

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations

Graduate Studies

5-2020

Teacher Questions in the Classroom: The Effects of Using a Low- to High-Level Questioning Sequence on the Text-Based Reading Comprehension Outcomes of Low-Performing Students

Shannon Harris Brown
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>

 Part of the [Special Education and Teaching Commons](#)

Recommended Citation

Harris Brown, Shannon, "Teacher Questions in the Classroom: The Effects of Using a Low- to High-Level Questioning Sequence on the Text-Based Reading Comprehension Outcomes of Low-Performing Students" (2020). *All Graduate Theses and Dissertations*. 7766.

<https://digitalcommons.usu.edu/etd/7766>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



TEACHER QUESTIONS IN THE CLASSROOM: THE EFFECTS OF USING A LOW-
TO HIGH-LEVEL QUESTIONING SEQUENCE ON THE TEXT-BASED READING
COMPREHENSION OUTCOMES OF LOW-PERFORMING STUDENTS

by

Shannon Harris Brown

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

Benjamin Lignugaris/Kraft, Ph.D.
Major Professor

Charles L. Salzberg, Ph.D.
Committee Member

Timothy A. Slocum, Ph.D.
Committee Member

Cindy Jones, Ph.D.
Committee Member

Ray Reutzel, Ph.D.
Committee Member

Richard S. Inouye, Ph.D.
Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2020

Copyright © Shannon Harris Brown 2020

All Rights Reserved

ABSTRACT

Teacher Questions in the Classroom: The Effects of Using a Low- to High-level Questioning Sequence on the Text-based Reading Comprehension Outcomes of Low-Performing Students

by

Shannon Harris Brown, Doctor of Philosophy

Utah State University, 2020

Major Professor: Dr. Benjamin Lignugaris/Kraft, Ph.D.
Department: Special Education and Rehabilitation

Teacher questioning may be an effective instructional procedure for building students' reading comprehension. Strategically asking questions at two different levels, low-level (*text explicit*) and high-level (*text implicit*), may be needed to assist students to engage in higher order thinking skills.

The purpose of this study was to evaluate the effectiveness of a low- to high-level questioning sequence *without or with linking prompts* on the text-based reading comprehension outcomes of fifth-grade students who evidenced poor reading comprehension. A secondary analysis was used to determine whether the questioning sequence was effective regardless of students' interest in the narrative stories used in the reading lessons.

Eleven fifth-grade students across three groups participated in this repeated measures study that consisted of two reading comprehension measures: response quantity

and comprehension accuracy. In addition, a multiple baseline design was applied across the lowest-performing students ($n = 5$). Groups of students engaged in reading lessons where one condition consisted of the low- to high-level questioning sequence and the other condition consisted of high-level questions only. Student outcomes for both reading comprehension measures were assessed immediately following each reading lesson. All students completed a student interest survey to identify their preference for the narrative stories.

Students increased the quantity and accuracy of their responses when the questioning sequence *with linking prompts* was implemented. This result was also found for four of the five lowest-performing students. Further, the questioning intervention was effective for increasing students' performance on both reading comprehension measures regardless of student interest in the narrative stories. Students preferred the high-level questions only condition but indicated that the low- to high-level questioning sequences helped them remember the stories better. Students also reported that they were better readers and liked reading the stories out loud in small groups, but had mixed ratings about leaving their classrooms to participate in the study.

Potential confounds and limitations of the study are discussed, specifically regarding the elements of the low- to high-level questioning sequences and study procedures as well as the need to further develop reading comprehension measures and student interest measures. Considerations for future investigations are also discussed.

(228 pages)

PUBLIC ABSTRACT

Teacher Questions in the Classroom: The Effects of Using a Low- to High-level
Questioning Sequence on the Text-based Reading Comprehension Outcomes of Low-
Performing Students
Shannon Harris Brown

Teacher questioning may be an effective instructional procedure for building students' reading comprehension. Strategically asking questions at two different levels, low-level (*text explicit*) and high-level (*text implicit*), may be needed to assist students to engage in higher order thinking skills.

The purpose of this study was to evaluate the effectiveness of a low- to high-level questioning sequence *without or with linking prompts* on the text-based reading comprehension outcomes of fifth-grade students who evidenced poor reading comprehension. A secondary analysis was used to determine whether the questioning sequence was effective regardless of students' interest in the narrative stories used in the reading lessons.

Eleven fifth-grade students across three groups participated in this repeated measures study that consisted of two reading comprehension measures: response quantity and comprehension accuracy. In addition, a multiple baseline design was applied across the lowest-performing students ($n = 5$). Groups of students engaged in reading lessons where one condition consisted of the low- to high-level questioning sequence and the other condition consisted of high-level questions only. Student outcomes for both reading comprehension measures were assessed immediately following each reading lesson. All

students completed a student interest survey to identify their preference for the narrative stories.

Students increased the quantity and accuracy of their responses when the questioning sequence *with linking prompts* was implemented. This result was also found for four of the five lowest-performing students. Further, the questioning intervention was effective for increasing students' performance on both reading comprehension measures regardless of student interest in the narrative stories. Students preferred the high-level questions only condition but indicated that the low- to high-level questioning sequences helped them remember the stories better. Students also reported that they were better readers and liked reading the stories out loud in small groups, but had mixed ratings about leaving their classrooms to participate in the study.

Potential confounds and limitations of the study are discussed, specifically regarding the elements of the low- to high-level questioning sequences and study procedures as well as the need to further develop reading comprehension measures and student interest measures. Considerations for future investigations are also discussed.

ACKNOWLEDGMENTS

The privilege for me to further my education has first and foremost shaped me as a person, which I value greatly. I recognize that every success, setback, resource, opportunity, and relationship in my educational path have influenced who am I striving to become, as well as the direction and purposes for my life. To complete my doctoral degree and this dissertation fills me with gratitude and acknowledgement for many people. To my major professor, Dr. Benjamin Lignugaris/Kraft, you inspired me to continue my path at the beginning of it all, back in my undergraduate years, and I thank you for being an excellent mentor, teacher, and colleague. Thank you for your guidance through two pilot studies and this dissertation. I have never taken your investment in my behalf for granted and appreciate every session of feedback, shaping, and time you have given. I am also indebted to Dr. Timothy Slocum, Dr. Charles Salzberg, and Dr. David Forbush for whom I could always count on for wisdom, instruction, and encouragement.

I would like to specifically thank department staff members Cosette Brunner, Tammy Soto, Teresa Simonsen, Kris Wengreen, Becky Morgan, and Darcie Peterson for their advisement and happy friendship throughout my years at Utah State University.

To Monica Lewis—what an honor to conduct research with you! This study would not have happened without your talent and support. To Anne Desjardins—one of the most inspiring women I have ever known, thank you for your passion, your dedication to teaching children, and for believing in me. To John Copenhaver—your support has meant so much and you have inspired me to continue in this important work. To my doctoral colleagues—it was a pleasure to work alongside you and I celebrate your successes and accomplishments!

I cannot possibly put in writing how incredible my parents are and the many details of how they have always supported me. To Mom—thank you for believing in me and giving me room to experiment and discover. Thank you for instilling in me a love of reading, from library trips in my youth to your example of life-long learning today as you continue to devour books and learn new things. The confidence I have to try new things and to create comes from you. I cherish all of your one-arm fist pumps and cheerful shouts of celebration these many years in school. I love you. To Dad—thank you for your consistency. You were always there to listen, mentor, and encourage. It has meant everything to have your support and trust for the decisions I’ve made in my life. I’ll never stop implementing the lessons you’ve taught me about “attacking with enthusiasm” and “showing up” and I daily look to you as my greatest example. I love you.

I feel blessed beyond measure for my family. To Jason, my husband, thank you for your steady support at the finish line. It means everything that you believe in me and trusted my process. I lovingly dedicate this dissertation to you. To Rhett, my son, you have renewed my dreams and ambitions—I hope to make you proud!

Finally, I want to acknowledge God in my life and in all that I do.

Shannon Harris Brown

CONTENTS

	Page
ABSTRACT.....	iii
PUBLIC ABSTRACT	v
ACKNOWLEDGMENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
CHAPTER	
I. INTRODUCTION	1
Teacher Questioning as an Instructional Tool	2
Levels of Processing for Comprehension	4
Question Levels	6
Questioning Strategies	9
Student Interest	13
Purpose & Research Questions.....	15
II. LITERATURE REVIEW	17
Reviews on Teacher Questioning	19
Search and Selection Criteria.....	23
Methodological Quality	59
Discussion.....	66
III. METHODS	71
Participants.....	71
Settings.....	77
Materials	78
Measures: Dependent Variables	85
Treatment: Independent Variable	91
Experimental Design.....	92
Procedures.....	95
Reliability.....	98
Treatment Fidelity.....	101

IV.	RESULTS	103
	Response Quantity Results	103
	Comprehension Accuracy Results	105
	Response Quantity & Comprehension Accuracy Results: Lowest Performers ...	106
	Student Interest Results.....	113
	Social Validity Results.....	124
V.	DISCUSSION	127
	Potential Confounds and Limitations.....	128
	Future Research	135
	REFERENCES	138
	APPENDICES	149
	Appendix A: Quality Indicators Coding Sheet: Correlational Studies	150
	Appendix B: Quality Indicators Coding Sheet: Experimental Studies.....	153
	Appendix C: Letter of Informed Consent	156
	Appendix D: Sample Reading Lesson Materials	161
	Appendix E: Scoring Rubric for Reading Comprehension Measures	173
	Appendix F: Social Validity Questionnaire.....	188
	Appendix G: Example: Comprehension Assessment with Scripted Directions ..	194
	Appendix H: Student Interest Survey	196
	Appendix I: Treatment Fidelity Checklist	201
	Appendix J: Student Interest Analysis.....	203
	Appendix K: Opportunities to Respond: Lowest Performing Students.....	205
	CURRICULUM VITAE.....	207

LIST OF TABLES

Table	Page
1. Demographics: Correlational Studies	29
2. Interventions, Measures, and Outcomes: Correlational Studies	32
3. Demographics: Experimental Studies.....	38
4. Interventions, Measures, and Outcomes: Experimental Studies.....	44
5. Ratings for Essential Quality Indicators: Correlational Studies	60
6. Ratings for Essential Quality Indicators: Experimental Studies.....	61
7. Literature Review: Overall Evidence Ratings	65
8. Year-end Proficiency Lexile® Ranges	73
9. Assessment Scores of Student Participants.....	74
10. Title, Lexile®, and Word Count for Narrative Stories	79
11. Question Packages: High-level Questions Only & Low- to High-level Question Sequence.....	82
12. High-level Questions: Percentage of Question Type per Narrative Story	83
13. Properties for High-level Questions Only and Low- to High-level Questions.....	84
14. Character/Trait, Event, and Idea References for Assessment Questions.....	87
15. Question Packages for Treatment Conditions.....	92
16. Lesson Schedule Across Groups of Students.....	94
17. IOA Results for Response Quantity and Comprehension Accuracy	100
18. Students Outcomes: Response Quantity and Comprehension Accuracy.....	105
19. Interest Survey Results for each Student	115
20. Percentage of High-interest Stories per Condition for each Student	117

21.	Response Quantity and Comprehension Accuracy Means and Standard Deviations for High-interest Stories	118
22.	Percentage of Low-interest Stories per Condition for each Student.....	121
23.	Response Quantity and Comprehension Accuracy Means and Standard Deviations for Low-interest Stories.....	122
24.	Social Validity Questionnaire Results	125
25.	Student Interest Analysis: Recentness Effects Across All Students.....	204
26.	Opportunities to Respond: Lowest Performing Students.....	206

LIST OF FIGURES

Figures	Page
1. Questioning in the Classroom: 3-Part Questioning Sequence	3
2. Overview of the Construction-Integration (CI) Model of Text Comprehension	6
3. Ranges for Student Lexile® Scores and Text Lexile® Scores	80
4. Example of Assessment Questions with Corresponding Student Response: Part 1 (P1) and Part 2 (P2).....	88
5. Response Quantity: Lowest-performing Students	108
6. Comprehension Accuracy: Lowest-performing Students.....	108

CHAPTER I

INTRODUCTION

Reading ability is critical for one's success in school and throughout life. For some, learning to read seems effortless and rapid, yet, for others, it is a difficult and frustrating process. The Research and Development (RAND) Reading Study Group (RRSG, 2002) determined that developing reading comprehension skills is one of the most pressing issues in literacy and emphasized that understanding how to improve reading comprehension outcomes for all students, especially low performers, is critical for future literacy research (Snow, 2002).

Students face increasing academic challenges for comprehending complex text as they advance through grade levels, making the task for teachers to increase text-based reading comprehension outcomes even more essential and ongoing. The RRSB (2002) determined that good instruction is the most powerful means of fostering the development of proficient comprehenders as well as preventing reading comprehension problems. To promote increases in students' reading outcomes, teachers must deliver evidence-based comprehension instruction.

The National Reading Panel (NRP)(2000) identified question answering (defined as "readers answer questions posed by teachers") as one of seven scientifically based reading comprehension instructional strategies. When reading in the classroom, teachers frequently ask students questions during ongoing verbal discussions to help students build their understanding of reading material (Borich, 1980; NRP, 2000; Wasserman, 1991). Used in this manner, teacher questions function primarily as an instructional tool to teach new content and secondarily as an assessment tool to monitor student learning

during a lesson. Throughout this process, teachers may clarify ideas, redirect students to the text, and confirm understanding. Moreover, teacher questions may also serve primarily as an assessment tool and secondarily as an instructional tool. Following the lesson, teachers may ask students to remember information and demonstrate understanding. This process allows teachers to confirm what students have learned as well as to highlight lesson material that may need additional instruction.

Teacher questions are essential to engage learners, foster critical thinking skills, deliver feedback, and monitor understanding (Caram & Davis, 2005). However, some teachers may not have proficiency in, or even access to, questioning procedures that help students engage in higher order thinking skills for answering complex questions (Bulgren, Marquis, Lenz, Deshler, & Schumaker, 2011; Raudenbush, Rowan, & Cheong, 1993). As a result, these teachers are not adequately prepared to strategically lead classroom discourse for building students' text-based reading comprehension (i.e., questioning as an instructional tool) or to determine what students have learned as a result of instruction (i.e., questioning as an assessment tool).

Teacher Questioning as an Instructional Tool

Both teachers and students stand to benefit from thoughtful questioning in the classroom. For teachers, questions provide opportunities for students to respond, promote higher student engagement, and deliver feedback (Gall, 1970; Levin & Nolan, 2004). For students, questions set the stage for continuous discourse, reinforce new learning, and promote high levels of critical thinking (Caram & Davis, 2005; Gall, 1970). When teachers thoughtfully deliver questions, the likelihood increases that students will focus

their attention on learning and monitor their own understanding (Rosenshine, Meister, & Chapman, 1996).

Sinclair and Coulthard (1975) defined instructional questioning in the classroom as a three-part sequence where “a typical exchange in the classroom consists of an *initiation* by the teacher, followed by a *response* from the pupil, followed by *feedback* to the pupil’s response from the teacher” (p. 21). This exchange can also be described as “the teacher asks a question, the learner gives an answer, and the teacher makes a comment” (Lynch, 1991, p. 201). Each element of the three-part sequence is important for general classroom learning, but the questions teachers initiate for building reading comprehension are especially critical, as they need to be strategically constructed and integrated into classroom discourse to increase student-centered learning (Dillon, 1981; Ellis, 1993; Roth, 1996). Figure 1 illustrates what teachers might think about when approaching classroom discourse via the three-part questioning sequence, with specific focus on what to prepare for when implementing the first step.

