WALKER, RONDA, Ph.D. The Effect of Response Format and Presentation Conditions on Comprehension Assessments for Students with and without a Reading Disability. (2017)
Directed by Alan G. Kamhi, 52 pp.

Previous studies (Collins, 2015; Kennan & Meenan, 2014) have shown how variations in text and task factors and individual reader skills affect performance on reading comprehension assessments. The present study examined whether different presentation conditions (silent reading, watching a video) and response formats (open-ended vs. multiple-choice questions) influenced comprehension performance for students with and without reading disabilities. In addition, measures of word-level reading, vocabulary, working memory, listening comprehension, and prior knowledge were also assessed to determine the best predictors of performance on comprehension assessments.

Participants were 32 fifth grade students, 17 with reading disabilities (RD) and 15 typically developing (TD) students. All students were initially administered measures of word-level reading, vocabulary, listening comprehension, working memory, and decoding. Students were then administered four passages. Two of the passages were read silently and two were presented with videos. For each condition (text and video), comprehension was assessed with open-ended and multiple-choice questions. All assessments were administered individually to each student across two 60-minute testing sessions.

All students were found to perform significantly better on the multiple-choice questions than the open-ended questions. As expected, the TD group had significantly higher comprehension scores on all measures. Presentation condition did not

significantly affect performance for either group. Listening comprehension, working memory, and prior knowledge contributed unique variance to performance on the different response formats. For the open-ended questions, 67% of the variance was explained by the measures of listening comprehension and prior knowledge. In contrast, only 38% of the variance was explained by working memory for the multiple-choice questions.

Even though students performed better on the multiple-choice questions, the regression analyses indicated that the open-ended questions were better reflections of basic language abilities and prior knowledge. Open-ended questions appear to provide a better measure of reader and text factors than multiple-choice questions which are more influenced by task factors. Future studies should continue to examine how reader, text and task factors influence comprehension performance.

THE EFFECT OF RESPONSE FORMAT AND PRESENTATION CONDITIONS ON COMPREHENSION ASSESSMENTS FOR STUDENTS WITH AND WITHOUT A READING DISABILITY

by

Ronda Walker

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2017

Approved by

_____
Committee Chair

To W. Lloyd & Mary Jo Walker and Kesha Harris

APPROVAL PAGE

This dissertation written by Ronda Walker has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Alan Kamhi for his guidance, patience, and unwavering support. I would also like to thank Dr. Pam Williamson for her constructive suggestions during the planning of this research project. My grateful thanks are also extended to Dr. Robert Mayo and Dr. Jill Chouinard for their time and support. The statistics assistance provided by Jeffrey Labban was greatly appreciated. A big thank you goes to the parents and students who gave of their time to participate in this research project. Finally, thank you to my family, specifically my parents and my sister. This degree and dissertation is truly a family earned process. I am eternally grateful for your love, encouragement, and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

STATEMENT OF THE PROBLEM

The assessment of reading has been researched and analyzed for years with many variations and revisions of assessment being based on a variety trends and philosophies that happened to dominate the educational conversation at the time (Pearson & Hamm, 2005). Through all of the educational changes and reforms that have occurred, there has been a constant effort to create authentic performance assessments that support challenging curriculum and teaching (Pearson, Valencia, & Wixson, 2014). Scholars and educators have consistently researched the best way to create assessments that effectively assess students' reading comprehension ability. In recent years, the focus on reading comprehension assessment has become more pronounced as new educational reforms and policy debates on school curriculum, teacher instruction, and high stakes accountability testing have come into focus (Pearson et al., 2014; Valencia, Wixson, & Pearson, 2014).

Recent educational policies brought attention to reading comprehension of complex text and the need for rigorous assessments of reading comprehension. Specifically, 46 states and the District of Columbia have adopted the Common Core State Standards (CCSS). The CCSS stresses the importance of reading comprehension in all subjects and requires students to engage in reading more complex texts throughout their academic careers (National Governors Association Center for Best Practice, Council of Chief State School Officers, 2010). Under the CCSS, teachers are required to challenge

their students with more text-based question and answer sessions and provide the students with opportunities to have rigorous conversations and writing tasks that depend on assigned text. However, it is not only the CCSS that is leading to more rigorous curriculum and assessment. Schools across the country are focusing on reforming curriculum to introduce higher order thinking and deeper learning (Pearson et al., 2014). These curricula stress the importance on students comprehending complex text. Because of these new policies, and coupled with high stakes assessment, more focus is being placed on the reader skills that students need and text complexity in assessment (Valencia et al., 2014).

Some studies have questioned whether reading comprehension tests measure the actual comprehension of text, or other, underlying skills needed to complete the assessment (Keenan, Betjemann, & Olson, 2008). Skills that are needed to comprehend text include but are not limited to word-level reading, word recognition, working memory, and attention (RAND Reading Study Group, 2002). In addition, readers also need to be able to build upon previously learned information, access prior knowledge, and make inferences in order to fully understand the text (Clark & Kamhi, 2014). The effect on these skills can vary depending on which reading assessment a student is being asked to complete (Collins, 2015; Keenan et al., 2008). The complexity of reading comprehension makes it difficult to measure and properly assess comprehension (Keenan & Meenan, 2014). The increased emphasis on text complexity makes assessment of comprehension more difficult (Valencia et al., 2014).

In addition to text complexity, there has been a specific focus on how tasks can affect the perceived complexity of text (Pearson et al., 2014; Valencia et al., 2014). In the past, many studies have examined ways to improve students' abilities to comprehend text. More recent studies have analyzed how perceived reading comprehension ability can vary based on the assessment being used (Kennan & Meenan, 2014). Specifically, studies have shown that certain dimensions of assessment, such as response format and text length, can alter student performance (Collins, 2015; Francis, Fletcher, Catts, & Tomblin, 2005). There have been a few studies that have investigated different assessment dimensions (Collins, 2015; Francis et al., 2005; Keenan et al., 2008), but more research is needed on this topic.

Consideration must also be given to reader skills. Individual reading skills and differences in assessment dimensions (i.e., response format) can cause children to test as a poor reader on some comprehension assessments while testing within normal limits on others (Francis et al., 2005; Keenan & Meenan, 2014). This can lead to a misdiagnosis of a reading problem where there is none. It may also affect long term reading interventions that are used to facilitate and improve reading comprehension. It is important to investigate how individual reading skills are assessed through different response formats and how dimensions of comprehension assessments may cause performance variations.

Another factor to consider is how information is presented when measuring comprehension. Comprehension is the ultimate goal of reading and the culmination of all reading acts. It is a coordination of cognitive, linguistic, and text specific processes and involves decoding, word recognition, and understanding of vocabulary among many

other skills (Kamhi & Catts, 2012; Kamhi & Catts, 2014). A student's skill level in just one of these skill areas can drastically affect the way they perform on a reading assessment (Collins, 2015). If a student has difficulty with decoding and word reading, those difficulties can drastically inhibit his or her ability to understand information (Kintsch & Kintsch, 2005). Since decoding and word recognition can negatively affect student performance on reading comprehension assessments (Keenan & Meenan, 2014; Spear-Swerling, 2004), other methods of presenting information should be considered (Sweet & Snow, 2003). One method may include having an instructor read aloud as students follow along reading silently with their own text (Williamson, Carnahan, Birri, & Swoboda, 2015). Another method is exposing students to video or picture images that help the students to understand information (Choi & Johnson, 2007). As new literacies are being incorporated into everyday life, the use of digital technology is an alternative to traditional text that should be considered as a way to transport information and domain-specific knowledge to students (Alvermann & Eakle, 2007).

