ranges/>

APPENDIX B: TEST BLUEPRINT

Process Objective	Content Areas		
	A. Introductory segment, 40%	B. Leo and Charlie at school, 40%	C. Awards ceremony, 20%
1. Forms a coherent textbase.	Leo finding Charlie Leo's family Leo's feelings toward Charlie	Leo's reason for bringing Charlie to school The other presentations impact on Leo Leo's presentation style	Leo's mindset before he receives his award Leo's mindset after he receives his award Leo's award The impact of the award on Leo
50%	6 questions	6 questions	3 questions
2. Forms a situation model grounded in the text.	Leo's general affect Leo's mental abilities	The class's mood during Leo's presentation The change in Leo	Leo's mood before he receives his award Leo's mood after he receives his award
50%	6 questions	6 questions	3 questions
Number of items	12	12	6
Total time for test—30 minutes			Total number of items—30

APPENDIX C: ITEM WRITING GUIDELINES

General item writing guidelines:

1. Use concise, grammatically correct sentences.
2. Use simple statements, focusing on only one part of the construct at a time.
3. Number the items.
4. List multiple choice responses vertically.
5. Include concise directions, written at the appropriate reading level.
6. Do not change types of questions without including a new set of directions.
7. Avoid double negatives.
8. Group all question types together.
9. Do not tie one item to another—i.e., it should not be necessary to complete one item to be able to respond to another.

Specific item writing guidelines:

1. Use 3 choices per multiple choice question.
2. Arrange for the first half of SVT sentences to come from first half of the story.
3. Keep reading level of stems, responses, and directions centered around 5th grade level (i.e., two years below the level of the population).
4. Shoot for 15-20 multiple choice items to be included in the final, so write 30 for review.
5. Shoot for 16 sentences (64 test sentences) for the SVT section.
6. Items should not occur in the same order as the plot of the text occurs.

APPENDIX D: INFERENTIAL ANALYSIS OF “SLOWER THAN THE REST”

c = coherence-type inference

e = elaborative-type inference

Leo was the first one to spot the turtle, so he was the one who got to keep it. They had all been in the car, driving up Tyler Mountain to church, when Leo shouted, "There's a turtle!" and everyone's head jerked with the stop.

Leo's father grumbled something about turtle soup //(e – Leo's father's personality), but Leo's mother was sympathetic toward turtles, so Leo was allowed to pick it up off the highway and bring it home. Both his little sisters squealed when the animal stuck its ugly head out to look at them, and they thought its claws horrifying, but Leo loved it from the start. He named it Charlie.

The dogs at Leo's house had always belonged more to Leo's father than to anyone else, and the cat thought she belonged to no one but herself, so Leo was grateful for a pet of his own. He settled Charlie in a cardboard box, threw in some lettuce and radishes, and declared himself a happy boy //(c – happy because of the pet).

Leo adored Charlie, and the turtle was hugged and kissed as if he were a baby. Leo liked to fit Charlie's shell on his shoulder under his left ear, just as one might carry a cat, and Charlie would poke his head into Leo's neck now and then to keep them both entertained.

Leo was ten years old the year he found Charlie. He hadn't many friends because he was slower than the rest. That was the way his father said it: "Slower than the rest." Leo was slow in reading, slow in numbers, slow in understanding nearly everything that passed before him in a classroom. As a result, in fourth grade Leo had been separated from the rest of his classmates and placed in a room with other children who were as slow as he. Leo thought he would never get over it. He saw no way to be happy after that //(c – being designated “slow” led to unhappiness).

But Charlie took care of Leo's happiness, and he did it by being congenial. Charlie was the friendliest turtle anyone had ever seen. The turtle's head was always stretched out, moving left to right, trying to see what was in the world //(c – head out equals outgoing) (e – unusual turtle behavior; Charlie was exceptional). His front and back legs moved as though he were swimming frantically in a deep sea to save himself, when all that was happening was that someone was holding him in midair. Put Charlie down and he would

sniff at the air a moment, then take off as if no one had ever told him how slow he was supposed to be // (c – “slow” used to describe both Charlie and Leo).