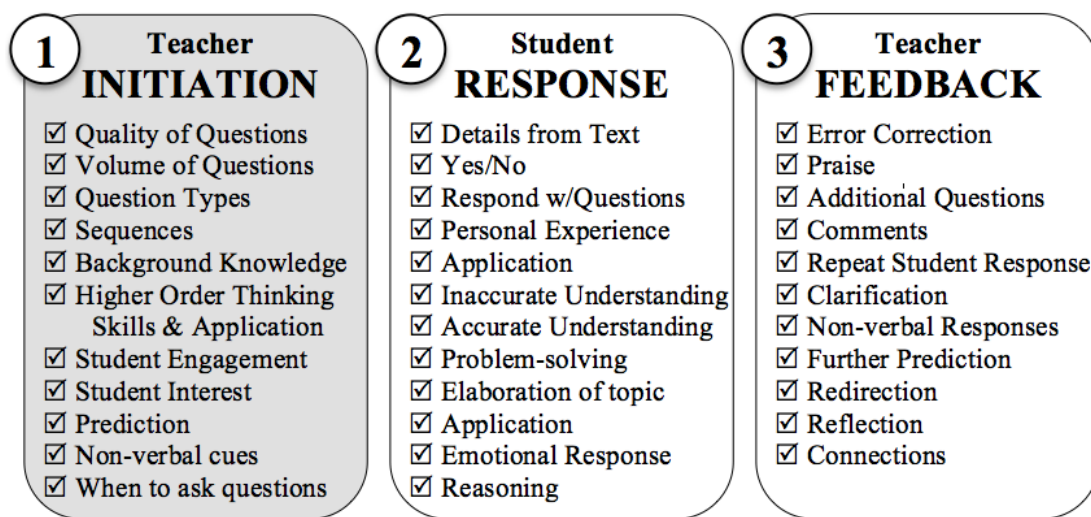


Figure 1. Questioning in the Classroom: 3-part Sequence.

Note: Definitions adapted from Sinclair & Coulthard (1975) and Lynch (1991).

Levels of Processing for Comprehension

Comprehension processes (e.g., creating mental images, making inferences) are important for understanding how to build reading comprehension and how to make the task of improving text-based reading comprehension easier (Pressley, 2001; Reutzel, Smith, & Fawson, 2005; Reutzel, 2014). These processes may be described using schemas. A schema is a cognitive framework that helps organize and interpret information (Cherry, 2019). However, some schemas may lead to the exclusion of pertinent information and only confirm pre-existing ideas. According to Piaget (1952), a schema is both the category of knowledge as well as the process of acquiring that knowledge (Anderson & Pearson, 1984). In schema theory, all knowledge is organized into units and is hierarchically categorized and connected into complex relationships (Piaget, 1952). The critical units that aid comprehension include the reader's knowledge about language, text, and the world around them. Schema theory is a conceptual system and helps teachers understand how knowledge is represented and how it is used. The fundamental element in the relationship of schema theory and reading comprehension is the assumption that written text does not carry meaning by itself, only the direction for readers to retrieve or construct meaning using their background knowledge (Seymour, 2017). Thus, schema theory provides an initial understanding of how readers' background knowledge is a critical element in the process of comprehending text as packages of knowledge stored in long-term memory and how it can be retrieved to aid in the comprehension of text (Reutzel, 2014). However, schema theory does not account for the role of text in the process of comprehension.

Kintsch's (1988; 1998; 2004; 2013; 2018) Construction-Integration (CI) Model of Text Comprehension provides the most fully-developed explanation of how background knowledge as well as other processes support text comprehension (Duke, Pearson, Strachan, & Billman, 2011; Graesser, 2007; NRP, 2000; Reutzel & Cooter, 2012; Reutzel, 2014; W. Kintsch, 2013; Wilkinson & Son, 2010). Kintsch (1988) asserts that text-based comprehension is a multi-leveled process. The CI model provides a framework for how teachers can support comprehension in text-based discourse in the classroom (See Figure 2). In this model, there are two major comprehension processes, *construction* and *integration*. The *construction* phase (lower-level processes) is the initial stage of reading comprehension in which ideas and concepts activate the reader's associations and simple inferences with the text, creating a *microstructure*, which represents the literal meaning of the text. The *integration* phase (higher-level processes) is when readers arrive at the final meaning of a text by strengthening relevant associations and dismissing non-relevant ones, creating a *macrostructure*, which is the global organization of these ideas into higher order units. The *microstructure* and *macrostructure* form the *textbase*, where the meaning of the text is represented as a network of concepts. When the *textbase* elements are combined with readers' background knowledge, the *situation model* is produced. Situation models are a form of inference where the reader essentially interprets what the text means (Reutzel, 2014). Within the situation model, students demonstrate that they can engage in low-level processing (e.g., recalling information) and then engage in high-level processing in order to evaluate and apply their knowledge (Almasi, 2003). The situation model that readers construct depends on their goals in reading the text as well as the amount of background knowledge they have.

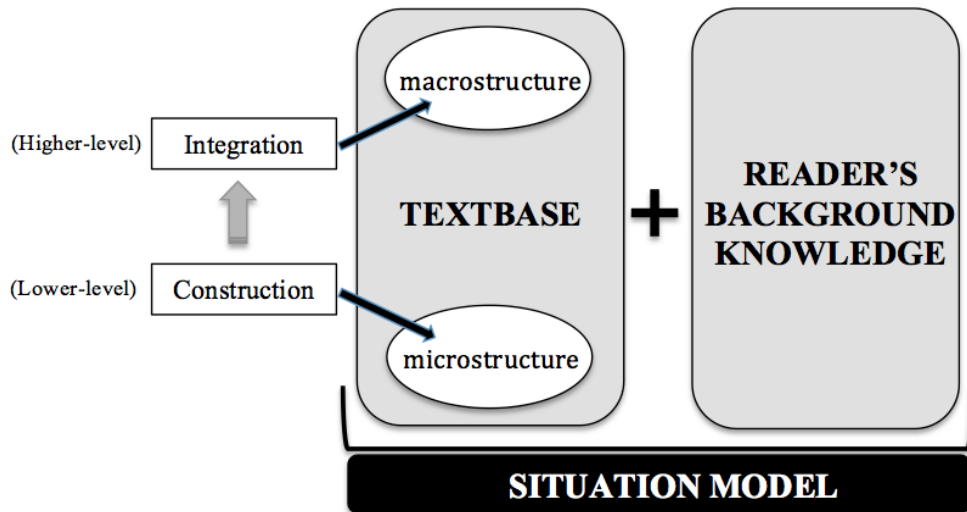


Figure 2. Overview of the Construction-Integration (CI) Model of Text Comprehension.

Reutzel (2014) identified several advantages of the CI model. First, the CI model emphasizes that text-based comprehension is a multi-leveled process. Second, the CI model positions text at the center of comprehension instruction rather than the reader's background knowledge. Finally, the CI model might be used to design instructional approaches that address the multiple levels of comprehension processes.

Question Levels

When using questioning as an instructional tool for building reading comprehension, teachers need to analyze questions at varying levels. Cognitive hierarchies are one of the most common ways to classify levels of questions. Gall (1970) estimated that there were at least eleven classification systems used to categorize questions as well as different category descriptions (e.g., recall, analytic thinking, creative thinking) used for analysis. In their classification system, Tienken, Goldberg, &

DiRocco (2009) used the terms *reproductive* and *productive* for low- and high-level questions, respectively. They found that teachers used reproductive questions (i.e., those that focus on low-level processes) 76% of the time and productive questions (i.e., those that focus on high-level integrative processes) 24% of the time.

In 1956, Bloom and Krathwohl developed a hierarchal system of ordering thinking skills known as Bloom's Taxonomy. This classification system is useful for discriminating between questions that focus on low-level processes and questions that focus on high-level processes. The focus of this classification system includes six major categories: *Knowledge*, *Comprehension*, *Application*, *Analysis*, *Synthesis*, and *Evaluation*. According to Kintsch's theory of discourse processing, students express *micro-level* and *macro-level* comprehension when responding to *Knowledge* and *Comprehension* questions (the first two categories of Bloom's Taxonomy). The remaining levels—*Application*, *Analysis*, *Synthesis*, and *Evaluation*—require high-level processing where students to engage in deeper levels of thinking, and demonstrate comprehension at the *situation level*. That is, students show how text relates to personal experiences and events. This is especially important when students must apply their knowledge for solving problems or judging the surrounding world (E. Kintsch, 2005; Kintsch, 2004).

In a study utilizing three different cognitive classification systems, Mills, Rice, Berliner, & Rosseau (1980) analyzed 54 typed transcripts to determine the percent of correspondence between the cognitive level of teacher questions and the cognitive level of student responses. The transcripts were coded based on the following classification systems: Bloom's Taxonomy, Aschner-Gallagher, and Smith and Meux. In their analysis,

percent of correspondence was based on the number of student responses that were coded in the same cognitive category as the teacher questions that elicited the responses (Mills et al., 1980). Mills et al. coded up to 3,483 episodes of correspondence between teacher questions and student responses and found that the Aschner-Gallagher classification system yielded the highest level of correspondence (56.1%) followed by Smith and Meux (51.4%), and Bloom's Taxonomy (51.3%). Even though the Aschner-Gallagher system yielded the greatest correspondence between teacher questions and student responses, Bloom's Taxonomy was the most useful for discriminating between low- and high-level teacher questions and low- and high-level students responses.

Raphael and Pearson (1985) described a questioning taxonomy directly linked to the text and the reader's schemata or background knowledge. Like Bloom's taxonomy, Raphael and Pearson describe question levels to support readers' comprehension before, during, and after reading. *Text explicit* questions (low-level) are "right there" questions often using who, what, where, and when prompts (What did Sally do when she heard the thunder?). *Text implicit* questions (moving from low- to high-level) require inference based on story details, often using why or how prompts (Why did she do that?). Importantly, these questions cannot be answered without the text, but they also cannot be answered by using exact words from the text. *Script implicit* questions (high-level) require an answer based upon the reader's background knowledge and experience with the topic (Why do you think people forget to plan for bad weather?). This is where text and background knowledge come together, promoting higher-order thinking (i.e., situation model) (Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018). This taxonomy considers the active role of the learner for understanding the processes

involved in answering questions. The processes, particularly for determining questioning levels, can in turn inform teachers' instructional approach for using questioning to improve student's reading comprehension.

Questioning Strategies

The definition of instructional questioning by Sinclair and Coulthard (1975) provides a basic understanding for how to think about questioning discourse in the classroom. Additional research for promoting deeper comprehension through strategies of linking questions is needed, especially for helping teachers to discriminate between and to strategically engage student's low- and high-level processes. Strategies that guide learning, support problem solving and reasoning, and refine comprehension through discourse are essential for actively building reading comprehension (E. Kintsch, 2005; Gholson & Craig, 2006; Pressley et al., 1992). Gallagher and Aschner (1963) found that the kind of thinking students engage in depends upon the kind of questions teachers ask. Thus, teachers who discriminate proficiently among question types and questioning strategies can help students understand what is read as well as make personal connections to the text in order to engage in higher-order thinking skills (Craig, Sullins, Witherspoon, & Gholson, 2006; Mangano & Benton, 1984; Redfield & Rousseau, 1981; Taboada & Guthrie, 2006; Wilen, 1991; Wimer, Ridenour, Thomas, & Place, 2001). Further, students may benefit from questioning strategies because they are designed to support the retention and transfer of information (Campbell & Mayer, 2009).

Different kinds of texts place different demands on learners and teachers must understand the critical differences between expository (informational, non-fiction) and

narrative text (stories and novels) when delivering reading comprehension instruction (NRP, 2000). This is important for teachers to know how to ask *text explicit*, *text implicit*, and *script implicit* questions for both texts because they differ in vocabulary use, organizational features, and analytic structures (e.g., comparisons and contrasts) (NRP, 2000).

Another factor that may influence comprehension of both narrative and expository text is student interest. Interest is defined as “a psychological state of having an affective reaction to and a focused attention for particular content” (Renninger & Hidi, 2002, p. 174). Student interest may influence how well they comprehend the text but not necessarily govern their reading comprehension overall.

Strategic Application of Questioning

Understanding how to move from *text explicit* to *text* and *script implicit* questions can be an effective strategy when instructing large or small groups of students (Goodwin, Sharp, Cloutier, Diamond, & Dalgaard, 1983). Further, Moyer and Milewicz (2002) concluded that teachers who question at various levels are more adept at assessing the range and depth of students’ thinking. Many students need support identifying salient story details (*text explicit* information), building accurate background knowledge (*schemata*), and then applying it appropriately (*text* and *script implicit*) (Schirmer & Woolsley, 1997).

The ability to identify and label questioning patterns may help teachers evaluate students’ understanding, determine students’ instructional levels, and develop strategies for promoting critical thinking (Buschman, 2001; Moyer & Milewicz, 2002; Ellis & Worthington, 1994; Sindelar, Bursuck, & Halle, 1986; Stronge, 2010). These skills are

essential for individualizing instruction to meet the needs of diverse groups of students (e.g., high-achieving, low-achieving, students with specific learning disabilities). Two types of questioning sequences teachers may use include (1) moving from high- to low-level questions or (2) moving from low- to high-level questions.

High- to low-level questions. In a high- to low-level questioning sequence, teachers start with a *text* or *script implicit* (high-level) question and move toward *text explicit* (low-level) questions contingent on students' responses. That is, high-level questions may be followed by low-level questions if students require more *text explicit* facts or details to establish their answer. Teachers may also use this sequence to prompt students to support or defend their answer or to understand how students arrived at the answer, similar to showing one's work for solving a math problem. The goal of this approach is to determine if students can engage in higher order situational thinking when first responding to what they read. Gall et al. (1978) conducted a study to determine the effects on student learning when asking questions in a high- to low-level sequence. Teachers asked students sixteen questions (eight high- and eight low-level questions) using a high- to low-level questioning pattern. The authors reported that teachers' use of this questioning strategy did not facilitate knowledge acquisition or improve responding to high-level questions as measured by the *Comprehensive Test of Basic Skills*. In a second study, the researchers varied the percentage of teachers' high-level questioning. Sixteen questions were included with each recitation; however different percentages of high-level questions (25%, 50%, and 75%) were in each treatment. In general, students who received only 25% high-level questions (and 75% low-level questions) outperformed students in the 75% and 50% high-level questions treatments. The findings from this

study suggest that questioning patterns that include *text explicit* or low-level questions lead to higher student achievement for responding to *script implicit* or high-level questions more than questioning patterns that emphasize high-level questions only.

Low- to high-level questions. A low- to high- level questioning sequence is an approach where teachers begin by asking *text explicit* (low-level) questions and move toward *text* or *script implicit* (high-level) questioning contingent on student responding. This direction ensures that students have the foundation knowledge needed to establish their answer. The goal of this approach is to reduce student errors and increase instructional and behavioral momentum. In essence, after ensuring that students have acquired essential *text explicit* knowledge, teachers can increase the demand placed on students to apply, synthesize, and evaluate what they have learned.