The proposed study will examine the influence of response format and presentation condition on comprehension performance in 5th grade students with and without reading disabities. The study will also determine the influence of listening comprehension, working memory, prior knowledge, and vocabulary on comprehension performance. This study is unique from previous studies in two ways: (a) videos will be used to present information as well as texts, (b) participants will be 5th grade students with and without RD. In the upcoming sections, I will discuss assessment dimensions, individual reader skills, and presentation of information.

CHAPTER II

REVIEW OF THE LITERATURE

Dimensions of Comprehension Assessments

The RAND Corporation Reading Study Group suggested that the reading comprehension process involved the interaction of four primary elements: reader, text, activity or task, and context (RAND, 2002). Each one of these elements can drastically alter performance on a reading comprehension assessment. Past studies suggest that variations in assessment dimensions such as text length (García & Cain, 2013; Keenan & Meenan, 2014; Spear-Swerling, 2004), time constraints (Garcia & Cain, 2013), oral versus silent reading (Garcia & Cain, 2013), and text genre (Best, Floyd, & Mcnamara, 2008; Collins, 2015) can lead to significant variations in assessment performance. The dimension that is most important for the purpose of this study is response format. The response format, or the method of collecting answers from students, varies depending on which comprehension assessment is being used. The problem with using different response formats to assess reading is that these tasks require different levels of proficiency in certain individual skills. That can cause variations in student performance across tests (Collins, 2015; Francis et al., 2005; Kobayashi, 2002; Kulesz, Francis, Barnes, & Fletcher, 2016; Pearson & Hamm, 2005).

Collins (2015) examined performance variations on reading comprehension assessments. In the study, three different response formats used to measure reading

comprehension were analyzed. Seventy-nine fourth graders read a combination of three narrative and three expository passages from the *Qualitative Reading Inventory-Fifth Edition* (QRI-5) and then completed a comprehension assessment of each passage. Each assessment had a varying response format which was either multiple-choice, open-ended, or retell. Results revealed statistically significant differences between open-ended and multiple choice questions with participants having higher accuracy with multiple-choice questions. Other studies (Ozuru, Briner, Kurby, & McNamara, 2013) have also shown that response format affects comprehension performance for both typically developing (TD) students and students with reading disabilities (RD) and contribute to the achievement gap between those two groups of students.

A recent meta-analysis of 82 studies compared performance on reading comprehension assessments for students with and without RD in kindergarten through 12^{th} grade (Collins, Lindstrom, & Compton, 2017). Six reading comprehension response formats were used, including multiple-choice, open-ended, retell, cloze, sentence verification, retell, and picture selection. As expected, students with RD did not perform as well as TD students on reading comprehension assessments, but the gap between students with RD and TD varied widely across the different studies. The gap was consistently larger on open-ended questions and picture selection than on retell. The findings from this study confirm that response format has a significant impact on comprehension performance.

Presentation of Information

It is generally assumed that comprehension problems in students with RD are limited to word-level reading (Keenan et al., 2008), but there is now considerable evidence that many of these students also have problems in listening comprehension (Roch, Florit, & Levorato, 2011; Wise, Sevcik, Morris, Lovett, & Wolf, 2007). The widely popular Simple View of Reading (Hoover & Gough, 1990), in fact, proposes that reading comprehension is the product of word-level decoding skills and language or listening comprehension. According to the Simple View of Reading, students can exhibit reading problems if they have deficiencies in language and listening comprehension regardless of their word-level reading abilities. Difficulties in word-level reading thus do not necessarily predict a student's ability to comprehend information presented verbally or through different media (Kendeou, Bohn-Gettler, White, & van den Broek, 2008). It may be the case that some of these students have difficulty understanding and learning regardless of how information is presented. In some cases, however, students might show better learning when information is presented with videos, such as YouTube.

Multimedia and digital technologies have been incorporated in classroom instruction (Hall & Stahl 2012). While multimedia in the classroom has become more commonplace in the past ten years, the theory behind this practice has been studied for years (Clark & Paivio, 1991). Paivio's Dual Coding Theory (DCT; 1986) explains how human behavior and experience operate on a wide network of modality specific verbal and nonverbal (i.e., imagery) representations. Central to the DCT is the premise that the brain makes connections between verbal and nonverbal input when both representations

are presented simultaneously. Those connections then strengthen the brain's ability to process and remember new information, thus allowing learning to become more effective (Clark & Paivio, 1991; Hall & Stahl, 2012). The DCT helps to explain why students may benefit from verbal and nonverbal support while working on high processing such as reading comprehension.

Literacy teachers have long used verbal and nonverbal representations in reading instruction. Picture books are an oft used tool that combines printed words with imagery (Hall & Stahl, 2012; Kendeou et al., 2006). As digital technologies continue to advance, it is important to see how new media can be used to teach and enhance high concept skills such as deep thinking and comprehension of information and text. Past research (Hall & Stahl, 2012; Silverman & Hines, 2009) has examined comprehension across different media.

An early study that examined the relationship between multimedia and comprehension focused narrative comprehension for young, at-risk children (Sharp et al., 1995). The study specifically examined if a multimedia environment with dynamic visual support would facilitate language comprehension when children listened to short stories. Eighteen kindergarteners were asked to remember and recall information from a series of stories. The stories were told in three conditions. In the "helpful video" condition, dynamic, silent video accompanied the beginning of stories. In the "minimal video" condition, static images of characters and places accompanied the beginning of stories. In the "no video" condition, children only heard the stories. The children were better able remember sentences when stories were presented in the "helpful video"

condition than either of the other two conditions. These results suggest that dynamic visual supports provide an effective framework for understanding and remembering linguistic information.

Another study examined the relation of children's inference generation skills and narrative comprehension in different media from preschool to early elementary school (Kendeou et al., 2008). The researchers followed two cohorts of children aged four and six as they turned six and eight years old, respectively, and compared their narrative comprehension and inference generation in aural, televised and written stories. For both assessment times two narratives were used as part of the narrative comprehension assessment. One narrative was presented aurally (on an audiotape) and the other audio-visually (on a television). The findings showed that children's inference generation skills were highly inter-related across different media for both cohorts and at both time points. The results indicated that young children's inference skills can be assessed via the use of aural or televised stories. The results also indicated that narrative comprehension of aural and television narratives at age six can directly predict reading comprehension at age eight.

The use of multimedia to improve comprehension is not only effective in young, developing readers. The influence of video technology on comprehension has also been researched at the college level. One study investigated the effects of video and group discussion components of problem-based video instruction (PBVI) on college students' learning (Choi & Johnson, 2007). The study specifically examined whether or not PBVI could improve learner satisfaction, comprehension, and retention. Participants included

147 college students enrolled in a social science course. The students participated in three different experiments. In the PBVI, students viewed a video that represented how challenges were related to the week's lesson then participated in the small-group discussion. The problem based text instruction (PBTI) was identical to the PBVI except that the real-life challenges were presented in a text format. The PBVI without group discussion was identical to the PBVI except that the problem-solving activity was individually conducted without group discussion. The results of the study showed significant differences between the PBVI and PBTI variables, with students reporting better understanding and comprehension of information from the PBVI lesson. The information from the PBVI lesson was also retained longer than the PBTI lesson. This suggests that video technology is effective in college courses and can enhance student satisfaction, comprehension and delayed retention.