Every day, Leo came home from school, took Charlie to the backyard to let him explore and told him about the things that had happened in fifth grade // (c – Leo has had Charlie for a year now). Leo wasn't sure how old Charlie was, and, though he guessed Charlie was probably a young turtle, the lines around Charlie's forehead and eyes and the clamp of his mouth made Leo think Charlie was wise the way old people are wise // (e – archetype of elderly as wise). So Leo talked to him privately every day.

Then one day Leo decided to take Charlie to school.

It was Prevent Forest Fires week and the whole school was making posters, watching nature films, imitating Smokey the Bear // (e – the students were preparing projects on forest fires). Each member of Leo's class was assigned to give a report on Friday dealing with forests. So Leo brought Charlie.

Leo was quiet about it on the bus to school // (e – Leo did not want it to be known he had a turtle. It is supposed to be a surprise). He held the covered box tightly on his lap, secretly relieved // (e – Leo really did not want it to be known he had a turtle on the bus) that turtles are quiet except for an occasional hiss. Charlie rarely hissed in the morning; he was a turtle who like to sleep in.

Leo carried the box to his classroom and placed it on the wide windowsill near the radiator and beside the geraniums. His teacher called attendance and the day began.

In the middle of the morning, the forest reports began. One girl held up a poster board pasted with pictures of raccoons, and squirrels, rabbits and deer, and she explained that animals died in forest fires. The pictures were too small for anyone to see from his desk. Leo was bored // (e – the reports lacked appeal to Leo).

One boy stood up and mumbled something about burnt-up trees. Then another got up and said if there were no forests, then his dad couldn't go hunting, and Leo couldn't see the connection in that at all // (c Leo is still bored).

Finally it was his turn. He quietly walked over to the windowsill and picked up the box. He set it on the teacher's desk.

"When somebody throws a match into a forest," Leo began, "he is a murderer. He kills trees and birds and animals. Some animals, like deer, are fast runners and they might

escape. But other animals"-he lifted the cover off the box-"have no hope. They are too slow. They will die." He lifted Charlie out of the box. "It isn't fair," he said, as the class gasped and giggled at what they saw // (e – the class was surprised by the turtle's appearance, and reacted the way typical fifth graders would). "It isn't fair for the slow ones."

Leo said much more // (e – about forest fires). Mostly he talked about Charlie, explained what turtles were like, the things they enjoyed, and what talents they possessed. He talked about Charlie the turtle and Charlie the friend, and what he said and how he said it made everyone in the class love turtles and hate forest fires. Leo's teacher had tears in her eyes // (c – Leo's teacher was made emotional by Leo's presentation) (e – this is unusual/unexpected behavior from Leo).

That afternoon, the whole school assembled in the gymnasium to bring the special week to a close. A ranger in uniform made a speech, then someone dressed up like Smokey the Bear danced with two others dressed like squirrels // (c – related to forest fire week). Leo sat with his box and wondered if he should laugh at the dancers with everyone else. He didn't feel like it // (c – Leo is still emotional from his presentation).

Finally, the school principal stood up and began a long talk. Leo's thoughts drifted off. He thought about being home, lying in this bed and drawing pictures, while Charlie hobbled all about the room. He did not hear when someone whispered his name // (c – Leo was daydreaming). Then he jumped when he heard, "Leo! It's you!" in his ear. The boy next to him was pushing him, making him get up.

"What?" Leo asked, looking around in confusion.

"You won!" they were all saying. "Go on!"

Leo was pushed onto the floor. He saw the principal smiling at him, beckoning to him across the room // (e – something good is about to happen to Leo). Leo's legs moved like Charlie's-quickly and forward.

Leo carried the box tightly against his chest // (e – Leo felt protective of Charlie and/or comforted by Charlie). He shook the principal's hand. He put down the box to accept the award plaque being handed to him. It was for his presentation with Charlie. Leo had won an award for the first time in his life, and as he shook the principal's hand and blushed and said his thank-you's, he thought his heart would explode with happiness.

That night, alone in his room, holding Charlie on his shoulder, Leo felt proud. And for the first time in a long time, Leo felt *fast*.