In a series of studies, Bulgren et al. (2009; 2011; 2013) utilized a Question Exploration Routine (QER) to support seventh-grade students' ability to respond to high-level questions. The use of the QER was compared to the traditional lecture-discussion method to determine if asking low- to high-level questions supported student achievement in reading comprehension. Students in the lecture-discussion method simply copied notes from the information a teacher provided using an overhead projector and students using the QER method sequenced low- to high-level information as part of the QER graphic organizer and strategic questioning. On assessments that varied with multiple choice, short-answer, and matching items, the QER condition resulted in students performing 26 points higher when compared to students participating in the traditional lecture-discussion format (QER, $M = 71.7$; Lecture, $M = 45.9$) (Bulgren, 2011). This suggests that teachers who focus on *text explicit* (low-level) questions prior to

text or *script implicit* (high-level) questions may effectively increase student reading comprehension outcomes.

Student Interest

Some researchers suggest that student interest plays a critical role in students' comprehension (Belloni & Jongasma, 1978; Haggard, 1986; Stevens 1980; Worthy, 2002). Methods for building reading comprehension may be more effective when high-interest materials are used (Belloni & Jongasma, 1978; Schiefele, 1996; Schiefele & Krapp, 1996; Stevens, 1980; O'Flynn, 2016). While the literature base on the relationship between student interest and comprehension is generally descriptive (Alexander & Jetton, 1996; Sauer, 2012; Subramaniam, 2009; Wigfield, Guthrie, Tonks, & Perencevich, 2004), researchers suggest that student interest correlates with deep understanding of text in contrast to a surface-level understanding. That is, students who are interested in a text tend to grasp details and can more readily apply, analyze, synthesize, and evaluate information, thus creating a strong situation model for the reader's comprehension (see Figure 2) (Almasi, 2003; Baldwin, Peleg-Bruckner, & McClintock, 1985; Sauer, 2012; Schiefele, 1996; Schiefele & Krapp, 1996; Taboada & Guthrie, 2006; Kintsch 1988; 1998; 2004; 2013; 2018). Importantly, student interest in a text may influence how well students comprehend, but it does not govern comprehension.

Within the literature base, *interest* is often addressed in its relationship with *motivation* but these terms should not be used synonymously. Rather, interest is a subset, or component of motivation, and the majority of available literature focuses on how motivation plays a critical role in overall learning (Alexander & Jetton, 1996; Gambrell,

2011; Hidi, 2006; Subramaniam, 2009; Wigfield & Guthrie, 1997; Wigfield, Guthrie, Tonks, & Perencevich, 2004). For example, Guthrie et al. (2006) proposed that reading engagement involves interactions with text that are motivated and strategic, concluding that motivation predicts how much a child reads (which is a predictor of reading comprehension). Further, Gambrell (1996) conducted an exploratory study to determine student's motivation to read using a Motivation to Read Profile (MRP) for gathering questionnaire and interview data, determining four key motivation factors for reading: access, choice, familiarity, and social interaction.

There are few experimental studies that specifically examine how students interest in a topic effects reading comprehension outcomes. Belloni and Jongsma (1978) conducted an experimental study to examine the relationship between reading interest and reading comprehension with low-performing students. Taking readability into account, students were tasked to read titles and abstracts of stories and select those they would most like to read and those they would least like to read. After reading stories, students completed a cloze test. The researchers found that students comprehended the high-interest stories better than they comprehended the low-interest stories. Several years later, Stevens (1980) assessed the interest of 25 topics using a verbal inventory questionnaire with fifth- and sixth-grade students. Reading passages were taken from a basal reader aligned with topics from the questionnaire and students completed a multiple-choice test after reading each story. Stevens found that students read significantly better under the high-interest condition than under the low-interest condition.

While the available research suggests that interest may impact student comprehension, there is little evidence of how student interest for the text interacts with teachers' application of comprehension strategies and its affect on text-based comprehension. Thus, it is important to gain an understanding of the general impact that a reading comprehension strategy (e.g., teacher questioning) might have on student learning, regardless of students' interest in the text material.

Purpose and Research Questions

Given the importance of questions in the classroom, researchers need to empirically investigate how student achievement might be effected as a result of specific approaches to teacher questioning. The current study extends the basic definitions of questioning provided by Sinclair and Coulthard (1975) and Lynch (1991) and examines the effects of integrating a system of linking questions or question sequences into classroom discourse. The strategic implementation of *text explicit*, *text implicit*, and *script implicit* questioning sequences during classroom discourse may improve students' reading comprehension (Raphael & Pearson, 1985). This approach may also help students engage in higher-order thinking skills when given assessment probes after the lesson.

The investigations for the current study are two-fold. First, as part of our primary analysis, we aim to reject the null hypothesis and propose that low- to high-level question sequences that are integrated into classroom discourse have an effect on students' text-based reading comprehension. We chose this sequence to ensure that student participants had *text explicit* knowledge (low-level) before responding to *script implicit* (high-level)

questions. Second, as part of our secondary analysis, we aim to explore how student interest may be a contributing factor in building reading comprehension, which is important for teachers to design instruction and help students comprehend text.

The purpose of this study was to evaluate the effectiveness of a low- to high-level questioning sequence *without or with linking prompts* on the reading comprehension outcomes of fifth-grade students who evidenced poor reading comprehension. Further, a secondary analysis was used to determine whether the low- to high-level questioning sequence was effective regardless of students' interest in the narrative story content.

The research questions addressed in this study were:

1. To what extent do low- to high-level questioning sequences increase fifth-grade students' response quantity and comprehension accuracy on a post-reading curriculum-based reading comprehension measure?
 - a. With low-performing students, to what extent do low- to high-level questioning sequences *without or with linking prompts* increase response quantity and comprehension accuracy on a post-reading curriculum-based reading comprehension assessment?
 - b. Given either high-interest or low-interest stories, to what extent do low- to high-level questioning sequences increase fifth-grade students' response quantity and comprehension accuracy?
2. To what extent do student participants rate their overall experience in the study and its impact on their learning and reading ability?

CHAPTER II

LITERATURE REVIEW

This literature review provides an examination of investigations on teacher questioning and its relationship to student achievement. Summaries included in this review highlight the current empirical research base on teacher questioning that align with the CI Model of Text Comprehension (Kintsch, 1988; 1998; 2004; 2013; 2018) as well as Raphael and Pearson's (1985) *text explicit*, *text implicit*, and *script implicit* taxonomy for the purpose of determining the best available evidence for implementing questioning strategies. Definitions for question levels and an overview of initial investigations of teacher questioning are provided below. The research question for this synthesis is: What are the effects of low- and/or high-level questions on the academic outcomes of students in Grades 4-12?

Definition: Low- and High-level Questions

Low-level or *text explicit* questions are those where the teacher is seeking literal, direct answers of factual information. These questions engage readers in the initial stage of reading comprehension, the *construction* phase, where the readers learn the literal meaning of the text (*microstructure*) (Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018). Questions at this level are usually either “right or wrong” and essentially involve recall of specific facts that are located directly in the text. For example, if a teacher asked the question, “What did Chandler hide in his sleeping bag?”, there is only one acceptable answer based on the story (his stuffed dog). For low-level questions, there are a limited number of acceptable answers for teachers to anticipate. For

example, a teacher might ask, “Where did the family put the old kitchen table?” where the anticipated response would not vary far from the limited correct answers, “in the living room,” “in another room,” or “in the corner.”

High-level, or *text* or *script implicit* questions require students to engage in higher-level processes of reading comprehension, the *integration* phase, where readers organize ideas in order to evaluate and apply their knowledge (Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018). High-level questions also require inference based on story details and background knowledge. The teacher is seeking more indirect and evaluative responses from students and there are many acceptable answers teachers may not anticipate. An example of a high-level question might be, “Why do you think some people cherish furniture so much?” where students must use their background knowledge to formulate a response. Another example of a high-level question might be, “How can objects, like jewelry, dream catchers, and toys, help give people courage to face their fears?” where students must engage in evaluative thinking that draw upon inference skills.

Teachers’ strategic use of low- and high-level questions assist readers in creating a *textbase* (*microstructure* and *macrostructure*). Importantly, when combined with a *textbase*, readers’ background knowledge plays an important role in building reading comprehension, allowing readers to recall information and evaluate and apply their knowledge (*situation model*) (Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018).

Reviews on Teacher Questioning

Systematic investigations of classroom dialogue began in the 1960s where it was determined that teachers did the majority of the talking in the classroom and did not ask many high-level questions to promote critical thinking; rather, the questions teachers emphasized focused on facts from the text (i.e., low-level knowledge) (Alexander, 2004; Gall, 1970; Topping & Trickey, 2007). One of the first researchers to empirically investigate teacher questioning based on Bloom's Taxonomy was Hunkins (1968), who found that asking high-level questions lead to higher student achievement for responding to high-level questions. In her review of teacher questioning, Gall (1970) proposed that educators need to first identify learning objectives and to determine which types of questions to ask (i.e., engage in didactic training) and then should consider how to use effective questioning strategies (i.e., prescribed sequences) that help students meet those objectives.

Winne (1979) completed a critical narrative review of 18 experimental and quasi-experimental studies conducted in the 1960s and 1970s to determine the effects of teacher questioning on student achievement. Higher cognitive questions (i.e., *text* or *script implicit* or divergent questions) were defined as those requiring that students manipulate previously learned information to create an answer with logically reasoned evidence, creating a *textbase*. Lower cognitive questions (i.e., *text explicit* or convergent questions) were defined as those calling for verbatim recall or recognition of factual information previously read or presented by a teacher.

Findings from this meta-analysis were inconclusive due to issues with methodological quality, the most erroneous being researchers' failure to document details

about the independent and dependent variables, problems with data collection for student achievement, and discrepancies between the definitions of low- and high-level questions. Winne determined that it was difficult to compare conclusions across studies because of the wide variation in dependent variables and the lack of information reported in many studies. Moreover, Winne concluded that the predominant use of either low-level or high-level questions made little or no difference on student achievement.

Several years later, Redfield and Rousseau (1981) conducted a meta-analysis of experimental research findings on the effects of teacher questioning on student achievement. Like Winne (1979), they used the same categories for qualifying studies (*Training* and *Skills*) as well as Campbell, Stanley, and Gage's (1966) criteria for internal validity. Of the 20 experimental and quasi-experimental studies that qualified, 18 were taken from Winne's (1979) narrative analysis. Redfield and Rousseau examined effect sizes for all studies. The researchers contacted the original authors for more data where possible and only 14 of the 20 studies provided the data necessary for establishing effect sizes.

Redfield and Rousseau concluded that there was a positive effect size on student achievement across studies (+.7292) when teachers' predominantly used high-level questions (*text* or *script implicit* questions) in the classroom. The effect size for *Training* studies was +.2245 and no effect size could be calculated for *Skills* studies due to issues with sample size, which the researchers acknowledged in their analysis. This finding supported those of Gall (1970) that teachers' use of high-level questions does lead to gains in student achievement, so long as teachers are trained in questioning skills and the implementation of instruction is accurate.

In contrast to Winne, the studies that Redfield and Rousseau (1981) reviewed had to include a standardized measure of student achievement. Moreover, they only examined one measure, a student achievement measure, from each of the qualifying studies.

Winne's (1979) procedure inflated the findings from each study (as more than one dependent measure was used in many of the studies) and therefore produced more comparisons and findings overall than Redfield and Rousseau (1981), who determined only one finding from each qualifying study.

Importantly, Redfield and Rousseau's (1981) review was more extensive and detailed than Winne's review. Redfield and Rousseau provided effect sizes, analysis of experimental validity, and coding of question types. However, their conclusion that asking more high-level questions produces greater student outcomes is limited because they did not separate high- and low-performing students in their analysis, so it is not clear if the predominant use of high-level questions is effective across low- and high-performing students. For example, Ellis (1993) suggests that low-level questioning is critical for improved achievement with low-performing students.

In 1987, Samson, Strykowski, Weinstein, and Walberg conducted a quantitative synthesis of the effects of teacher questioning levels on student achievement. Forty-four study variables (e.g., grade level, reliability) were coded for the 14 *Training and Skills* studies included in Redfield and Rousseau's analysis, particularly for whether researchers used low- and/or high-level questioning practices. A one-way analysis of variance (ANOVA) was conducted for each study variable and only 18 (41%) reached the .05 level of significance (subjects, dependent measure, treatment characteristics, and design factors). Further, some studies did not provide adequate information to code the study

variables. Samson et al. concluded that teachers' predominant use of high-level questions had a small effect on student achievement (mean effect size = .26) in contrast to Redfield and Rousseau's results (mean effect size = .73). Further, Samson et al. found that the treatments and conditions for producing large effects were unclear and that researchers were not able to produce large replicable effects. Based on the small effect size from their analysis, Samson et al. did not agree with Redfield and Rousseau's (1981) estimate of questioning effects.

Samson et al.'s (1987) review mainly scrutinized high-level questions and their effect on student achievement. Although high-level questions are essential for helping students engage in the critical thinking skills needed for deeper comprehension (Redfield & Rousseau, 1981; Samson et al., 1987; Topping & Trickey, 2007), findings on the combined effects of low-level and high-level questions were not reported. Further, these reviews did not provide a distinctive analysis where they separated students' performance based on their ability levels. For example, Redfield and Rousseau's (1981) findings favored the predominate use of high-level questions for building reading comprehension, but did not discuss how low-level questions might be needed for low performers (see Gall, 1970; Ellis, 1993). Further, there is limited understanding of how low- and high-level questions might be utilized together in various questioning sequences to scaffold students' high-level comprehension (Bulgren et al., 2011; Dantionio & Beisenherz, 2001; Graesser, Person, & Hu, 2002; Wilen & Clegg, 1986).

The National Reading Panel (NRP) (2000) conducted a review of the research on comprehension instruction from the years 1980-2000, finalizing a total of 203 qualifying studies published in scientific journals. Importantly, studies were required to have an

experiment that involved at least one treatment and a control group. The NRP classified and grouped studies into 16 kinds of instruction used (e.g., mental imagery, prior knowledge, mnemonic, story structure) and established two categories for student-lead questioning: *question answering* (for readers to improve skills for answering questions) and *question generation* (for readers to learn to generate and answer inferential questions during reading). Unfortunately, no studies were reviewed where the treatment involved teacher-lead questioning. Reliability, replication, and generality were the main criteria used when evaluating each strategy. Effect sizes could not be calculated for almost all of the studies and only two researchers reported effect size data.

The Panel concluded that “more information is needed on the effective ways to teach teachers how to use proven strategies for instruction in text comprehension” (NRP, 2000, p. 4-52). Teachers who implement strategies (e.g., strategic use of low- and high-level questions) that build text comprehension during reading can help increase student comprehension outcomes (NRP, 2000). The following scientific review examines teacher-lead questioning strategies and their relationship to student achievement.

Search and Selection Criteria

Since Samson et al.’s (1987) quantitative synthesis, a number of intervention studies examining levels of questioning in the classroom have been published. This review examines the research base for teacher questioning studies that have been published since 1988. Qualifying studies were identified through an electronic search of the EBSCO Host database and Academic Search Premier, ERIC, and PsycINFO databases published between January 1988 and August 2016. Twenty-nine different

combinations of the following descriptors were used: independent variable descriptors (*question* techniques, teacher question*, instructor question*, educator question*, question* behavior, convergent question*, divergent question*, low-level question*, high-level question*, question* levels, level of question, cognitive question*, intervention*); dependent variable descriptors (*academic achiev*, student outcomes*); and population descriptors (*LD, learning disab*, special education, student* with disab**). This initial search resulted in 542 articles. After conducting abstract screenings and removing duplicates, 55 potential articles were retained. This total was then narrowed down to 15 potential articles, from which citation searches were conducted, resulting in only one additional article. Articles that met the following selection criteria were included in this review:

1. The study was published in a peer-reviewed English language journal between January 1988 (after the Samson et al. (1987) review) to August 2016.
2. At least one dependent variable addressed student academic achievement or student outcomes.
3. The independent variable or intervention addressed question levels delivered by teachers or researchers to students (i.e., no student self-questioning strategies were included).
4. Participants included general education students or students with disabilities in Grades 4-12.
5. The study was experimental, quasi-experimental, correlational, or a single-subject design; Treatment-comparison studies needed to include a control or comparison group. Correlational designs were included as intriguing results could lead to an experimental study for future research.