The results of these studies demonstrate that comprehension skills can transfer across different types of media (Kendeau et al., 2006). Assessment of comprehension does not need to be restricted to one specific medium (Hall & Stahl, 2012). These research studies show promising findings on video technology and comprehension, but the research is not complete. Much of the research on multimedia and comprehension focused on early reading skills and narrative comprehension (Kendeau et al., 2006; Sharp et al., 1995). Choi and Johnson's (2007) research of college students indicates that video technology can be used for complex, higher level comprehension. While there has been more of an effort to recognize digital media and incorporate new literacies into the

classroom (Leu et al., 2011), there is still little research on using video technology to assess and teach comprehension of information.

It is important to understand how the presentation of information can be an important factor in how information is retained and understood. The reading of text is the traditional way for children to gain knowledge and be exposed to new information. However, the individual dimensions of reading comprehension (decoding, word recognition, processing) may also play a part the breakdown of information. As more technology starts to be incorporated into the school system, it is important to examine how new literacies and means of transferring information to students can intersect with their individual skills in order to best assess their level of knowledge. It is also important to consider how individual skill level and characteristics may affect comprehension. This will be explored in the next section.

Reader Characteristics and Skills

Over the years, different models of comprehension and learning have shown how different skills can play significant role in how a reader is able to understand and comprehend information (Baddeley, 2000; Kintsch, 1998; RAND Reading Study Group, 2002; Zwann, & Radvansky, 1998). Individual differences in reader characteristics and skill level can also cause variance in performance on comprehension measures. Some scholars have examined how executive functions, such as attention and working memory, can affect student performance on comprehension assessments (Arrington, Kulesz, Francis, Fletcher, & Barnes, 2014; Collins, 2015; Eason, Goldberg, Young, Geist, &

Cutting, 2012). Studies have shown that an individual's background knowledge of a subject can strongly influence student performance (Best et al., 2008). Other studies have examined decoding and word recognition as predictors of performance on reading comprehension tests (Keenan et al., 2008).

In their recent study, Kulesz et al. (2016) investigated how different dimensions of reader skills of middle and high school students affected performance on standardized comprehension assessments. Reading skills assessed included word reading, working memory, background knowledge, and vocabulary. Results of their study revealed that vocabulary and background knowledge were the most important reader characteristics needed for the best performance on comprehension assessments.

The work-level reading skills of the reader have also been proven to be important. Approximately 30% of school-aged children have difficulty understanding what they read (Kulesz et al., 2016). Studies on reading comprehension difficulties have often focused on decoding and word reading. However, there are students who have age appropriate word reading accuracy but have comprehension ages below both their chronological ages and their word reading accuracy age (Cain, Oakhill, & Bryant, 2000). Specifically, students who are have difficulty with reading comprehension often have difficulty answering questions that require interpretation or questions where the answers are not explicitly stated in the text (Tirado & Saldaña, 2016).

Collins (2015) examined the cognitive, linguistic, and behavioral skills of participants in relation to their performance on reading comprehension tasks. Specific skills assessed included listening comprehension, nonverbal reasoning, domain

knowledge, vocabulary, working memory, word recognition, decoding, attention, and reading strategies. Skills that were statistically significant predictors of reading comprehension performance included working memory, listening comprehension, attention, and word recognition.

In the ongoing quest to best assess and comprehension level, it is important to know which response questions best measure knowledge, which presentation conditions best convey information, and how individual reader skills affect the ability for a child to comprehend text and information.

Purpose of the Current Research

The purpose of the present research was to examine how linguistic skills, cognitive skills, presentation conditions, and response formats contribute to comprehension of information in students with and without RD. The students with RD tested within normal limits for word-level reading skills but had deficits in reading comprehension. In previous studies (Best et al., 2008; Keenan & Meenan, 2014), comprehension was assessed only through reading. The present study included the addition of a video condition when presenting information. This study also included students with RD and sampled a slightly older population (5^{th} graders) than the 4^{th} graders studied by Collins (2015). The specific research questions and hypotheses addressed in this study were:

1. Are there group differences in student comprehension across different response formats and presentation conditions? Hypothesis: TD students will perform

better on comprehension assessment regardless of response format or presentation condition.

2. How will response format of questions (multiple choice vs. open ended questions) affect comprehension performance? Hypothesis: For both the TD and the RD groups, students will have a higher score on multiple course questions than on open-ended questions.
3. How will presentation condition (reading vs. video) influence comprehension performance? Hypothesis: There will be no difference in presentation conditions for the TD group. For the RD group, participants will perform better in the video condition.
4. What is the relative contribution of working memory, listening comprehension, background knowledge, and word-level reading on comprehension performance for the two response formats? Hypothesis: Listening comprehension will be the best predictor of comprehension performance for both response formats.

CHAPTER III

METHOD

Participants

Flyers were distributed to 5th grade students and parents of 5th grade students at five public and charter schools in Greensboro, North Carolina and Greenville, South Carolina. All children came from schools that qualified for federal lunch aid. Interested parents contacted the researchers to set up testing times. Participants were 32 5th grade students; 17 had RD (mean age = 11.18, range = 10.0-12.1.) and 15 were TD (mean age = 10.79, range 10.11-11.80). Groups were roughly matched for sex and race (see Table 1). To qualify for the RD group, students had to have a standard score within one standard deviation of the *Sight Word Efficiency* and *Phonemic Decoding Efficiency* subtests of the *Test of Word Reading Efficiency-Second Edition* (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012). This was a requirement to make sure that the students were capable of word-level reading and any reading comprehension failure was not caused by a deficit in word-level reading ability. Students in the RD group also had a standard score below 85 on both the *Listening Comprehension* and *Passage Comprehension* subtests of the *Woodcock Reading Mastery Tests, Third Edition (*WRMT-III, Woodcock, 2011). All students in this group had documented reading comprehension deficits in their school setting, either through an Individualized Education Plan (IEP) or through enrollment in a school-wide intervention program (i.e., Response to Intervention (RTI)).

All students were English speakers not currently enrolled in any English as Second Language program. Students were excluded from the study if they are identified as having significant disabilities such as intellectual disabilities, hearing impairment, visual impairment, brain injury, behavioral disabilities, Attention Deficit Hyperactive Disorder (ADHD) or Autism Spectrum Disorder (ASD). Students enrolled in gifted programs were also excluded from this study. Students in the TD group had no history of speech and language disorders with parents reporting that there were no documented problems with reading curriculum.

Table 1. Demographics of Participants

Variable	Typically Developing (*n*=15) Frequency	Reading Disability (*n*=17) Frequency
Sex		
Male	10	13
Female	5	4
Race		
Black	10	12
White	5	5

Standardized Test Measures

Word level reading. Word recognition and decoding was assessed with the *Sight Word Efficiency* and *Phonemic Decoding Efficiency* subtests of the *Test of Word Reading Efficiency-Second Edition* (TOWRE-2; Torgesen et al., 2012). For this test, students were given 45 seconds to read a list of real or nonsense words. Test-retest reliability for both

the *Sight Word Efficiency* subtest and the *Phonemic Decoding Efficiency* subtest of the TOWRE-2 are .90 and .91, respectively.

Working memory. Research studies have revealed a link between working memory and reading comprehension ability (Collins, 2015). When considering individual reader skills' role in comprehension, it is important to consider the impact of individuals' working memories. For the present study, working memory was assessed using the *Automated Working Memory Assessment (*AWMA; Alloway, 2007). In this computerized assessment, students completed verbal memory tests involving digit and non-digit stimuli. Test reliability of the AWMA was measured on 128 individuals randomly selected across schools and universities aged between 4.10 years to 22.5 years and ranged from .69 to .90.

Vocabulary. Research continues to stress the importance of receptive vocabulary on reading comprehension (Collins, 2015; Lesaux, Crosson, Kieffer, & Pierce, 2010). Receptive vocabulary in this study was assessed using the *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-4; Dunn & Dunn, 2007). For this test, the examiner presented a series of pages that contains four pictures. Each picture was numbered. The examiner then said a word and the student identified the number of the picture that best corresponded to the word. The split-half reliability ranges from 0.60 to 0.80 while the test-retest reliability ranges from 0.70 to 0.90. The validity was established in comparison to other vocabulary assessments and ranges between 0.68 and 0.76. The PPVT is consistently used in research studies on reading comprehension to measure the receptive vocabulary of research participants ranging from the age of two to adulthood

(Drijbooms, Groen, & Verhoeven, 2016; Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015; Rice, & Hoffman, 2015)

Listening comprehension. Listening skills were measured with the *Listening Comprehension* and *Passage Comprehension* subtests of the *Woodcock Reading Mastery Tests, Third Edition (*WRMT-III, Woodcock, 2011). On the *Listening Comprehension* subtest, students listened to passages read by the examiner and verbally responded to questions about their content. On the *Passage Comprehension* subtest, a modified cloze procedure was used where students were asked to read a passage in which a single word had been removed. Students were then asked to provide one word to complete the sentence. Median reliability coefficient alphas for all age groups for the standard battery ranged from .81 to .94. The clusters median for split half reliability is .95 with a range of .87 to .98. and the total median for split half reliability is .97 with a range of .86 to .99. The tests median for internal consistency is .91 with a range of .68 to .98.

Comprehension Measures

Four expository passages were selected from the ReadWorks.org website ("ReadWorks.org," 2016). ReadWorks.org provides research-based (RAND, 2002), Common Core aligned texts used to measure reading comprehension for students in grades kindergarten through the 12th grade. The website has over 12,000 Lexile leveled reading passages that come with accompanying multiple-choice and open-ended questions. These passages were used to assess comprehension in two conditions: reading and video.

All selected passages were analyzed using Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix is a computational tool that analyzes linguistic and discourse representations of a text through multiple measures. These measures are then characterized into eleven groups: (1) descriptive, (2) text easability principal component scores, (3) referential cohesion, (4) latent semantic analysis, (5) lexical diversity, (6) connectives, (7) situation model, (8) syntactic complexity, (9) syntactic pattern density, (10) word information, and (11) readability.

Coh-Metrix indices were used to see if the passages chosen were comparable. Lexical diversity refers to the variety of unique words (types) that occur in a text in relation to the total number of words (tokens). Type-token ratio (TTR) is a measure of lexical diversity that is the number of unique words, or types, divided by the number of tokens of those words. TTR for content words for the four passages ranged from .50-.70 with a mean of .6189. TTR for all words ranged from .32-.50 with a mean of .44.

Coh-Metrix provides an incidence score for all connectives which plays an important role in the creation of cohesive links between ideas and clauses and provide clues about text organization (Graesser et al., 2004). The connectives score ranged from 82-105 with a mean of 94.01. Coh-Metrix also has several measures for syntactic complexity, including the mean number of words before the main verb of the main clause in sentences which is a good index of working memory load. This measure ranged from 2.2-3.6 with a mean of 2.92.

Coh-Metrix uses the Flesch Reading Ease formula and the Flesch-Kincaid Grade Level formula to assess readability of a text. The Flesch Reading Ease formula is a

number from 0 to 100, with a higher score indicating easier reading. The Flesch-Kincaid Grade Level formula converts the Reading Ease Score to a U.S. grade-school level ranging from 0-12. The higher the number, the harder it is to read the text. The texts chosen to assess comprehension for this study had a Flesch Reading Ease formula that ranged from 66.05 to 75.28 with a mean of 69.44 and a Reading Ease Score that ranged from 6.2-6.7 with the mean Reading Ease Score being 6.469. By selecting passages one reading grade higher than the 5th grade participants, the possibility of a ceiling affect for TD participants was lessened.

Each passage had a video created for the purpose of this study. In the video condition, students were shown a PowerPoint slide show of pictures that correspond with an audio recording of the passage. The audio recording of each video was narrated by the lead researcher in the study. Screencast-o-matic software will be used to create the videos. To ensure that videos correspond to the text, a trained student from the Communication Sciences and Disorders (CSD) department at the University of North Carolina Greensboro (UNCG) used a 3-point scale (i.e. 1 - does not correspond, 2 - not sure, 3 – does correspond) to rate the correspondences of the videos with the text. All instances of unclear correspondence were modified and verified by the trained graduate student.

There were two response formats: (a) open-ended questions, and (b) multiple-choice questions.

Questions. Following each passage, students answered six open ended questions or six multiple choice questions. The questions that followed the selected passages were

used to assess students' level of comprehension. Each set of questions was research-based and followed the same structure so that the question types were used for each passage (Readworks.org, 2016). Students were allowed to look back over the text when responding. The questions were read aloud by the examiner as the student followed along with a written copy of the questions. Answers were scored as either right or wrong by following the answers provided by the ReadWorks.org website. Every set of questions included:

1. A text-based, literal question that focused on important information that was necessary for understanding the main idea or theme of the passage.
2. A question on text structure that assessed the students' awareness and understanding of the expository text structure.
3. A question that assessed students' ability to draw conclusions based on important information and evidence in the text.
4. A question that assessed students' inference abilities based on important information related to the main idea or theme.
5. A question that focused on the use of specific language or elements that were used to create meaning in the text.
6. A question that assessed the students' understanding of the main idea or theme question.

Examples of multiple-choice and open-ended questions are in Appendix C.

Multiple-choice questions. The selected passages came with accompanying multiple choice questions created for each passage. All questions and answer options

were read aloud by the examiner as the student followed along to minimize the potential effects of word recognition difficulties for students in the RD group.

Open-ended questions. The open ended questions consisted of the item stems from the multiple-choice questions without the answer options. Passage wording was changed to reflect an open-ended question as opposed to a multiple-choice question (e.g., Instead of "which of the following…," a question would be changed to "What evidence from the text…"). All questions were read aloud by the examiner as the students followed along. The students then wrote down an answer to the question. Student responses for open-ended questions were analyzed for correctness. The student responses did not have to match exactly the correct multiple-choice answer option provided by Readworks.org. The correct responses had to show accurate knowledge and understanding of the text and the questions being asked.

Procedures

Consent and assent was asked for each participant. Once consent and assent has been obtained, assessment batteries were administered in two 60 minute sessions. Testing sessions were conducted one-on-one by the examiner. In the first session, students completed the *Sight Word Efficiency* and *Phonemic Decoding Efficiency* subtests of the *Test of Word Reading Efficiency-Second Edition* (TOWRE-2; Torgesen et al., 2012), the *Automated Working Memory Assessment (*AWMA; Alloway, 2007), the *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-4; Dunn & Dunn, 2007), and

the *Listening Comprehension* and *Passage Comprehension* subtests of the *Woodcock Reading Mastery Tests, Third Edition (*WRMT-III, Woodcock, 2011).

In the second session, the four selected passages were administered to all students. The second testing session was completed within one week of administration of the first assessment battery. For each passage, the order of administration was: concept questions, reading/watching passage, answer questions.

The session began with the examiner asking five concept questions about each passage to assess students' prior knowledge of the topic. Each concept questions was worth one point. A question was marked correct if the student answered with a definition, an example of the concept, specific attribute, or function of the passage content. No points were given if a student gave answers unconnected to the question. Examples of concept questions are in Appendix B.