The majority of articles were eliminated because they were descriptive studies, did not include student performance measures, or included interventions that focused on self-questioning strategies for students. Only eight studies in which researchers examined

the relationship between teacher questioning levels and student outcomes qualified for this research synthesis, resulting in six experimental designs and two correlational designs. Wolf, Crosson, and Resnick (2005) did not specify the *n* for each grade (1-8), but this study was retained because it still included participants in Grades 4-8. No studies were identified in which researchers used a single-subject experimental design.

Coding

Coding protocols were developed to determine the methodological quality of the available research, the study and intervention characteristics, and the overall strength of the available evidence. A coding document was developed and used to organize essential information about each type of study (see Appendices A & B). Methodological quality across studies was measured using indicators for the following: (a) student participant information, (b) teacher participant information, (c) language of instruction, (d) study design information, (e) summaries of the dependent and independent variables, (f) description of treatment and measures, (g) interobserver agreement and fidelity, (h) outcome measures, (i) data analysis, and (j) results and findings (indicators adapted from Gersten et al., 2005; Horner et al., 2005; Jitendra, Burgress, & Gajria, 2011; Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005). There were 22 indicators for correlational studies and 18 indicators for experimental studies. Similar indicators for both correlational and experimental studies included sufficient information for the student and teacher participant selection, a plausible rationale and/or research questions for the study, clearly defined measures, and reports of attrition rates and reliability blinding. Descriptions of measure reliability and statistically significant findings were also

evaluated for each research design. Finally, indicators that addressed statistical power and limitations of the study were used for indicator coding.

Indicators specifically for correlational research included reporting of score reliability coefficients, one or more effect sizes, and confidence intervals. Details regarding potential analysis errors that were unique to a particular statistical method were also essential for determining methodological quality (see Appendix A for correlational design coding sheet). For experimental designs, indicators focused on equivalence across groups and whether there was a description of the control condition as well as the treatment condition. Also, indicators addressed whether researchers collected fidelity of implementation information and the timeline for capturing the effects of the treatment (see Appendix B for experimental design coding sheet).

For each study, a methodological quality percentage score was calculated based on whether research indicators were present or not present for both study designs (e.g., statistical significance, measure of fidelity). After coding each article and evaluating methodological quality, a separate evidence rating was determined. Ratings of *compelling* (There is little debate about the truthfulness or value of this evidence), *suggestive* (The truthfulness or value of the study is open to debate on some points), *debatable* (The truthfulness or value can be debated on many points; experts might come to different conclusions) or *weak* (The truthfulness or value is not empirically sound; many instances of variability exists) were used for both correlational and experimental designs (Jitendra et al., 2010; Kratochwill et al., 2010). The criterion for determining the evidence rating involved two main areas of focus. First, coders examined the methodological quality percentage score. Second, coders examined the findings of each study to establish an

overall evidence rating, especially considering the effect size outcomes for experimental studies (based on the distribution of effect sizes for educational outcomes, Hattie (2009) suggests that an effect size of $d = 0.6$ or greater is *large*, $d = 0.4$ to 0.59 is *medium*, and $d = 0.2$ to 0.39 is *small*) (see also Lipsey et al., 2012). Studies with methodological quality scores of 85-100 percent and large effect sizes were considered *compelling*. Studies with methodological quality scores of 69-84 percent and medium to large effect sizes were considered *suggestive*, and studies with methodological quality scores 69-84 percent with small or no effect sizes were considered *debatable*. Studies with methodological scores of 67 percent or below, regardless of effect sizes, were considered *weak*, indicating serious methodological problems in the research.

Inter-rater Reliability

The researcher and a doctoral student who had experience conducting reviews and carrying out experimental research completed all coding. The researcher independently read each article and coded the methodological quality indicators. The doctoral student also followed this procedure for 38 percent of the articles. Interobserver agreement was established comparing scores from each coder to calculate the percentage of agreement (i.e., agreements divided by agreements plus disagreements). Mean percentage agreement was 87% (range = 80-93%).

Findings: Correlational Studies

The demographics of the correlational studies are presented in Table 1. Demographic information includes the authors, measures of reliability and fidelity, study

design, participants, grade level, duration of study, person(s) implementing the intervention, and setting. Studies are organized in ascending order by year.

The two correlational studies included a total of 447 students (males = 200; females = 247), no students with specific learning disabilities, and four students with profound hearing loss. Sample sizes ranged from 6 to 441. Scoring reliability was reported in both studies. Schirmer & Woolsey (1997) reported IOA results of 97.7% on their researcher-developed comprehension questions assessment, but failed to report IOA for their researcher-developed cloze test. Wolf, Crosson, & Resnick (2005) reported IOA results for one standardized measure (76%), as well as results for reliability coefficients (Cronbach's alpha, Spearman's). The duration of these two studies ranged from 21 to 32 sessions, lasting between 30 to 50 min each. The persons implementing treatment consisted of classroom teachers and one teaching-certified researcher.

Table 1

Demographics: Correlational Studies

Study	Study Design	Participants	Duration	Implementation
1. Schirmer & Woolsey (1997) Reliability: Point-by-point IOA Answers to high-level questions = 97.7% (range = NR) <i>Cloze test</i> = NR Fidelity: NR	Correlational	N = 6 males = 2 females = 4 4 students had profound hearing loss; 2 used speech and signing. Age and/or Grade: Ages 10.9-12.5	Daily: 30-45 min Weeks: 9 Sessions: 32	First author (N = 1) Setting: Small reading groups outside the general education classroom
2. Wolf, Crosson, & Resnick (2005) Reliability: Point-by-point IOA Academic Rigor = 76% (range = NR) Cronbach's alpha = .93 Spearman's = .88 ($p < .00$ at $\alpha = .05$) Accountable Talk = NR (range = 57-67%) Cronbach's alpha = .74 to .92 Spearman's = .62 to .83 ($p < .00$ at $\alpha = .05$) Fidelity: NR	Correlational	N = 441 males = 198 females = 243 (SWD = NR) (GE = NR) Age and/or Grade: Grades 1-8: n = NR	Daily: 45-50 min Weeks: -- Sessions: 21 (1 session per teacher)	Classroom Teachers from 10 schools (N = 21) Setting: General Education Classrooms

Note. GE = general education students; IOA = inter-observer agreement; NR = not reported; SWD = students with disabilities.

Characteristics of the correlational studies are presented in Table 2 and include descriptions of the intervention and comparison conditions, dependent variables, and outcomes for student performance.

Both correlational studies examined the relationship between the use of teacher *text* or *script implicit* (high-level questions) and student comprehension outcomes. However, Wolf et al. (2005) reported procedures in which teachers used *text explicit* (low-level) questions to help clarify high-level student responses. The teachers engaged students in low-level responding by asking them yes/no questions or directing them to look back to the text to find factual information. For the purpose of investigating classroom talk for their study, Wolf et al. were primarily interested in the degree to which *text* or *script implicit* questions only helped students engage in higher order thinking skills as opposed to identifying basic information in and outside the text. Therefore, the finding that teachers asked low-level questions and provided low-level content knowledge when students needed support for responding to the high-level questions was noteworthy. In essence, the teachers naturally engaged in a high- to low-level questioning sequence when prompting students to think more critically.

The first dependent variable for students in this study was *academic rigor*, the degree to which students had opportunities for high-level thinking and active use of knowledge. This measure determined the degree to which the teacher talk (i.e., how the teacher facilitated discussion) assisted students in deepening their comprehension of text (rather than by recalling facts via low-level questions). A correlation between *academic rigor* and teacher talk was found (0.79; significant at $\alpha = .01$), suggesting that there is a strong relationship between question type and the degree of rigor of the reading

comprehension lesson. The researchers concluded that students' responses to *text* or *script implicit* questions required a thorough understanding of the text via *text explicit*, or low-level knowledge.

The second dependent variable for students was *[student] accountable talk*, or how students communicated their knowledge. A strong correlation between how teachers asked for knowledge and how students provided knowledge was found (.90; significant at $\alpha = .01$). The researchers concluded that teachers they observed mostly initiated discussion via high-level questions and that students expressed how they arrived at their high-level response through low-level details. Wolf et al. (2005) concluded that there is a relationship between student responses and the types of questions asked by teachers and that it is more likely that students will communicate what they know contingent on how teachers prompt a response.

Table 2

Interventions, Measures, and Outcomes: Correlational Studies

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
<p>1. Schirmer & Woolsey (1997)</p> <p>Purpose: To examine the effect of high-level questions on reading comprehension of deaf children, with the hypothesis that their ability to respond to the high-level questions did not need to be supported by answering low-level questions.</p> <p>T: Using fables, folk tales, or stories written by authors of children's books, the investigator taught a mini-lesson using a modified Directed Reading Thinking Activity, which included:</p> <ol style="list-style-type: none"> 1. Introduction of sight words, vocabulary, idioms, and characters in the story. 2. Teacher-directed questions for students to make predictions and discuss them together. 3. Students read story silently (in segments). 4. After each segment, the investigator asked high-level comprehension questions. 5. After reading and discussing story, students completed a cloze test. 	<p>Implicit, high-level questions only</p>	<p>Standardized: None</p> <p>Researcher Developed: 1. Responses to high-level questions: Scored as correct or incorrect when compared to model answers.</p> <p>2. Cloze test: Required students to provide story details.</p> <p>Other Measures: None</p>	<p>No statistical tests used for each measure; only data for descriptive statistics provided. Calculations determined by author.</p> <p>Responses to high-level questions: No statistical significance M = 79.1, SD = 8.83</p> <p>For the first 3 lessons, students responded correctly more than 87% of the time; during the last 4 lessons, students decreased this performance to 69% (when reading stories above reading level).</p> <p>Cloze test: No statistical significance M = 85.6, SD = 7.42</p> <p>For the first 3 lessons, students responded correctly at least 93% of the time; during the last 4 lessons, students decreased this performance to 82% (when reading stories above reading level).</p> <p>Finding: No correlation coefficient was reported for correct responses to high-level questions and to story cloze for each child. Therefore, the authors collapsed the data of both measures across the 6 students, resulting in statistical significance.</p> <p>Correlation Results Statistical significance = .817</p>

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
<p>Stories started just below or at students' grade level and then were progressively more difficult throughout the study (up to one grade above reading levels).</p> <p>N = 6</p>			<p>(correlation coefficient used was NR)</p> <p>Finding: There was a strong relationship between the ability to answer implicit, high-level questions and the ability to answer questions about story details.</p>
<p>2. Wolf, Crosson, & Resnick (2005)</p> <p>Purpose: To investigate the relationship between the nature of the classroom talk and the degree of the rigor of reading comprehension lessons. Researchers focused on the quality of teacher talk and student talk through the open-ended questions teachers asked during instruction.</p> <p>T: Teachers were instructed to deliver a reading comprehension lesson that was "as typical as possible" that included these three components:</p> <ol style="list-style-type: none"> 1. a text read aloud to, with, or by the students 2. whole-group discussion for 20-min 3. independent work assignments <p>Two raters scored the quality of classroom talk during the lesson.</p>	<p>Implicit, high-level questions only (with variations of high- to low-level questions)</p>	<p>Standardized: Data collected using the Academic Rigor and Accountable Talk rubrics from Instructional Quality Assessment (IQA) tool used for K-12 settings.</p> <p>1. Academic Rigor: To holistically measure the degree to which the treatment assisted students in deepening their comprehension of text, as opposed to recalling, describing, or giving facts.</p> <p>2. Accountable Talk: Measures how teachers asked for student knowledge</p>	<p>A score of 3 reflects high levels of cognitive demands and rigor.</p> <p>Academic Rigor (AR): Statistical significance at $\alpha = .01$ The observed lessons were slightly beyond the level of comprehending the storyline & interpreting the text.</p> <p>Academic Rigor (AR) M = 2.67, SD = .966</p> <p>Finding: In addition to analyzing and interpreting the text, researchers found that responding to implicit, high-level questions required a more thorough understanding of the text via facts and details from the story (i.e., low-level knowledge).</p> <p>Accountable Talk (AT): Statistical significance at $\alpha = .01$ Strategic teacher talk plays an important role in generating interaction with students, especially in how questions are presented.</p> <p>Teachers Asking</p>

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
N = 441		and how students provided knowledge.	for Knowledge (TAK), M = 2.14, SD = .910, R ² = NR
		Researcher Developed: None	Students Providing Knowledge (SPK) M = 3.10, SD = 1.091, R ² = .70
		Other Measures: None	<p>Finding: Researchers found that teachers mostly initiated discussion through implicit, high-level questions and then prompted students to respond to <i>text explicit</i>, low-level questions to illustrate how they arrived at their answer (e.g., often directing students back to the text) or by repeating the low-level knowledge back to the students.</p> <p>Correlations between AR & AT: Statistical significance at $\alpha = .01$ for the relationship between AR and TAK = .79. Statistical significance at $\alpha = .01$ for the relationship between AR and SPK = .84. Statistical significance at $\alpha = .01$ for the relationship between TAK and SPK = .90.</p> <p>Finding: There is a relationship between student responses and the types of questions asked by teachers.</p>

Note. GE = general education students; IOA = inter-observer agreement; M = mean; NR = not reported; SD = standard deviation; SWD = students with disabilities; T = treatment.

Schirmer and Woolsey (1997) proposed that teachers should ask fewer *text explicit* (low-level) questions during instruction and maximize learning through more *text* or *script implicit* (high-level) questions. In their correlational study, they proposed that low-level knowledge may not be necessary for answering implicit, high-level questions, and the only questions needed are those that help students analyze, synthesize, and evaluate the story. Lessons using short stories and *text* or *script implicit* (high-level) questions only were implemented across 9 weeks (32 sessions) with students with profound hearing loss. As the study went on, the stories became increasingly more difficult and above students' reading level. Immediately following each lesson, students responded to high-level assessment questions and completed a cloze test that prompted them to provide story details (i.e., low-level knowledge). The researchers found that there was no relationship between responding to *text explicit* or low-level questions and student performance outcomes (i.e., students' ability to respond to implicit, high-level questions and complete cloze tests). However, the researchers examined the relationship between the ability to answer questions about story details on students' responses to high-level questions and the cloze tests and found a strong correlation (+.817).

Both Schirmer and Woolsey (1997) and Wolf et al. (2005) suggested that *text* or *script implicit* (high-level) questions may lead to stronger student outcomes. In both studies, however, students' higher-order responding was clearly related to their basic understanding of the text. While one might conclude that responding correctly to only high-level questions yields improved achievement, it would be erroneous to conclude that one might respond correctly to implicit, high-level questions without the basic understanding established through *text explicit* (low-level) questioning. In these studies, it

was apparent that the students understood the basic information in the text. It is not clear that this basic understanding was gained through the process of asking implicit, high-level questions or was gained as a result of simply reading the story in their absence.

Findings: Experimental Studies

The demographics of the experimental design studies are presented in Table 3. Demographic information includes the authors, measures of reliability and fidelity, study design, participants, grade level, duration of study, person(s) implementing the intervention, and setting. Studies are organized in ascending order by year.