After concept questions were answered, students were presented with a passage read. Passages were presented the following order: (a) read passage then answer questions, (b) watch video then answer questions, (c) read passage then answer questions, (d) watch video then answer questions. All passages and response formats were counterbalanced.

Reliability

Inter-rater and intra-rater reliabilities were calculated on 25% of the samples for each measure used and was reported as percent of agreement for all samples. Disagreements were resolved through discussion. Inter-rater reliability was 98% for all

standardized assessment measures of cognitive and linguistic skills. Intra- rater reliability was 99.99% for all standardized measures of cognitive and linguistic skills. Inter-rater reliability was 96% for concept questions and 95% for comprehension questions. Intra-rater reliability was 98% for concept questions and comprehension questions.

Data Analyses

Data were analyzed in terms of the total standard scores from the TOWRE, AWMA, and PPVT, the standard scores from the subtests of the WRMT, the total scores of the four sets of concept questions (range 0-5), and the total scores of the four comprehension measures (range 0-6). All descriptive data and statistical analyses were performed using IBM SPSS.

A 2 (group) x 2 (presentation condition) x 2 (response format) mixed-design, repeated measures ANOVA was run to test for differences and possible interactions between the three independent groups. A series of independent *t*-tests to analyze group differences on measures of word-level reading, vocabulary, listening comprehension, working memory, and prior knowledge. Cohen's *d* was used to calculate effect size for both the ANOVA analyses and the *t*- tests. Pearson correlations were conducted to examine the relationship between measures of word-level reading, vocabulary, listening comprehension, working memory, prior knowledge, and comprehension performance. Follow up regression analyses were carried out to determine the best predictors for each response format for the total sample.

CHAPTER IV

RESULTS

Table 2 presents the group means and standard deviations for the measures of decoding, vocabulary, listening comprehension, working memory, and prior knowledge. In this study, decoding scores and listening comprehension scores were used to determine with which group the students would be counted. To eliminate decoding as a reason for any reading deficits, students in both groups had to have a standard score within one standard deviation of 100 on the decoding measure. To qualify for the RD group, students had to have a standard score below 85 on the listening comprehension measure. An independent *t*-test confirms that there were significant differences among the groups. As expected, there were no significant group differences in decoding due to the selection criteria. Significant group differences were found for the measures of listening comprehension t (23, 8) = 3.27, p =.044, d = .91, working memory t (13, 18) = 4.85, p =.001 d = .78, and prior knowledge t (8, 23) = 3.827, p =.005. d = .57. In addition, the measure of working memory and prior knowledge also confirm the differences between the two groups. There were no differences in vocabulary. Cohen's *d* was again used to calculate effect size for the *t*- tests. Effect sizes were large for the measures of listening comprehension and working memory. Prior knowledge had a moderate effect size.

Table 3 presents the means and standard deviations for comprehension performance as a function of group, presentation condition, and response format. A 2

(group) x 2 (presentation condition) x 2 (response format) mixed-design, repeated measures ANOVA found main effects for group, F (1,30) = 65.67, $p < .001$, $d = .69$, and response format, F (1,30) = 23.18, $p < .001$ $d = .44$. Cohen's d was used to calculate effect size. As seen in figure 1, the TD group obtained significantly higher comprehension scores than the RD group. For both the TD and RD groups, comprehension performance was significantly better for the multiple-choice questions than the open-ended question. TD students did best on the multiple-choice questions with a mean of 5 out of 6 for the reading format and 4.6 out of 6 for the video format. Comparatively, the best performance of the RD group was on 3.4 out of 6 on the multiple choice video questions. The mean for the RD group's best performance is approximated to the worst performance for the TD group, which was 3.7 for the open-ended video questions. None of the interactions were significant. Effect sizes were moderate for group and response format.

The next series of analyses examined the relationship between the independent variables and the two response formats, multiple-choice and open-ended questions. Table 4 presents the Pearson correlation coefficients for these measures. Listening comprehension, working memory, and prior knowledge were all significantly related to performance on both question types. The correlations for listening comprehension and working memory were noticeably higher for the open-ended questions than the multiple-choice questions. Regression analyses confirmed that these variables explained more of the variance for open-ended questions than multiple-choice questions. For the open-ended questions, 67% of the variance was explained by two of the variables in contrast to

the multiple-choice questions where one variable accounted for only 38% of the variance. Listening comprehension explained 60% of the variance (adjusted *R* Square = .60, *p* <.001) for the open-ended questions with prior knowledge accounting for an additional 7% of the variance (adjusted *R* Square .67, $p < .001$). For multiple-choice questions, only 38% of the variance was explained by working memory (adjusted *R* Square .38, $p <$ .001). No other variable contributed significant variance for multiple-choice questions. The very high correlation between listening comprehension and working memory ($r =$ .82, see Table 5) meant that only one of these variables would contribute to unique variance to performance (listening comprehension for open-ended questions and working memory for the multiple-choice questions). Of note is that neither vocabulary nor decoding abilities contributed unique variance to comprehension performance.

Decoding ability was not significantly related to the multiple-choice measures nor any of the other dependent variables. Decoding did have a small, significant correlation with open-ended questions (r=.42, see Table 4). Vocabulary had a small, significant correlation with open-ended questions (r=.53, see Table 4u) as well as the listening comprehension variable (r=.43, see Table 5).

Table 2. Group Means and Standard Deviations Scores for Independent Variables

	Typically Developing (*n*=15)		Reading Disability (*n*=17)	
	Means	(SD)	Means	(SD)
Decoding	99.53	9.28	94.35	8.45
Vocabulary	99.80	6.51	96.06	4.84
Listening Comprehension	105.87	11.70	80.94	4.45
Working Memory	100.47	.95	79.41	.51
Prior Knowledge	3.18	.41	2.53	.08

note: SD=Standard Deviation; Prior Knowledge range 1-5

Table 3. Group Means and Standard Deviations for Comprehension

	Typically Developing (*n*=15)		Reading Disability (*n*=17)	
Variable	Means	(SD)	Means	(SD)
Reading				
Open-Ended	4.0	1.20	2.1	1.20
Multiple Choice	5.0	1.33	3.1	1.80
Video				
Open Ended	3.7	1.29	1.6	.93
Multiple Choice	4.6	.91	3.4	1.06

note: SD=Standard Deviation; Comprehension Score Range was 0-6; 6 assessment questions per passage

Table 4. Pearson Correlation Coefficients between Independent Variables and Multiple-Choice/ Open-Ended Comprehension Scores (*n*=32)

	Open-ended	Multiple-choice
Decoding	.42*	.17
Vocabulary	.53**	.20
Listening Comprehension	.78***	.59***
Working Memory	.73***	.63***
Prior Knowledge	.67***	.63***

* *p*<.05 ** p<.01 ***p<.001

Table 5. Pearson Correlation Coefficients between Independent Variables (*n*=32)

	Decoding	Vocabulary	Listening Comprehension	Working Memory	Prior Knowledge
Decoding		.24	.34	.34	.25
Vocabulary	.24		.43*	.28	.27
Listening Comprehension	.34	.43*		.82**	.63**
Working Memory	.34	.28	.82**		.66**
Prior Knowledge	.25	.27	.63**	.66**	