The six experimental studies included a total of 545 student participants (83 students with disabilities) with sample sizes that ranged from 30 to 180. Five of the six researchers reported that 55% ($n = 180$) of the participants were male and 45% ($n = 149$) were female. In only one study researchers did not report participants' gender. In three studies, researchers reported that participants were in high school (Grades 9-12) (Bulgren et al., 2009; 2013; Lenz et al., 2007), in two studies participants were in seventh grade (Ward-Lonergan et al., 1998; Bulgren et al., 2011), and in one study participants were in fourth grade (Topping & Trickey, 2007). Half of the studies took place in inclusive classrooms ($n = 3$) (Bulgren et al., 2011; 2013; Topping & Trickey, 2007). Locations for the remaining studies were either not reported, in the library, cafeteria, or resource room setting. When reported, the duration of the studies ranged from two to nine weeks, consisting of 25 sessions. The persons implementing the experiments included classroom teachers ($n = 11$), researchers ($n = 2$), and a certified substitute teacher ($n = 1$). Reliability was reported in all six studies with IOA scores ranging from 90.5% to 100%. Treatment

fidelity was reported in two studies (Bulgren et al., 2011; 2013), yet no scores were reported from the fidelity implementation checklists.

Table 3

Demographics: Experimental Studies

Study	Study Design	Participants	Duration	Implementation
<p>1. Ward-Lonergan, Liles, & Anderson (1998)</p> <p>Reliability: IOA accuracy of transcriptions = 100%</p> <p>Point-by-point IOA calculated on students' responses to each question type.</p> <p>Literal Questions: LLD = 95%, WD = 91%</p> <p>Inferential Questions: LLD = 92%, WD = 91%</p> <p>Fidelity: NR</p>	Treatment Comparison	<p>N = 49 males = 49 females = 0</p> <p>T (n = 20) All 20 students had a language-learning disability (LLD)</p> <p>C (n = 29)</p> <p>Age and/or Grade: Grade 7: n = 49 (Ages 12.5-14.7)</p>	<p>Daily: 45-min</p> <p>Weeks: --</p> <p>Sessions: 2 consecutive sessions</p>	<p>Person(s) conducting sessions were NR; Lectures were delivered via video recording</p> <p>Setting: NR</p>
<p>2. Lenz, Adams, Bulgren, Pouliot, & Laroux (2007)</p> <p>Reliability: Point-by-point IOA calculated for 100% of assessments = 100%; no range reported.</p> <p>Fidelity: NR</p>	<p>Treatment Comparison</p> <p>Repeated Measures</p> <p>Pretest-posttest</p>	<p>N = 30 males = 20 females = 10</p> <p>T (n = 10)</p> <p>C (n = 10)</p> <p>All students had a learning disability</p> <p>Age and/or Grade:</p>	<p>Daily: 90-min</p> <p>Weeks: 3</p> <p>Sessions: 15</p> <p>Mon & Fri:</p>	<p>10th grade substitute teacher certified to teach Language Arts (N = 1)</p> <p>Setting: Resource Room</p>

Study	Study Design	Participants	Duration	Implementation
		Grade 9: n = 20 Grade 10: n = 7 Grade 11: n = 2 Grade 12: n = 1	Testing Tues-Thurs: Lessons	
3. Topping & Trickey (2007) Reliability: Range for point-by-point IOA reported for all measures together (85-97%). Fidelity: NR	Treatment Comparison Quasi-Experimental Pretest-posttest	N = 180 T (n = NR) 4 intervention classrooms C (n = NR) 2 comparison classrooms Age and/or Grade: All students age 10	Daily: 30-min Weeks: -- Sessions: 2 Sessions: 1. October 2. May	Classroom teachers (N = 6) Setting: General Education Classroom
4. Bulgren, Marquis, Lenz, Schumaker, & Deshler (2009) Reliability: Point-by-point IOA on student essays: Writing score = 99.1% (range = NR) Content score = 98.3% (range = 80-100%) Fidelity: NR	Treatment Comparison Randomized Control Trial Pretest-posttest	N = 36 (SWD = 18) (GE = 18) T (n = 19) males = 9 females = 10 (SWD = 10) males = 6 females = 4 C (n = 17) males = 9	Daily: 89-min Weeks: 2 Sessions: 2 (5 days apart)	Study researchers (N = NR) Setting: Lesson 1: T Library C Library Lesson 2: T Library C Cafeteria

Study	Study Design	Participants	Duration	Implementation
		females = 8 (SWD = 8) males = 7 females = 1		
		Age and/or Grade: Grade 9: n = 8 Grade 10: n = 27 Grade 12: n = 1		
5. Bulgren, Marquis, Lenz, Deshler, & Schumaker (2011)	Counterbalanced Design	N=116 (SWD = 17) males = 11 females = 6 (GE = 99) males = 50 females = 49	Daily: 50-min Weeks: NR Sessions: 4	Researcher (first author) (N = 1) Setting: General Education Classroom for assigned school subject (science & social studies)
Reliability: Point-by-point IOA calculated for 26.7% of students' Marching & MC assessments (100%) and SA assessments (97.6%)	Repeated Measures			
Fidelity: An Implementation Checklist to ensure all components of instruction were covered was completed and scored with points during each lesson; however, results were not reported.		T (n = NR) C (n = NR) Students were divided into two groups and received both treatments		
		Age and/or Grade: Grade 7: n = 116		
6. Bulgren, Marquis, Deshler, Lenz, & Schumaker (2013)	Treatment Comparison	N = 134 (SWD = 18)	Daily: NR	Classroom Teachers

Study	Study Design	Participants	Duration	Implementation
<p>Reliability: Point-by-point IOA calculated for short answer assessment = 90.5% (range = NR) No IOA calculated for multiple-choice assessment.</p> <p>Fidelity: An Implementation Score Sheet was completed during each lesson to ensure fidelity of implementation; however, results were not reported.</p>		<p>males = 16 females = 2 (GE = 116)</p> <p>males = 48 females = 68</p> <p>T (n = 64) males = 14 females = 41 (SWD = 9)</p> <p>C (n = 52) males = 18 females = 25 (SWD = 9)</p> <p>Age and/or Grade: Grade 9: n = 112 Grade 10: n = 4</p>	<p>Weeks: 9</p> <p>Sessions: NR</p>	<p>(N = 5)</p> <p>Setting: General Education Classroom</p>

Note. GE = general education students; LLD = language learning disability; IOA = inter-observer agreement; MC = multiple-choice; NR = not reported; SWD = students with disabilities

Characteristics of the group experimental studies are presented in Table 4 and include descriptions of the intervention and comparison conditions, dependent variables, and outcomes for student performance. In three of the six experimental studies, researchers used language arts content (Bulgren et al., 2013; Lenz et al., 2007; Topping & Trickey, 2007), in two studies researchers used science content (Bulgren et al., 2009; 2011), and in one study researchers used social studies content (Ward-Lonergan et al., 1998). All researchers assessed student outcomes using researcher-developed measures and only one researcher reported student satisfaction data (Bulgren et al., 2013). Further, the direction of questioning sequences (i.e., low- to high-level; high- to low-level) was reported in five of the six studies. Topping and Trickey (2007) investigated the use of high-level questions only.

In three of the six studies, teachers did not receive training for question levels and were instructed to deliver a “typical lecture discussion” to students in the comparison group (Bulgren et al., 2011; 2013; Topping & Trickey, 2007). In three studies, the teachers told students in the comparison group what information to include in their notes rather than engaging students in a discussion of the content (Bulgren et al., 2011; 2013; Lenz et al., 2007). Comparison students in three studies did not complete their own note guide or graphic organizer (Bulgren et al., 2013; Lenz et al., 2007; Ward-Longergan et al., 1998) when compared with students in the comparison groups of the remaining three studies. Teachers in two studies displayed information on an overhead projector during the lesson with students in the comparison group (Bulgren et al., 2011; 2013).

Effect sizes were reported or sufficient information was provided to calculate effect sizes in four of the six experimental studies (Bulgren et al., 2009; 2011; 2013;

Ward-Lonergan et al., 1998). Across the studies, a total of 26 effect sizes were calculated that ranged from $d = 0.05$ to 3.26. Seventy-four percent ($n = 20$) of the effect sizes came from Bulgren et al.'s (2009; 2011; 2013) research and 18 of the 20 effect sizes that Bulgren et al. reported were *large*. Overall, 81 percent of effect sizes were *large*, 4 percent of effect sizes were *medium*, and 15 percent of effect sizes were *small*. The majority of large effect sizes favored a low- to high-level questioning sequence as part of the treatment condition.

At least one measure of statistical significance favoring the treatment condition was found in each experimental study with 12 instances of statistical significance found overall. In three studies (Ward-Lonergan et al., 1998; Topping & Trickey, 2007; Bulgren et al., 2009) researchers reported at least one finding of no statistical significance.

Bulgren et al. (2013) were the only researchers to use a social validity measure to determine student participant satisfaction with the intervention and the extent to which participants felt prepared for tests. Results ranged from feeling *neutral* to feeling *somewhat satisfied*.

Table 4

Interventions, Measures, and Outcomes: Experimental Studies

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study																																		
<p>1. Ward-Lonergan, Liles, & Anderson (1998)</p> <p>Purpose: To compare the listening comprehension and story recall outcomes for students with language-learning disabilities (LLD) and general education students (GE). Two different types of expository discourse structures, comparison and causation, were implemented individually with all students via two videotaped social studies lectures on a fictitious country called "Lifeland." Sessions were conducted as follows:</p> <ol style="list-style-type: none"> 1. 1-min pre-lecture video 2. 5.5-min comparison or causation video 3. Students answer questions 4. 25-min break 5. 5.5-min of video not previously seen 6. Students answer questions <p>Discourse Structures:</p> <p>Comparison Overall structure contrasted opposing points. Critical elements focused on opposing view, explanation of opposing view, favored view, and explanation of favored view. Comparative topics included housing, education, employment, and population growth in Lifeland.</p> <p>Causation Overall structure consisted of</p>	<p>No linear direction; both low- and high-level questions were mixed together</p> <p>For this study, literal questions were defined as text explicit or low-level; inferential questions were defined as implicit, or high-level.</p>	<p>Standardized: None</p> <p>Researcher Developed: 40-question assessment was developed to measure listening comprehension and recall performance; 20 literal and 20 inferential comprehension questions were written for each discourse structure to make up four question types:</p> <ol style="list-style-type: none"> 1. Comparison literal 2. Comparison inferential 3. Causation literal 4. Causation inferential <p>Other Measures: An analysis was also conducted to measure which group (LLD or GES) performed better.</p>	<p>Means & Standard Deviations GE performed better than LLD on all both question types for both discourse structures</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th style="text-align: center;">LLD n = 20 M (SD)</th> <th style="text-align: center;">GE n = 29 M (SD)</th> </tr> </thead> <tbody> <tr> <td colspan="3">Comparison</td> </tr> <tr> <td>Lit</td> <td style="text-align: center;">6.05 (2.35)</td> <td style="text-align: center;">9.97 (3.32)</td> </tr> <tr> <td></td> <td colspan="2" style="text-align: center;">d = -1.36</td> </tr> <tr> <td>Infer</td> <td style="text-align: center;">4.15 (3.76)</td> <td style="text-align: center;">8.62 (3.90)</td> </tr> <tr> <td></td> <td colspan="2" style="text-align: center;">d = -1.16</td> </tr> <tr> <td colspan="3">Causation</td> </tr> <tr> <td>Lit</td> <td style="text-align: center;">5.95 (3.38)</td> <td style="text-align: center;">10.03 (3.84)</td> </tr> <tr> <td></td> <td colspan="2" style="text-align: center;">d = -1.11</td> </tr> <tr> <td>Infer</td> <td style="text-align: center;">5.70 (3.87)</td> <td style="text-align: center;">9.86 (3.49)</td> </tr> <tr> <td></td> <td colspan="2" style="text-align: center;">d = -1.14</td> </tr> </tbody> </table> <p style="text-align: center;">MANOVA Significant main effect favoring GE F = 23.87, df = 1, 47, p < .001</p>			LLD n = 20 M (SD)	GE n = 29 M (SD)	Comparison			Lit	6.05 (2.35)	9.97 (3.32)		d = -1.36		Infer	4.15 (3.76)	8.62 (3.90)		d = -1.16		Causation			Lit	5.95 (3.38)	10.03 (3.84)		d = -1.11		Infer	5.70 (3.87)	9.86 (3.49)		d = -1.14	
	LLD n = 20 M (SD)	GE n = 29 M (SD)																																			
Comparison																																					
Lit	6.05 (2.35)	9.97 (3.32)																																			
	d = -1.36																																				
Infer	4.15 (3.76)	8.62 (3.90)																																			
	d = -1.16																																				
Causation																																					
Lit	5.95 (3.38)	10.03 (3.84)																																			
	d = -1.11																																				
Infer	5.70 (3.87)	9.86 (3.49)																																			
	d = -1.14																																				

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
<p>antecedents and consequents. Critical elements focused on antecedent, explanation of antecedent, consequent, and explanation of consequent. Causation topics included early inventions, ship building, written language, and architecture.</p> <p>N = 49</p>			<p>d = .34</p> <p>Significant interaction between question types and discourse structures, favoring both groups being more accurate on comparison literal & causation inferential F = 9.51, df = 1, 47, $p < .01$ d = 0.17</p> <p>No significant group interaction by lecture F = .28, df = 1, 47</p> <p>No significant group interaction by question type F = 1.88, df = 1, 47</p> <p>No significant 3-way group by lecture by question type F = 1.14, df = 1, 47</p>
<p>2. Lenz, Adams, Bulgren, Pouliot, & Laroux (2007)</p> <p>Purpose: To examine the effects of the Question Exploration Routine (QER) compared to traditional periodic reviews of repeated information with high-school students with learning disabilities. The QER followed a low- to high-level question instructional framework aimed to increase students' critical thinking skills and resulted in students completing a</p>	<p>Low- to high-level questions (explicit to implicit questioning)</p>	<p>Standardized: None</p> <p>Researcher Developed: Pretests and posttests assessing all 3 lessons (45 total items): matching = 15 fill-in-the-blank = 15</p>	<p>Statistical significance favoring treatment from a one-way ANOVA with repeated measures</p> <p>T Posttest M = 6.43 (SD = 2.54)</p> <p>C Posttest M = 3.27 (SD = 1.80)</p> <p>A Tukey post-hoc analysis showed that students earned higher scores in</p>

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
QEG (Question Exploration Guide) with the teacher. All students participated in each study condition		multiple-choice = 15 Other Measures: None	the QER condition than in the traditional periodic reviews condition, $p < .001$, $d = 1.46$.

T: QER treatment consisted of three phases. First, a guiding question (implicit, high-level question) was posed. Students listed information and other supporting questions that were needed before the guiding question could be answered. The teacher wrote the list on an overhead transparency. Next, three times during the lesson, the teacher stopped and directed students to the list and prompted them to determine if this knowledge had been acquired. Finally, at the end of the lesson, the answer to the guiding question was constructed together while the teacher prompted students to review the list.

C: Traditional periodic review lessons consisted of three phases. First, the teacher stated the objective and topics of the lesson. Next, three times during the lesson, the teacher stopped and reviewed the critical information presented by repeating it. Finally, at the end of the lesson, the teacher again repeated critical information that had been presented in the lesson. Importantly, no visual graphics, organizers, or guiding questions were provided in this condition and the critical information was not visually displayed.

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
<p>Language Arts lessons: 1: Personification 2: Characterization 3: Plot</p> <p>N = 20</p> <p>3. Topping & Trickey (2007)</p> <p>Purpose: To investigate the impact of an intervention package that utilized open-ended questioning (implicit, high-level questions) on students' verbal behavior for the average duration of student utterances and proportion of student talk to teacher talk.</p> <p>T: Teachers received initial and follow-up professional development to learn and implement the "Thinking Through Philosophy" process from The Philosophy for Children program (P4C), whose main feature was the use of open-ended teacher questioning. For both study sessions, the teacher read the same Greek fable out loud to students and asked high-level questions afterwards to explore its meaning (4 treatment classrooms; n = NR).</p> <p>C: Teachers continued to receive regular professional development experiences scheduled for the academic year. For both study sessions, the teacher read the same Greek fable as used with intervention students and implemented "traditional</p>	<p>Implicit, high-level questions only</p>	<p>Standardized: None</p> <p>Researcher Developed: Data from the first 10-min of 30-min video recorded sessions were scored.</p> <p>Measure 1: the amount of time the students talked vs. the amount of time the teacher talked.</p> <p>Measure 2: mean duration of student utterances</p> <p>Other Measures: None</p>	<p>Measure 1: Proportion of student talk to teacher talk Borderline statistical significance favoring treatment</p> <p>T $p = .05$; student talk increased from 41% to 66% due to longer and more elaborated responses, not due to quantity of comments.</p> <p>C $p > .05$; no significant gains</p> <p>Measure 2: Mean duration of student utterances No statistical significance</p> <p>T Pretest (M = .32) Posttest (M = .48) increase in duration</p> <p>C Pretest (M = .24) Posttest (M = .15) decrease in duration</p>

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study										
instruction” techniques using Main Idea Guidelines to explore its meaning (2 comparison classrooms; n = NR).													
N = 180 across T and C													
4. Bulgren, Marquis, Lenz, Schumaker, & Deshler (2009)	Low- to high-level questions	<p>Standardized: None</p> <p>Researcher Developed: 1. Content Knowledge Score: 5 points (rubric scoring system: name problem, cause of problem, effect of problem, solution, main idea statement)</p> <p>2. Writing Scores: 30 points (based on Six Traits of Writing: ideas, organization, voice, word choice, fluency, conventions)</p> <p>Other Measures: None</p>	<p>Content Knowledge: Statistical significance favoring treatment from an analysis of covariance (ANCOVA) F (1, 33) = 15.90, $p < .001$, $d = .74$</p> <p>T Pretest M = 1.63 (SD = 1.54) C Pretest M = 1.88 (SD = 1.32)</p> <p>T Posttest M = 3.16 (SD = 1.83) C Posttest M = 1.71 (SD = 1.49)</p> <p>Content Knowledge: GE Statistical significance favoring treatment F (1, 15) = 17.96, $p = .001$, $d > 2.0$</p> <p>Content Knowledge: SWD No statistical significance F (1, 15) = 1.78, $p = .20$, $d = .69$</p> <table border="0" data-bbox="1577 1133 1875 1235"> <tr> <td></td> <td>GE</td> <td>SWD</td> </tr> <tr> <td></td> <td>n = 18</td> <td>n = 18</td> </tr> <tr> <td></td> <td>M (SD)</td> <td>M (SD)</td> </tr> </table> <p>Pretest T 3.00 (1.00) 0.40 (0.52)</p>			GE	SWD		n = 18	n = 18		M (SD)	M (SD)
	GE	SWD											
	n = 18	n = 18											
	M (SD)	M (SD)											
<p>Purpose: To measure student performance via written expression of content knowledge rather than by multiple-choice and short answer measures as in Bulgren et al. (2002), where a graphic organizer, the Question Exploration Guide (QEG) was used to help students answer a high-level question through the development of low-level questions. The QEG with its associated Question Exploration Routine (QER) was implemented to measure student learning of content. Additionally, researchers examined a combination of the QEG, QER, and of writing prompts to determine if these supports provided the resources needed for students to convey content knowledge in written form.</p> <p>Prestudy: As one group, T and C students were instructed to take notes as they normally would during a 30-min lesson. Students then used notes to complete a pretest essay.</p> <p>T: For the study lesson, implementation of the QER and QEG was identical to the intervention in</p>	(explicit to implicit questioning)												

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study
<p>Bulgren et al. (2002) with two differences. First, the name of the Cue-Do-Review Sequence was changed to Phase 1, 2, and 3; the procedures in each phase were the same in both studies. The researcher used the QER to fully develop the QEG in a 30-min lesson and students were prompted to take notes during this process using a blank QER as guidance. Second, additional instruction was provided after completing the QEG via a 4-min explanation on how the it could be used to write an essay (e.g., how to use information from low-level questions to develop topic paragraphs).</p>			<p>$d = 3.26$</p> <p>C 2.00 (1.23) 1.75 (1.49) $d = 0.18$</p> <p>Posttest</p> <p>T 4.67 (0.71) 1.80 (1.40) $d = 2.59$</p> <p>C 1.67 (1.58) 1.75 (1.49) $d = -0.05$</p>
<p>C: For the study lesson, the QER and QEG were not used with the control group and students did not receive the same researcher-delivered lesson as the experimental group. Instead, students watched the 30-min film on which the prestudy lesson had been based and were instructed to take notes as they normally would. A 4-min lesson on how to write a good 5-paragraph essay was provided afterwards (e.g., how to write a concluding paragraph).</p>			<p>Writing Scores: Statistical significance favoring treatment from an analysis of covariance (ANCOVA) $F(1, 33) = 17.14, p < .001, d = 1.44$</p> <p>T Pretest $M = 2.68$ ($SD = 0.81$) C Pretest $M = 2.70$ ($SD = 0.63$)</p> <p>T Posttest $M = 3.33$ ($SD = 0.93$) C Posttest $M = 2.47$ ($SD = 0.62$)</p>
<p>Science lessons (topics related to the Earth's atmosphere): Prestudy: Depletion of the ozone layer Study: NR $N = 36$</p>			<p>Writing Scores: GE Statistical significance favoring treatment $F(1, 15) = 6.49, p = .022, d = 1.32$</p> <p>Writing Scores: SWD Statistical significance favoring treatment $F(1, 15) = 6.48, p = .022, d = 1.32$</p>

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study							
			GE n = 18 M (SD)	SWD n = 18 M (SD)						
			Pretest							
			T	3.26 (0.50) 2.17 (0.69) d = 1.80						
			C	2.83 (0.65) 2.54 (0.62) d = 0.46						
			Posttest							
			T	3.81 (0.90) 2.90 (0.76) d = 1.09						
			C	2.65 (0.62) 2.27 (0.60) d = 0.62						
<p>5. Bulgren, Marquis, Lenz, Deshler, & Schumaker (2011)</p> <p>Purpose: To measure student knowledge at different levels of thinking as a result of two different instructional methods: the Question Exploration Routine (QER), which included the Question Exploration Guide (QEG), and a traditional lecture-method discussion. All students received both interventions for the same scripted lesson topic. Similar to Bulgren et al. (2009), the purpose of this study was to measure the effects of using a graphic organizer (i.e., QEG) to help students answer a high-level question through the</p>	<p>Low- to high-level questions (explicit to implicit questioning)</p>	<p>Standardized: None</p> <p>Researcher Developed: Content Test for both topics to assess comprehension and retention of information (40 points possible). 16 matching, 20 MC, 4 SA (reported as % correct)</p> <p>Other Measures:</p>	<p>Chemical weapons test: Statistical significance favoring QER treatment vs. the traditional lecture-discussion format via a general linear mixed model analysis. $F(1, 5.7) = 27.8, p = .002, d = 1.42$</p> <table> <thead> <tr> <th data-bbox="1583 1078 1671 1180">QER n = 50 M (SD)</th> <th data-bbox="1772 1078 1875 1180">Lecture n = 66 M (SD)</th> </tr> </thead> <tbody> <tr> <td data-bbox="1562 1224 1703 1248">71.7 (18.6)</td> <td data-bbox="1751 1224 1896 1248">45.9 (16.9)</td> </tr> <tr> <td colspan="2" data-bbox="1667 1256 1770 1281">d = 1.45</td> </tr> </tbody> </table>		QER n = 50 M (SD)	Lecture n = 66 M (SD)	71.7 (18.6)	45.9 (16.9)	d = 1.45	
QER n = 50 M (SD)	Lecture n = 66 M (SD)									
71.7 (18.6)	45.9 (16.9)									
d = 1.45										

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study										
development of low-level questions and also to test knowledge at different levels of thinking.		None	SWD in the QER group scored 15 mean points higher (56) than SWD in the traditional lecture-discussion format (41).										
<p>T: A Cue-Do-Review Sequence adapted from a research report from Bulgren et al. (2002) was implemented to assist the researcher and students to complete the QEG together. Teachers developed a QEG with the students through an interactive process using an overhead projector called the Linking Steps (to enhance understanding and thinking process to finalize the QEG) and the Cue-Do-Review Sequence. Students completed individual QEGs during this process. The focus in the Cue phase was to introduce the QEG and emphasize importance of note taking and participation. The construction of the QEG by the students with teacher guidance took place in the Do phase. Importantly, this phase included the low-level questions that helped answer the high-level question. It also included an additional instructional strategy of six thinking steps to guide students in cognitive processing. The Review phase involved reviewing the QEG, checking students' understanding, and discussing the content to provide an answer to the high-level question.</p>			<p>Biological weapons test: Statistical significance favoring QER treatment vs. the traditional lecture-discussion format via a general linear mixed model analysis. $F(1, 10.2) = 18.7, p = .001,$ $d = 1.16$</p> <table border="0" data-bbox="1554 714 1911 941"> <tr> <td>QER</td> <td>Lecture</td> </tr> <tr> <td>n = 66</td> <td>n = 50</td> </tr> <tr> <td>M (SD)</td> <td>M (SD)</td> </tr> <tr> <td>69.9 (19.9)</td> <td>48.3 (17.5)</td> </tr> <tr> <td colspan="2" style="text-align: center;">$d = 1.15$</td> </tr> </table> <p>SWD in the QER group scored 18 mean points higher (59) than SWD in the traditional lecture-discussion format (41).</p>	QER	Lecture	n = 66	n = 50	M (SD)	M (SD)	69.9 (19.9)	48.3 (17.5)	$d = 1.15$	
QER	Lecture												
n = 66	n = 50												
M (SD)	M (SD)												
69.9 (19.9)	48.3 (17.5)												
$d = 1.15$													
<p>C: The lecture-discussion method involved the distribution of a note-taking sheet and the researcher telling students the information to write down (rather than discussing and completing the</p>													

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study																					
<p>graphic organizer together like for the QEG). Similarly, an overhead projector was used to assist students in copying information onto their note-taking sheets.</p> <p>Lesson and assessment topics:</p> <ol style="list-style-type: none"> 1. Biological weapons 2. Chemical weapons <p>N = 116</p>																								
<p>6. Bulgren, Marquis, Deshler, Lenz, & Schumaker (2013)</p> <p>Purpose: To explore the use of a Question Exploration Routine (QER) on the performance of SWD in inclusive classrooms. A graphic organizer, the Question Exploration Guide (QEG) was used to help students answer a high-level question through the development of low-level questions.</p> <p>T: Teachers developed a QEG with the students following the same instructional treatment as in Bulgren et al., 2011, the Cue-Do-Review Sequence.</p> <p>C: Teachers provided “traditional instruction” using Main Idea Guidelines where students received the same information as in the QEG used in the experimental group. An overhead projector and “typical classroom discussion” was used to guide students towards an understanding of the main idea only. Teachers were asked to conduct the</p>	<p>Low- to high-level questions (explicit to implicit questioning)</p>	<p>Standardized: None</p> <p>Researcher Developed: Two tests to measure student understanding of information in the lessons, each with 16 MC items and 4 short-answer questions.</p> <p>Other Measures: Student Satisfaction Questionnaire (SS) of the QER</p> <p>Student Confidence Questionnaire (SC) for confidence level of preparedness for tests</p>	<p>Prejudice lesson: Statistical significance favoring treatment from a general linear mixed model analysis $F(1, 10.9) = 11.20, p = .007, d = 0.94$</p> <table data-bbox="1570 857 1885 966"> <tr> <td></td> <td>GE</td> <td>SWD</td> </tr> <tr> <td></td> <td>n = 112</td> <td>n = 18</td> </tr> <tr> <td></td> <td>M (SD)</td> <td>M (SD)</td> </tr> </table> <p>Posttest</p> <table data-bbox="1451 1019 1885 1182"> <tr> <td>T</td> <td>.80 (.15)</td> <td>.69 (.23)</td> </tr> <tr> <td></td> <td colspan="2">d = .57</td> </tr> <tr> <td>C</td> <td>.65 (.18)</td> <td>.68 (.15)</td> </tr> <tr> <td></td> <td colspan="2">d = .18</td> </tr> </table> <p>Impetuous Behavior lesson: Statistical significance favoring treatment from a general linear</p>		GE	SWD		n = 112	n = 18		M (SD)	M (SD)	T	.80 (.15)	.69 (.23)		d = .57		C	.65 (.18)	.68 (.15)		d = .18	
	GE	SWD																						
	n = 112	n = 18																						
	M (SD)	M (SD)																						
T	.80 (.15)	.69 (.23)																						
	d = .57																							
C	.65 (.18)	.68 (.15)																						
	d = .18																							

Independent Variable (Teacher questioning)	Direction (Questions)	Dependent Variables (Student outcomes)	Findings from Study	
discussion as they usually did.			mixed model analysis	
Language Arts lessons (topics from <i>Romeo and Juliet</i>):			F (1, 9.32) = 24.27, $p = .0007$, d = 1.23	
1: Prejudice			GE	SWD
2: Impetuous behavior			n = 127	n = 17
N = 134			M (SD)	M (SD)
			Posttest	
			T	.88 (.15) .74 (.25) d = .68
			C	.66 (.19) .57 (.25) d = .41
			Other Measures:	
			SS: Results reported for T only (Likert-type scale of 1-7)	
			M = 4.8: Between “neither satisfied nor dissatisfied” and “somewhat satisfied”	
			SC: (Likert-type scale of 1-7)	
			T M = 5.1	C M = 3.8

Note. C = comparison/control; QEG = Question Exploration Guide; GES = general education students; LLD = language learning disability; M = mean; MC = multiple-choice; NR = not reported; SD = standard deviation; SWD = students with disabilities; T = treatment.

Ward-Lonergan et al. (1998) used low- and high-level questions after a lesson to compare the listening comprehension and story recall outcomes for general education students and students with a language-learning disability. Two different types of expository discourse structures (comparison and causation) were implemented individually with each student, where students listened and received information (i.e. no discussion) via two videotaped social studies lectures on a fictitious country. After completing each session, students responded to 40 assessment questions where half were *text explicit*, or low-level questions, and the other half were *text* or *script implicit* (high level) questions. The questions did not follow a linear direction and instead were mixed together. Mean scores, standard deviations, and some effect sizes were reported. Both groups of students responded more accurately to low-level questions than high-level questions, but general education students outperformed students with language-learning disabilities on both question types for both discourse structures ($d = 0.34$) resulting in a statistically significant main effect. Another effect was a significant interaction between question types and discourse structures ($d = 0.17$). Researchers found that students were more accurate in the comparison lecture for both low- and high-level questions. There were three non-effects reported for no interaction by lecture, no interaction by lesson type, and no interaction of group (students) by lecture by lesson type.