Note: * *p*<.05 **p<.001

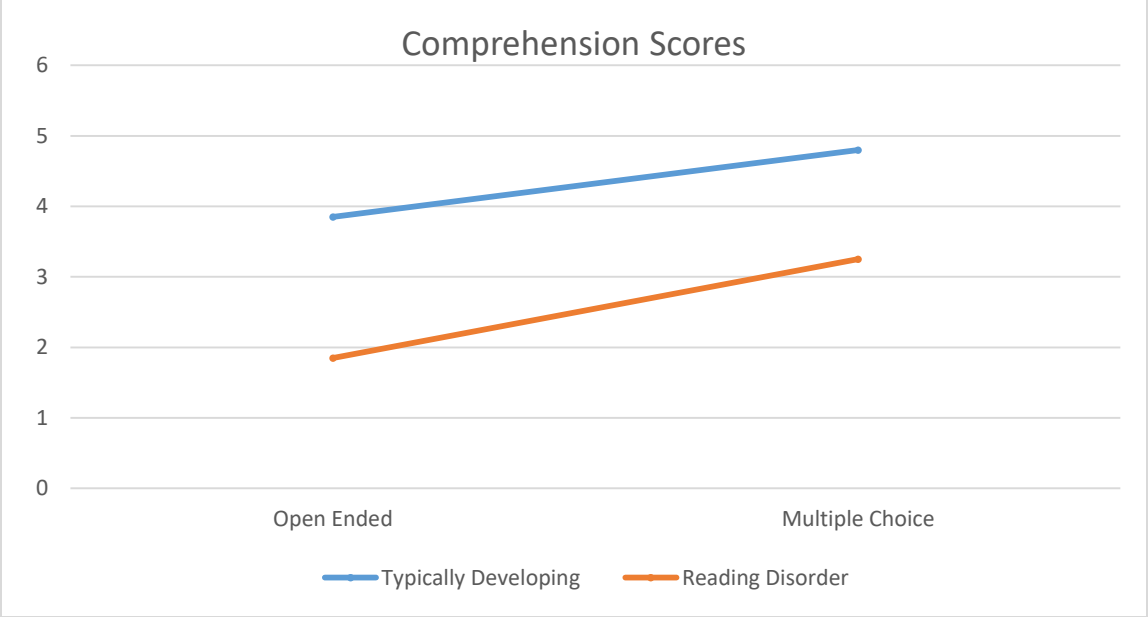

Figure 1. Mean Comprehension Scores for Typically Developing Students and Students with Reading Disabilities

CHAPTER V

DISCUSSION

The purpose of this study was to investigate the role of response formats and presentation conditions on comprehension for students with and without a reading disability. This study added to the ongoing research on reading comprehension by examining how different response formats and presentation conditions can affect comprehension assessment performance. While there have been other studies (e.g., Collins, 2015) that have examined the role of response formats in reading comprehension assessments, this study was the first to compare video and text presentations of information in TD students and students with a reading disability. As in Collins (2015), two response formats were used, open ended and multiple choice questions.

As expected, the TD group had significantly higher comprehension scores than the RD group across the two presentation conditions and response formats. This finding is not surprising given that the two groups differed significantly on measures of listening comprehension, prior knowledge, and working memory. The more interesting question was whether the RD group's comprehension performance was enhanced by the video presentation. It was not. There were no significant differences in comprehension performance between the two presentation conditions for either group.

The differences in memory demands of the two conditions might explain this finding. The memory demands in the text condition were lower than those in the video

condition because students were allowed to look back at the text when they were answering the comprehension questions. For the video condition, answers had to be based on what students remembered about the videos. Previous studies (Choi & Johnson, 2007: Kendeau et al., 2006) have found some benefit to comprehension with video presentations, but only when the videos were accompanied with text that presumably allowed for look backs. Discussion also facilitated comprehension performance of videos, but discussion also improved comprehension performance of texts (Applebee, Langer, Nystrand, & Gamoran, 2003; Choi & Johnson, 2007).

As predicted, students in both groups obtained significantly higher comprehension scores on the multiple-choice questions than the open-ended questions. These findings were consistent with the findings in previous studies by Collins (2015) and Ozuru et al. (2013). Pearson and Hamm (2005) have suggested that answering open-ended questions often requires more higher-level thinking skills than answering multiple-choice questions. Supporting this suggestion, listening comprehension and prior knowledge were the best predictors of performance on the open-ended questions. Listening comprehension (60%) and prior knowledge (7%) accounted for two-thirds of the variance for the open-ended questions. The strong relationship between listening comprehension and performance on the open-ended questions should not be surprising because the measure of listening comprehension used in this study, the oral comprehension subtest from the *Woodcock Reading Mastery Tests, Third Edition* (WRMT-III, Woodcock, 2011), also required responses to open-ended questions. These analyses indicate that open-ended questions provide a more accurate measure text comprehension than multiple-choice questions.

Multiple-choice questions are easier than open-ended questions because they rely less on language and prior knowledge. A correct response simply requires selecting the most likely answer. While past research indicates that multiple-choice assessments can do an adequate job of measuring a students' understanding of target material, there are studies that have revealed that multiple-choice questions provide wide variations of student knowledge and understanding of material (Simkin & Kuechler, 2005). Supporting this view, listening comprehension and prior knowledge contributed no significant variance for the multiple-choice questions. Working memory was the only significant predictor of performance on the multiple-choice questions, but it only explained 38% of the variance. This was 30% less that the variance explained by language and prior knowledge for the open-ended questions.

The important role of working memory in comprehension performance has been well-documented in theories of reading comprehension. In Kintsch's (1998) construction-integration model, which is regarded as the most comprehension model of comprehension, adequate working memory capacity is necessary to integrate knowledge related to text content with information from the text. In the current study, there were moderate-strong relationships with working memory and performance on open-ended (r = .73) and multiple-choice questions (r = .63). There was a strong relationship with working memory and listening comprehension (r = .82). As noted above, working memory was also the best predictor of performance on the multiple-choice questions.

The influence of prior knowledge on comprehension assessments has also been well documented in previous studies (Best et al., 2008; Kulesz et al., 2016). The present

study found moderate relationships between prior knowledge and students performance on open-ended (r = .67) and multiple-choice (r = .63) questions. Regression analysis revealed that prior knowledge contributed 7% unique variance to open-ended questions but no unique variance for multiple choice questions. Prior knowledge also had a moderate correlation with listening comprehension (r = .63) and working memory (r = .66) so the unique variance for response formats was absorbed by working memory instead of prior knowledge.

Results from the measures of word-level reading and vocabulary did not deviate between the two groups. These two variables did not contribute unique variance to the two response formats. This was expected for word-level reading since, to rule out a deficit in word-level reading, all participants had to score within one standard deviation on the *Sight Word Efficiency* and *Phonemic Decoding Efficiency* subtests of the TOWRE (Torgesen et al., 2012). Vocabulary was not in the selection criteria, but the two groups still performed comparably. Vocabulary is an important reader skill that can influence how students perform on reading comprehension assessments (Kulesz et al., 2016). Word-level reading (r = .42) and vocabulary (r = .53) did have a significant moderate correlation with performance on the open-ended questions, but neither variable was significantly related to performance on the multiple-choice questions. This does not mean that word-level reading and vocabulary do not affect comprehension. However, for 5th grade students whose word-level reading and vocabulary skills are within the standard range, decoding and vocabulary have little to no impact on comprehension compared with other measures. The reading difficulties for the students in the RD group were at the

sentence and discourse or text level. They were able to recognize words and read at the word level. For students that have poor comprehension skills but are adequate word-level readers, listening comprehension, working memory, and prior knowledge will help determine comprehension performance regardless of format.