Bulgren and colleagues (2009; 2011; 2013) established a line of research to explore the use of a graphic organizer, the Question Exploration Guide (QEG), to help students answer implicit, high-level questions through the collaborative development of *text explicit* (low-level) questions. The instructional routine in all three studies was the same, with several modifications for each replication. In the 2009 study, teachers

developed a QEG with the students through an interactive process using an overhead projector called the *Linking Steps* (to enhance understanding and the thinking process for completing the QEG) and the *Cue-Do-Review Sequence*. Students completed their own QEGs during this process. The focus in the *Cue* phase was to introduce the QEG and emphasize the importance of note taking and participation. The construction of the QEG by the students with teacher guidance took place in the *Do* phase. Importantly, this phase included the *text explicit*, low-level questions that supported the *text* or *script implicit*, high-level questions. The *Review* phase involved reviewing the QEG, checking students' understanding, and discussing the content to provide an answer to the high-level question. In contrast, teachers in the control condition provided "traditional instruction" using *Main Idea Guidelines* where students received the same information as in the QEG used in the treatment condition. An overhead projector was used to guide students towards an understanding of the main idea only and students did not complete their own graphic organizer. Mean scores, standard deviations, and some effect sizes were reported for measuring students' content knowledge scores and essay writing scores (rubric adapted from the *6-Trait Model of Writing Instruction*). Five of six effects were statistically significant favoring the treatment for both measures. The only non-effect was for students with disabilities on content knowledge when compared with the performance of students without disabilities (i.e., general education students) ($d = .69$). Across both measures, student without disabilities outperformed students with disabilities ($d = 2.59$; $d = 1.09$).

In the second study, Bulgren et al. (2011) replicated the 2009 study with one modification to the *Do* phase where an instructional strategy of six thinking steps was added to the *Cue-Do-Review Sequence* for two science lessons. The six steps were

implemented to guide students in processing information, including prompts like, “Search for supporting questions,” and “Relate the main idea to today’s real world.” In contrast to the 2011 study, students in the control condition only copied information from an overhead projector onto their note-taking sheets and the teacher told the students what information to write down rather than facilitating discussion. Student outcomes were determined using only one measure, a content test with matching, multiple-choice, and short answer items. Mean scores, standard deviations, and some effect sizes were reported. Effects on both lessons were statistically significant favoring the treatment condition (Lesson 1: $d = 1.42$; Lesson 2: $d = 1.16$). The authors did not report descriptive data when comparing outcomes of students with and without disabilities, but did report that students with disabilities in the treatment group (Lesson 1: mean points = 56; Lesson 2: mean points = 59) outperformed other students with disabilities in the control group on both science lessons (Lesson 1: mean points = 41; Lesson 2: mean points = 41).

In the third study, Bulgren et al. (2013) extended their research into the area of language arts with two lessons from *Romeo and Juliet*. This study replicated the 2009 study but did not include the measure to write a short essay aligned with the *6-Trait Model of Writing Instruction*. Mean scores, standard deviations, and some effect sizes were reported. Effects on both lessons were statistically significant favoring the treatment condition (Lesson 1: $d = 0.94$; Lesson 2: $d = 1.23$). The authors also compared outcomes of students with and without disabilities. On both lessons, students without disabilities were outperformed by their classmates with disabilities (Lesson 1: $d = .57$; Lesson 2: $d = .68$). When comparing students with disabilities in the treatment and control conditions, there was no difference in effect for participants in the treatment and the control

conditions for Lesson 1 ($d = .05$) but those in the treatment condition outperformed those in the control condition in Lesson 2 ($d = .68$).

Lenz et al. (2007) adapted Bulgren et al.'s QEG and compared the adapted QEG application with traditional periodic reviews of information. Elements of the QEG were adapted from a research report by Bulgren et al. (2002). All students were in high school and had a learning disability. Instead of completing the QEG via the *Cue-Do-Review Sequence*, students followed a three-phrase instructional routine to gather information, respond to *text explicit* (low-level) questions, and construct a response together with the teacher for the *text or script implicit*, high-level question. Notably, the teacher stopped three times during the lesson to determine if students had the low-level knowledge needed as they moved toward the high-level question. Students in the control condition did not complete a graphic organizer and did not respond to any low-level questions during the lesson. The teacher only stated the objective and topic of the lesson and told students the information they needed to know, often repeating it throughout instruction. One effect of statistical significance from a one-way ANOVA was found favoring treatment. The mean for students with disabilities who used the QEG was $M = 6.43$ while the mean for students with disabilities who did not use the QEG was $M = 3.27$. A Tukey post-hoc analysis also confirmed that scores were higher for those in the QEG condition ($d = 1.46$).

Topping & Trickey (2007) were the only researchers that investigated the effects of *text or script implicit*, high-level questions only on student responding, as measured by the amount of time students talked, or elaborated in their responses, and the mean duration of their utterances. Six classroom teachers received training to implement the

Thinking Through Philosophy process (Cleghorn, 2002) whose main feature was the use of open-ended teacher questioning. For the pretest (October) and posttest (May) conditions, the teachers asked high-level questions after reading the same Greek fable out loud to students. Teachers of students in the control condition continued to receive regular professional development experiences throughout the academic year and used the same Greek fable to implement reading lessons. However, instead of open-ended questions, these teachers used techniques for determining the main idea, exploring meaning, and discussing the text. Descriptive data and results for calculating effect sizes were not reported. The authors reported that there was a statistically significant effect favoring students in the treatment group for the percentage of time they contributed to the discussion (66% from 41%) when compared to the total classroom talk. However, the researchers did not report the statistical method used for their findings. It is likely that student talk in the treatment condition increased not because students commented more often, but because they elaborated more in their responses, indicating that teachers were talking less and asking more open-ended questions. In contrast, students in the comparison groups showed no significant gains. The researchers did not address the issue of scientific control that increased student performance may have been due to practice on the same Greek fable rather than the effects of the intervention. There was also one non-effect of statistical significance for this study when measuring the mean duration of student utterances, even with the increased talk for students in the treatment group ($M = .32$ (pretest); $M = .48$ (posttest)).

Methodological Quality

A list of the qualifying studies and their methodological quality scores and ratings based on correlational and experimental design indicators are presented in Tables 5 and 6. Studies are listed in rank order from highest to lowest overall rating. The mean methodological quality score for the two correlational studies was 47.5% (range = 38-57%) and 72.8% (range = 33-88%) for the experimental design studies.

Table 5

Ratings for Essential Quality Indicators: Correlational Studies

Methodological Quality Indicators	Wolf et al. (2005)	Schirmer & Woolsey (1997)
Rationale, Participants & Setting		
1 Plausible rationale	X	X
2 Student selection is replicable		X
3 Teacher/researcher selection		X
4 Description of physical setting		X
Measurement		
5 Depend variable(s) defined w/quantifiable index		X
6 Reliability coefficients	X	
7 Reliability evidence	X	NA
8 Reliability & validity rationale	X	X
Practical & Clinical Significance		
9 Effect size statistics identified & reported		
10 Interpretation of effect sizes		
11 Authors address limitations	X	X
Potential Analysis Errors		
12 Examination of coefficients	X	
13 Interval data not converted	X	X
14 Univariate methods not in place of outcomes	X	
15 Univariate methods not used post hoc	X	
16 Assumptions of statistical methods	X	
Confidence Intervals		
17 Confidence intervals reported		
18 Confidence intervals reported for effect size(s)		
Data Analysis		
19 Analysis aligned with research questions	X	
20 Attrition reported		
21 Statistically significant findings	X	
22 Adequate statistical power	NA	
Total	12/21	8/21
%	57	38

Note. X = indicator present. Indicators based on criteria proposed by Gersten et al. (2005), and Jitendra et al (2011), and Thompson, Diamond, McWilliam, Snyder, & Snyder (2005).

Table 6

Ratings for Essential Quality Indicators: Experimental Studies

Methodological Quality Indicators	Bulgren et al. (2013)	Lenz et al. (2007)	Bulgren et al. (2011)	Ward-Lonergan et al. (1998)	Bulgren et al. (2009)	Topping & Trickey (2007)
Intro & Participants						
1 Plausible rationale	X	X	X	X	X	X
2 Student selection	X	X	X	X	X	
3 Equivalence of groups	X	X	X	X	X	
4 Teacher selection	X	X				
Study Conditions						
5 IV described & implemented	X	X	X	X	X	
6 Comparison group described	X	X	X	X	X	X
7 Measures defined	X	X	X	X	X	
8 Reliability reported	X	X	X	X	X	X
9 Evidence of blinding						
10 Fidelity reported	X		X			
Measures & Data Analysis						
11 Measures align with intervention	X	X	X	X	X	X
12 Measurement schedule appropriate	X	X	X	X	X	
13 Unit of analysis aligned w/research question(s)	X	X	X	X	X	X
14 Limitations reported	X	X	X	X	X	X
15 Attrition reported						
16 Significant findings	X	X	X	X	X	
17 If not significant, was there adequate statistical power						
18 Statistics and/or effect sizes reported (or data provided)	X	X	X	X	X	
Total %	15/17 88	14/17 82	14/17 82	13/17 76	13/17 76	6/18 33

Note. X = indicator present. Indicators based on criteria proposed by Gersten et al. (2005) and Jitendra et al (2011); IV = independent variable

Correlational Studies

In the two correlational studies, researchers provided an adequate rationale for the need and purpose of the research, but only Schirmer & Woolsey (1997) provided specific details for student and teacher selections as well as a description of the physical setting. Wolf et al. (2005) reported evidence of reliability, correlational coefficients used, and rationale for their measurement, but did not describe their dependent variable with operational precision or the procedure for generating quantifiable results. Schirmer & Woolsey (1997) did operationally define their dependent variable, but did not run statistical tests for each measure, choosing only to report minimal descriptive data. In both studies, researchers addressed the influence of score reliability and validity on their study interpretations.

In terms of practical and clinical significance, both studies failed to identify and report effect sizes. However, the researchers in both studies did address the limitations of their research, pointing out issues of sample size and confounds with initial assessment for selecting student participants. Only Wolf et al. (2005) reported sufficient details for their statistical analysis, addressing potential analysis errors and how they examined their coefficients. In both studies, researchers failed to report confidence intervals and attrition rates.

In addressing findings, Wolf et al.'s (2005) study did produce statistically significant results that aligned with their research questions, but none of the data analysis indicators were present for Schirmer & Woolsey (1997), who failed to align their data analysis with the research questions, produce statistical significance, and ensure enough statistical power.

Experimental Studies

There are strengths and weaknesses in the limited number of experimental studies examined in this literature review that inform the research base for establishing future methodology. Collectively, there were more strengths than weaknesses across the studies. For example, in all six studies, researchers provided a plausible rationale for their research, included a description of treatment for the comparison groups and reported reliability, and used outcome measures that aligned with the intervention and demonstrated that generalizable skills were successfully taught. Further, all researchers employed data analysis techniques linked to their research question(s) and all researchers addressed research limitations.

In five of the six studies, researchers provided operational definitions of the measures used, employed an appropriate schedule to measure outcomes, and included clear descriptions of the independent variable and implemented the treatment as intended, with the exception of Topping and Trickey (2007) who failed to meet these criteria. Again, with the exception of Topping and Trickey (2007) who did not have adequate statistical power, the researchers in the remaining five studies produced statistically significant findings or reported effect sizes or provided enough descriptive data to calculate effect sizes.

Weaknesses in the research primarily stemmed from Topping and Trickey (2007), who, in addition to the issues mentioned above, did not meet the criteria for student and teacher selection or equivalence of groups and risked potential threats to internal validity (e.g., maturation, Hawthorne Effects) due to the length of time they took to measure the effects of their treatment (October to May).

In only two of the six studies, researchers provided specific details for student and teacher participant selection and established equivalence of groups (Bulgren et al., 2013; Lenz et al., 2007). In three studies, researchers met the criteria for these indicators, with the exception of providing sufficient information for teacher or interventionist selection (Bulgren et al., 2009; 2011; Ward-Longergan et al., 1998). None of the researchers reported evidence of blinding or reported attrition rates.

Overall evidence ratings. In Table 7, all studies are presented in rank order by methodological quality percentage scores. Descriptors include the study design, authors and year, quality indicators score, effects and non-effects, and evidence rating.

Table 7

Literature Review: Overall Evidence Ratings

Study Design	(Year) Author	Methodological Quality Score	Findings Effects	Effect Sizes	Evidence Rating
Experimental	(2013) Bulgren et al.	88%	2 effects 0 NE	2 large	<i>Compelling</i>
Experimental	(2007) Lenz et al.	82%	1 effect 0 NE	none	<i>Debatable</i>
Experimental	(2011) Bulgren et al.	82%	2 effects 0 NE	4 large	<i>Suggestive</i>
Experimental	(1998) Ward-Lonergan et al.	76%	2 effects 3 NE	4 large 2 small	<i>Debatable</i>
Experimental	(2009) Bulgren et al.	76%	3 effects 0 NE	11 large 1 medium 2 small	<i>Debatable</i>
Correlational	(2005) Wolf et al.	57%	5 effects 0 NE	1 large	<i>Weak</i>
Correlational	(1997) Schirmer & Woolsey	38%	1 effect 2 NE	none	<i>Weak</i>
Experimental	(2007) Topping & Trickey	33%	1 effect 1 NE	none	<i>Weak</i>

Note. NE = non-effects; *Compelling* = 85-100% and large effect sizes; *Suggestive* = 69-84% and medium to large effect sizes; *Debatable* = 69-84% and small to no effect sizes; *Weak* = 65% or below, regardless of effect sizes.

Only one study (Bulgren et al., 2013) had a *compelling* evidence rating, with the highest methodological quality score (88%) across all studies and large effect sizes on researcher-developed reading comprehension measures, only failing to report evidence of

blinding and attrition rates. Three studies (Bulgren et al., 2009; 2011; Ward-Lonergan et al., 1998) produced strong effect sizes (24) and demonstrated effects (7 effects, 3 non-effects) as a result of their treatment, but earned an evidence rating of *suggestive* due to methodological issues with unclear teacher selection, lack of fidelity measures, and like Bulgren et al. (2013), failed to report evidence of blinding and attrition rates. Lenz et al. (2007) was the only study that had a *debatable* evidence rating due to a lack of effect sizes and one demonstrated effect, even though the methodological percentage score was 82% (no blinding, fidelity, or attrition reported), suggesting that the treatment may not have been effective enough to increase student outcomes. Finally, three studies (Schirmer & Woolsey, 1997; Topping & Trickey, 2007; Wolf et al., 2005) were rated *weak*. Combined, these studies resulted in seven effects and three non-effects for student academic outcomes with one large effect size, but too many methodological indicators essential to research design were not present, especially for student and teacher selection and data analysis procedures.

Discussion

The purpose of this literature review was to examine the relationship between teacher questioning and student achievement from fourth grade through high school. Summaries included in this review highlight the current empirical research base on teacher questioning that align with the CI Model of Text Comprehension (Kintsch, 1988; 1998; 2004; 2013) as well as Raphael and Pearson's (1985) *text explicit, text implicit, and script implicit* taxonomy. The results from correlational studies and intervention research were analyzed and resulted in 17 effects and 6 non-effects across standardized and

researcher-developed comprehension measures. However, due to limitations in methodological quality, primarily for measuring student academic outcomes, the overall body of literature on the effects of teacher questioning on student comprehension outcomes is weak.