Educational Implications

Reading comprehension assessments are an important part of the K-12 school curriculum. The results of these assessments help determine a student's perceived strengths and deficits and can greatly influence a student's trajectory in school. This study expanded the literature on reading assessment dimensions and echoed past research findings that a number of factors contribute to how students perform on reading comprehension assessments (Collins, 2015; Francis et al., 2005; Keenan, 2013; Keenan et al., 2008). Specifically, this research study reiterated that students perform better with the multiple-choice format than with the open-ended format. That does not mean, however, that a multiple-choice assessment is the best measure of knowledge and skills. As noted, individual skills accounted over two-thirds of the variance for open-ended questions while individual skills accounted for only 38% of the variance for multiple-choice questions. This indicates that even though students may perform better on multiple-choice assessments, open-ended assessments may be the best measure of what students truly understand and comprehend.

Reading comprehension is complex and difficult to measure. In order to get the most complete idea of how much a student understands, it may be best to assess students

using more than one reading comprehension assessment. These different comprehension assessments may consist of multiple response formats and assess different types of reader skills. This may give classroom teachers and specialists a better understanding of what students understand in the classroom and therapy room.

Important decisions regarding educational placement and resource services are made based on the results of reading comprehension assessments. The findings from the current study showed that response format affected comprehension performance. Educators should consider using at least two different response formats in assessing a student's understanding of a particular text.

Limitations and Future Directions

There were some limitations in this study. The sample size was relatively small which prevented comparisons of more than two response formats and presentation conditions. A larger sample size would not only confirm the validity of the findings from the current study, but also allow comparisons of other response formats (e.g., verbal or written recall) and presentation conditions (e.g., listening-only or listening + video).

Future studies should examine how certain reader skills can affect reading comprehension performance for various ages. Reading comprehension is assessed across all developmental levels and grade levels. Much of the previous research on language and literacy skills and reading comprehension assessment variations has focused on students in the elementary grades (Arrington et al., 2014; Best et al., 2008). Future research should also examine the effect of language and literacy skills on adolescent performance of

reading comprehension assessments at the middle and high school level. Assessments at this stage are longer, more domain-specific, and more cognitively demanding (Applebee et al., 2003). Future studies should focus on different groups.

The current study included a video condition in addition to a reading condition. While there were no significant findings, the video condition could be investigated in different ways in future studies. Researchers could examine how video can enhance text and help explain information to students (Choi & Johnson, 2007). Future research could also focus on how a video component might affect reading comprehension performance for different populations (i.e., ASD, ADHD, intellectual disability). Even though there were no significant findings in the current study does not mean that the use of video cannot be used in future studies as a way of conveying information to students. Future studies should also investigate a condition where the examiner reads while students listen. In the current study, there was a significant relationship between listening comprehension and both response formats. It would be beneficial to examine how students would perform with the addition of a listening condition to present information.

Summary and Conclusions

The purpose of this study was to compare the influence of response format, presentation of information, and individual reader skills on comprehension performance of 5th grade students with and without RD. All students were found to perform significantly better on the multiple-choice questions than the open-ended questions. As expected, the TD group had significantly higher comprehension scores on all measures.

Presentation condition did not significantly affect performance for either group. Listening comprehension, working memory, and prior knowledge contributed unique variance to performance on the different response formats. For the open-ended questions, 67% of the variance was explained by the measures of listening comprehension and prior knowledge. In contrast, only 38% of the variance was explained by working memory for the multiple-choice questions. This finding indicates that open-ended questions may be the best way to measure what students truly understand because open-ended questions rely more on language and conceptual knowledge than multiple-choice questions. Future studies should continue to examine how presentation condition and response format affect performance on reading comprehension assessments.

REFERENCES

Alloway, T.P. (2007). *Automated Working Memory Assessment.* London: Harcourt.

Alvermann, D. E., & Eakle, A. J. (2007). Dissolving learning boundaries: The doing, re-doing, and undoing of school. In D. Thiessen and A.Cook-Sather (Eds.), *International Handbook of Student Experience in Elementary and Secondary School* (pp. 143-166). Dordrecht, The Netherlands: Springer.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40,* 3, 685-730.

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18*, 325–346. doi:10.1080/10888438.2014.902461

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Best, R. M., Floyd, R. G., & Mcnamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29,* 137–164.

Cain, K., Oakhill, J., & Bryant, P. (2000). Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing, 12,* 1, 31-40.

Center for Teaching Excellence, Virginia Commonwealth University. (2016). *Writing good multiple-choice questions test questions. Retrieved from* *https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/*

Cervetti, G. N., Bravo, M. A., Hiebert, E. H., Pearson, P. D., & Jaynes, C. A. (2009). Text genre and science content: Ease of reading, comprehension, and reader preference. *Reading Psychology, 30,* 6, 487-511.

Choi, H. J., & Johnson, S. D. (2007). The effect of problem-based video instruction on learner satisfaction, comprehension and retention in college courses. *British Journal of Educational Technology, 38,* 5, 885-895.

Clark, M. K., & Kamhi, A. G. (2014). Influence of prior knowledge and interest on fourth- and fifth-grade passage comprehension on the Qualitative Reading Inventory-4. *Language, Speech, and Hearing Services in Schools, 45,* 4, 291-301.

Clark, J.M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3(3), 149–210. doi:10.1007/BF01320076*

Collins, A. A. (2015). *Performance variations across response formats on reading comprehension assessments.* Unpublished doctoral dissertation. Vanderbilt University, Nashville, TN.

Collins, A., Lindstrom, E.R., & Compton, D.L. (2017). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities, 2*, 1-16. doi: 10.1177/0022219417704636

Dunn, L. M. & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: Pearson Assessments.

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology*, *104*, 515–528.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

García, J. R., & Cain, K. (2013). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, 1-38. doi:10.3102/0034654313499616

Ghelani, K., Sidhu, R., Jain, U., & Tannock, R. (2004). Reading comprehension and reading related abilities in adolescents with reading disabilities and attention-deficit/hyperactivity disorder. *Dyslexia, 10,* 4, 364-384.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36,* 193–202.

Graves, M., & Aminotte Liang, L. (2008). Four facets of reading comprehension in the middle grades. *Middle School Journal, 39,* 4, 36-45.

Haladyna, T. M. (1999) *Developing and validating multiple-choice test items* (2nd ed.) Lawrence Erlbaum Associates, Inc.

Hall, M., & Stahl, K. A. D. (2012). Devillainizing video in support of comprehension and vocabulary instruction. *The Reading Teacher, 65,* 6, 403.

Hoover, W. and Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160.

Hsiao-fang, C. (2005). A Comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. *Foreign Language Annals, 37,* 4, 544-555.

Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment & Evaluation in Higher Education, 37,* 1, 125-136.

Joughin, G. (1998). Dimensions of oral assessment. *Assessment & Evaluation in Higher Education, 23, 367–78.*

Kamhi, A. (2007). Knowledge deficits: the true crisis in education. *ASHA Leader*, *12*, 7, *28–29.*

Kamhi, A. G. & Catts H.W. (Eds.). (2012). *Language and Reading Disabilities* (3rd ed.). Boston, MA: Allyn & Bacon.

Kamhi, A., G., & Catts, H., W. (2015). Reading comprehension: It's not a simple thing. [PowerPoint slides]. ASHA presentation.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*, 281–300. doi:10.1080/10888430802132279

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, *47*, 125–135. doi:10.1177/0022219412439326

Kendeou, P., Bohn-Gettler, C., White, M.J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading, 31*(3), 259–272. doi:10.1111/j.1467-9817.2008.00370.x

Kendeou, P., Lynch, J.S., van den Broek, P., Espin, C.A., White, M.J., & Kremer, K.E. (2006). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal, 33*(2), 91–98. doi:10.1007/s10643-005-0030-6

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71–92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19,* 2, 193-220.

Kulesz, P. A., Francis, D. J., Barnes, M. A., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology,* Advance online publication. http://dx.doi.org/10.1037/edu0000126.

Leslie, L., & Caldwell, J. (2017). *Qualitative Reading Inventory–6.* Boston, MA: Pearson Education, Inc.

Leu, D. J., Gregory, M. V. J., Ian, O. B. W., Kiili, C., Zawilinski, L., Everett-Cacopardo, H., Kennedy, C., ... Forzani, E. (2011). The New Literacies of Online Reading Comprehension: Expanding the Literacy and Learning Curriculum. *Journal of Adolescent & Adult Literacy, 55,* 1, 5-14.

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67,* 359–370.

National Center for Education Statistics [NAEP], Institute of Education Sciences, US Department of Education. (2015). Fast facts: English language learners. Institute of Education Sciences, US Department of Education. http://nces.ed.gov/fastfacts/display.asp?id=96.

National Governors Association Center for Best Practice, Council of Chief State School Officers. (2010). Common Core State Standards. Washington, DC: Author. Available from http://corestandards.org.

NOVA website, http://www.pbs.org/wgbh/nova/zero/ (accessed January 25, 2016).

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency* (2nd ed.). Austin, TX: PRO-ED, Inc.

Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology, 67*, 3, 215-227.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices---Past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13-69). Mahwah, NJ: Erlbaum.

Pearson, P. D., Valencia, S. W., & Wixson, K. (2014). Complicating the world of reading assessment: Toward better assessments for better teaching. *Theory into Practice, 53*, 3, 236-246.

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

Readworks.org (2016). Retrieved from http://www.readworks.org/.

Roch, M., Florit, E., & Levorato, C. (2011). Follow-up study on reading comprehension in down's syndrome: The role of reading skills and listening comprehension. *International Journal of Language & Communication Disorders, 46,* 2, 231-242.

Sharp, D.L.M., Bransford, J.D., Goldman, S.R., Risko, V.J., Kinzer, C.K., & Vye, N.J. (1995). Dynamic visual support for story comprehension and mental model building by young, at-risk children. *Educational Technology Research and Development, 43*(4), 25–42. doi:10.1007/BF02300489

Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of english-language learners and non-english-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101(2),* 305–314. doi:10.1037/a0014217

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: what is the connection? *Decision Sciences Journal of Innovative Education, 3,* 73-98.

Spear-Swerling, L. (2004). Fourth graders' performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology an International Quarterly, 25,* 2, 121-148.

Sweet, A. P., & Snow, C. E. (2003). *Rethinking reading comprehension*. New York: Guilford Press.

Tirado, M. J., & Saldaña, D. (2016). Readers with autism can produce inferences, but they cannot answer inferential questions. *Journal of Autism and Developmental Disorders, 46,* 3, 1025-1037. doi 10.1007/s10803-015-2648-6

Torgesen, J. K., Wagner, R. K., & Rashotte, C.A. (2012). *Test of Word Reading Efficiency* (2nd ed.). Austin, TX: PRO-ED, Inc.

Valencia, S. W., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context. *The Elementary School Journal, 115*, 2, 270-289.

Walraven, N. (2013). Improving the reading comprehension of fifth grade students. Retrieved from https://asu.tk20.com/fileshighered/3e01a8c07974932f13bc096b351-2f0/100007fecfcd9106128f7f16-5ee9/3e01a8c01d3319df142b730aa2452fe/1386090450545//Improving%20the%20Reading%20Comprehension%20of%20Fifth%20Grade%20Students.pdf

Waterfield, J. and West, B. (2006). *Inclusive assessment in higher education: A resource for change*, Plymouth: University of Plymouth. http://www.plymouth.ac.uk/pages/view.asp?page=10494 (accessed May 28, 2016).

Williamson, P., Carnahan, C. R., Birri, N., & Swoboda, C. (2015). Improving comprehension of narrative using character event maps for high school students with autism spectrum disorder. *Journal of Special Education, 49,* 1, 28-38.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research, 50,* 4, 1093-1099.

Wolraich, M. L., Bard, D. E., Neas, B., Doffing, M., & Beck, L. (2013). The psychometric properties of the Vanderbilt attention-deficit hyperactivity disorder diagnostic teacher rating scale in a community population. *Journal of Developmental and Behavioral Pediatrics, 34,* 2, 83-93.

Wolraich, M. L., Feurer, I., Hannah, J. N., Pinnock, T. Y., & Baumgaertel, A. (1998). Obtaining systematic teacher reports of disruptive behavior disorders utilizing DSM-IV. *Journal of Abnormal Child Psychology, 26,* 141– 152.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2011). *Woodcock Reading Mastery Tests, Third Edition (WRMT-III),* Bloomington, MN: Pearson.

Yonelinas, A. P. (2002). The nature of recollection and familiarity. *Journal of Memory and Language, 46,* 441–517.

Zwann, R. A. & Radvansky, G. A. (1998). Situation models in language and reading comprehension. *Psychological Bulletin, 123*, 162-185.

APPENDIX A

INSTRUCTIONS FOR READING PASSAGES

Today you will read two passages and watch two videos. Before you read each passage or watch each video, I'm going to ask you some questions. You will read the passages silently to yourself. I cannot help you with the reading in any way. If you come to a word you do not know, just do the best you can and continue on. Afterwards I will also ask you some questions about what you have read and you will either write or choose an answer. You can look back over the passages or the video while answering questions. Ready?

APPENDIX B

EXAMPLES OF CONCEPT QUESTIONS

From *Dig This*:
What happens when a volcano erupts?

From *Ecosystem of the Forest*:
In the ecosystem, what is a consumer?

From *The World's First Shopping Mall*:
What is an architect?

From *Tornado Scientists*:
What equipment is used to study tornadoes?

APPENDIX C

EXAMPLES OF COMPREHENSION QUESTIONS

Open-Ended Questions

From *Dig This:* What do paleontologists study?

From *Ecosystem of the Forest:* What is the passage mainly about?

From *The World's First Shopping Mall*: Where was the first modern shopping mall built?

From *Tornado Scientists*: What evidence from the passage tells us that tornados are dangerous?

Multiple-Choice Questions

From *Dig This*:
How does the author describe Earth?

A solid and motionless
B moving and changing
C smooth and shiny
D dangerous and lifeless

From *Ecosystem of the Forest*:
What is an ecosystem?

A a living being, such as a human, that eats other living beings in order to survive
B the process by which the body of a living thing is broken down by decomposers
C one particular area where several kinds of organisms interact with each other
D an organism that breathes in oxygen and then breathes out carbon dioxide

From *The World's First Shopping Mall*:
The development team of Southdale Center wanted to encourage customers to stay and shop for longer. What feature did the developers implement to achieve this goal?

A. windows in the ceiling and electric lights to make the day seem longer
B. climate control to make sure customers were comfortable
C. a zoo to keep customers entertained
D. trees and plants to relax customers

From *Tornado Scientists*:
4. What can information about one tornado tell scientists?

A. Information about one tornado can tell scientists how old a tornado probe is.
B. Information about one tornado can tell scientists where another tornado may happen.
C. Information about one tornado can tell scientists how many people took shelter from it in their basement.
D. Information about one tornado can tell scientists whether closets or bathrooms are better for taking shelter in.