Only four studies in this review (Bulgren et al., 2009; 2011; 2013; Ward-Lonergan et al., 1998) were rated as *compelling* (n = 1) or *suggestive* (n = 3). The evidence from these studies suggests that a systematic approach to teacher questioning may lead to increased comprehension for students with initially low comprehension. Bulgren and colleagues (2009; 2011; 2013) used a low- to high-level questioning sequence as part of their treatment to improve student responding to *text* or *script implicit* (high-level) questions while Ward-Lonergan et al. (1998) mixed low- and high-level questions together with no linear direction. In all but one study (Ward-Lonergan et al. 1998), researchers used either visual supports or graphic organizers to support students' retention of information and to assist students in taking notes. These findings should be interpreted cautiously, however, as it is unclear to what extent the outcomes were due to the use of a graphic organizer or questioning sequence. Bulgren and colleagues (2009; 2011; 2013) emphasized that interactive dialogue to discuss content and make connections was critical for helping students engage in higher order thinking skills. In contrast, Ward-Lonergan et al., (1998) mixed high- and low-level questions without other supports and produced primarily large effects on student comprehension. Importantly, they noted that embedding explicit and implicit (low- and high-level) questions within a class discussion may help students think critically about the content.

Limitations and Future Research

Questioning in the classroom has long been considered to be one of the foundational skills for good teaching and learning (Gall, 1970; Bulgren, 2011). However, ongoing experimental research to establish strategies for teacher questioning as an evidence-based practice is scarce (NRP, 2000). Few studies where the level of teacher questioning was manipulated and researchers examined the effects on students' academic outcomes were located for this review. Also, the dependent variables for measuring student outcomes varied greatly. Examinations from this review highlight the need for researchers to look closely at how many facts and ideas students accurately remember after reading text.

More experimental research is needed to understand how strategically sequenced questions may impact student performance, particularly for low comprehenders. Further, a limitation across the studies in this literature review is that researchers did not compare a questioning sequence (low- to high-level; high- to low-level) or a mixed approach (low- and high-level questions) to high-level questions alone.

There is little empirical evidence of the effects of questioning strategies across students with different achievement levels (e.g., average-achievers, students with disabilities). In particular, there is little research for questioning strategies that effectively include low- and high-performing students in discussions while using challenging texts (Shanahan et al., 2012). Determining the instructional impact of questioning on diverse groups of students is important to help teachers individualize instruction and to ask questions at the appropriate learning level for each student within the group.

Limitations regarding methodological quality continue to be a concern in research on teacher questioning, especially in terms of controlling for student and teacher selection and for measuring student outcomes. Although researcher-developed measures are likely to be more sensitive than standardized measures for capturing the effects of a specific intervention, they are difficult to develop and may have biasing results. The variations in the available researcher-developed measures in this literature review make it difficult to interpret the findings.

There is an ongoing need for reliable and valid standardized measures for assessing reading comprehension that have diagnostic qualities and that can also be used to identify text processing differences among students. Keenan, Betjemann, and Olson (2008) found that traditional reading assessments (e.g., cloze procedures, open-ended response formats, multiple-choice) often provide students' decoding or word recognition outcomes, but not necessarily reliable reading comprehension outcomes. Currently, researchers are developing norm-based standardized measures that will reliably distinguish between good and poor comprehenders (August, Francis, Hsu, & Snow; 2006; Bintz, 2000; Carlson, Seipel, McMaster, 2014; Klingner, 2004; Magliano et al., 2011; Pike, Barnes, & Barron, 2010;).

It is possible that not all the available research on teacher questioning was identified for this review. The majority of questioning studies that surfaced while searching the database did not include a measure for student outcomes nor did it require that teachers manipulate question levels as part of the intervention. Also, much of the available research on questioning includes interventions for teaching students how to use self-questioning strategies.

Four of the six experimental studies in this review were from the same authors (Bulgren et al., 2009; 2011; 2013; Lenz et al., 2007). It is critical that other researchers replicate the treatments suggested by Bulgren et al., (2002; 2009; 2011) and Lenz et al., (2007). In addition, it is important to understand if including low-level questions to confirm poor comprehenders' understanding of text helps those students meaningfully engage with challenging grade level text (Shanahan et al., 2012).

More research is needed for determining how to effectively construct and deliver *text explicit*, and *text* or *script implicit* questions to build students' text-based reading comprehension (Bloom & Krathwohl, 1956; Gallagher & Aschner, 1963; Goodwin et al., 1983; Nassaji 2003). Effective teacher questioning would also assist teachers in the task to individualize reading comprehension instruction (Almasi, 2003; Kintsch, 1998; 2004; E. Kintsch, 2005). The study in the following chapter incorporates some of the methodological elements required for experimental research that were lacking in previous studies. Specifically, researcher-developed measures for how many comprehension ideas students remembered from the text (response quantity) and how many of those ideas were accurately remembered (comprehension accuracy) are introduced.

CHAPTER III

METHODS

Participants

Student participants. Eleven fifth-grade students were identified for and completed this study. They were selected from two elementary schools in the coordinating school district. The students were assigned to three reading groups based on their class schedule. Group 1 included two males and one female, Group 2 included four males, and Group 3 included two males and two females. The first language of all student participants was English.

Assessment data for each student are presented in Table 10. These data were used to identify potential student participants and to qualify them for the study. Initially, three primary assessment scores were used for student selection. First, a *Scholastic Reading Inventory* (SRI) Lexile® score (Scholastic, 1999) provided a specific reading comprehension measure for each student. Second, the Reading subtest score from the *Iowa Tests of Basic Skills* (ITBS) taken in students' first month of their fifth-grade year provided nationally normed student achievement results. Finally, the end-of-year fourth grade *Student Assessment of Growth and Excellence* (SAGE) scores for English-Language Arts provided curriculum-based assessment results as normed across the state.

A Lexile® score (e.g., 618L) represents the comprehension level of a reader within a range of 150 points (Lennon & Burdick, 2004). The lower end of that range (100 points below) represents the level at which the student reads independently. The upper

end of that range (50 points above) represents the student's instructional reading level (MetaMetrics, 2012; Scholastic, 1999).

Lexile® scores are also assigned to text as determined by two factors, word frequency and sentence length. Word frequency factors include the occurrence of commonly used words in a variety of contexts as well as the likelihood that the reader had previous contact with words in that text. Sentence length factors are based on the number of words per sentence. If a student's Lexile® score was 1240L, then the appropriate scale for selecting text is 1140L to 1290L (i.e., 100 points below and 50 points above). Thus, when a student's Lexile® score is matched to the Lexile® score of a text, an appropriate reading comprehension level is established. This level is based on a 75% comprehension rate for independently reading text. The comprehension rate increases if the student receives help (Scholastic, 1999). For example, a student with a Lexile® score of 700L is forecasted to comprehend approximately 75% of text with the same Lexile® measure (700L). This approach helps teachers or students select text at an appropriate level of challenge.

In 2012, target Lexile® scores for each grade level were established as part of the Common Core State Standards (CCSS) for English Language Arts (2016 Common Core Standards) (see Table 8). The Lexile® ranges at each grade level represent the *Proficient* level of performance at the end of each grade so the student is on track to be college and career ready upon graduating from high school (National Governors Association, 2012). A fifth-grade student with a Lexile® score of 865L to 980L is considered to be *Proficient* and on grade level for comprehending text (Scholastic, 1999).

Table 8
Year-end Proficiency Lexile® Ranges

Grade	Range
1	100L to 400L
2	450L to 620L
3	610L to 790L
4	770L to 885L
5	865L to 980L
6	955L to 1020L
7	996L to 1060L
8	1039L to 1155L
9	1080L to 1210L
10	1187L to 1305L
11	1215L to 1310L
12	1285L to 1355L

In order for a student to qualify for the study by this measure, a Lexile® score of *Well-below Proficient* (below 599L) based on the SRI Lexile® assessment administered in the third month of their fifth-grade year (see Table 9) was needed. Lexile® scores were available for 10 of the 11 fifth-grade students participating in this study and all were *Well-below Proficient*. Further, each student is identified throughout the study by his or her Lexile® score. Student 000L moved in after the initial SRI Inventory was administered and the cooperating school district did not administer an SRI inventory upon his arrival. The decision to obtain consent for 000L to participate was based on the strong recommendations of the school reading specialist, principal, and classroom teacher.

Table 9

Assessment Scores of Student Participants

	Lexile®	ITBS NPR	SAGE	IEP	R-CBM	CELF-5 NPR
Group 1						
388L (F)	388L	-- --	-- --	no	129 wpm	49 37 th
528L (M)	528L	197 38 th	284 BP	no	113 wpm	49 37 th
533L (M)	533L	221 72 nd	322 BP	no	150 wpm	47 37 th
Group 2						
000L (M)	000L	1 < 1 st	-- BP	yes	60 wpm	33 5 th
317L (M)	317L	175 13 th	291 BP	no	108 wpm	33 5 th
478L (M)	478L	175 13 th	228 BP	yes	101 wpm	46 25 th
527L (M)	527L	212 59 th	274 BP	no	114 wpm	50 37 th
Group 3						
283L (F)	283L	175 13 th	238 BP	no	119 wpm	30 5 th
471L (F)	471L	208 54 th	279 BP	no	139 wpm	57 63 rd
577L (M)	577L	197 38 th	312 BP	no	119 wpm	49 37 th
595L (M)	595L	193 33 rd	327 AP	yes	90 wpm	34 5 th

Note. M = male; F = female; wpm = words per minute; BP = *Below Proficient*; AP = *Approaching Proficient*; NPR = national percentile rank; IEP = Individualized Education Program.

The ITBS and SAGE provide information about participants' reading and English-Language Arts proficiency from a national and state perspective. The ITBS is a nationally normed standardized assessment that tests for mastery of academic skills and provides a standard score (SS) and a national percentile rank (NPR). Ten of eleven students completed the Reading subtest and seven of the ten students had scores below the 50th percentile. This was 388L's first year attending public school and therefore she had not taken these assessments. Further, only the SAGE proficiency level was documented for 000L (i.e., no standard score).

SAGE is the statewide-standardized assessment for math, science, and English. Students receive scores based on four proficiency levels: *Highly Proficient*, *Proficient*, *Approaching Proficient*, and *Below Proficient*. Ten of eleven students (again, scores were not available for 388L) completed the SAGE test in English-Language Arts. All student participants were categorized as *Below Proficient* except 595L (*Approaching Proficient*).

In summary, 11 students participated in this study (see Appendix C for Letter of Informed Consent). Ten of the eleven students qualified for the study with *Well-below Proficient* SRI Lexile® scores. From a national perspective, seven of the participating students had ITBS Reading subtest scores below the 50th percentile. Finally, from a statewide perspective, 10 students scored *Below Proficient* on the SAGE English-Language Arts subtest.

After qualifying student participants using the primary assessment data, three additional scores were obtained: (1) if students had current Individualized Education Program (IEP) as a result of qualifying for special education services, (2) the *Reading-Curriculum Based Measurement* (R-CBM) scores from the *Achievement Improvement*

Monitoring System (AIMSweb) taken halfway through the school year, and (3) the *Recalling Sentences* subtest from the *Clinical Evaluation of Language Fundamentals, Fifth Edition* (CELF-5) (Wiig, Semel, & Secord, 2013).

Three of the eleven student participants (000L, 478L, 595L) qualified to receive special education services under the Individuals with Disabilities Education Act (IDEA) and had a current IEP throughout the duration of the study.

The R-CBM was used to determine students' oral reading fluency on grade-level texts. Fluency serves as the bridge between decoding and reading comprehension; that is, fluency reflects the student's ability to decode words in a text. For this measure, students were prompted to read three different passages aloud for one minute each while the examiner recorded errors and calculated the number of words read correct per minute (Daniel, 2010). According to AIMSweb (2015), a fifth-grade student should be reading at a rate of approximately 128 words per minute at the mid-year assessment. All students participating in this study except three (388L, 533L, 471L) read below the 128 words per minute benchmark.

Since an oral reading comprehension measure was used in this study, the *Recalling Sentences* subtest from the CELF-5 was administered to student participants before beginning the study. The purpose of this assessment was to evaluate each student's ability to recall and orally reproduce sentences of varying length and syntactic complexity in order to determine their spoken language listening comprehension. The CELF-5 has a high degree of test-retest reliability (.90) and a sentence imitation subtest such as this has strong predictive ability in identifying children with language deficits (e.g., Catts, Fey, Zhang, & Tomblin, 2001; Gray, 2004; Nash & Donaldson, 2005;

Tomblin, Zhang, Buckwalter, & O'Brien, 2003). Only one student earned a raw score in the *Average* range (471L) while the remaining students scored *Below Average*. Four students (000L, 317L, 283L, 595L) scored at the 5th NPR, achieving a score that was only higher than 5% of the students in the national norm group (n = 2,380).

Secondary assessment data show that three students had current IEPs at the time of the study, eight students had an average oral reading rate below their peers at mid-year, and 10 students scored below the 40th percentile on the *CELF-5* listening comprehension measure.

Settings

Sessions for this study took place in available settings within the two participating elementary schools based on daily schedules. Sessions for Group 1 took place in three different rooms throughout the study: a portable classroom used for technology and testing, a conference room next to the principal's office, and a small curriculum room used to store books and desks. Fifty percent of sessions were held in the conference room, 35% in the portable classroom, and 15% in the curriculum room. Sessions for Groups 2 and 3 took place in two different rooms throughout the study; an extra classroom used for testing and trainings and a conference room next to the principal's office. Eighty-eight percent of sessions for Group 2 were held in the extra classroom and 12% were held in the conference room. For Group 3, 80% of sessions were held in the extra classroom and 20% were held in the conference room.

Sessions took place during school-wide reading instruction during the same hour each morning. The only individuals present during each session included the student

participants and the primary researcher. For each lesson, students sat in a semicircle facing the primary researcher.

Materials

Materials for this study included scripted reading lessons for each study condition. Other equipment included flipcams and iPads to record each reading session as well as student copies of each narrative story (see Appendix D).

Narrative stories were obtained from the *SRA Reading Laboratory*TM, (Parker, 2004), the *Reading Street*TM curriculum (Afflerbach et al., 2008), and from *Short Story Time*, an online library of short stories (www.short-story-time.com). While the comprehension of both narrative and expository texts is important, narrative texts were selected for this study. Given that the students were fifth graders, they had far more exposure to and a deeper history with narrative texts. Further, comprehension deficits on narrative texts were an ongoing problem for the students participating in this study.

The primary researcher reduced the length of all narrative stories so they ranged from 850 to 950 words and sentence length and vocabulary were modified so the stories generated a Lexile[®] that ranged from 865L to 980L (the *Proficient* Lexile[®] scale for fifth-grade students) (Scholastic, 1999). The Lexile[®] score for each narrative story was determined using the online Lexile[®] Analyzer software (www.lexile.com/analyzer). The average text Lexile[®] score for all 20 narrative stories was 913L and the average number of words per narrative story was 921 (see Table 10).

Table 10

Title, Lexile®, and Word Count for Narrative Stories

Narrative Story	Text Lexile®	Word count
1. Bones for Christmas	870L	943
2. Love Grows	870L	948
3. Mr. Pancake Turkey	870L	942
4. Picnic Food	870L	938
5. Trout Fishing	870L	890
6. Circumstantial Evidence	880L	942
7. Kitchen Table	880L	947
8. What Jo Did	880L	895
9. Chandler's Secret Weapon	890L	882
10. Blood for Chiaka	910L	949
11. A Pet for Sugar	920L	934
12. Lenny the Flying Inventor	920L	935
13. Butterflies are Free	930L	932
14. Amanda and Horace	940L	949
15. Shadow and Carly	940L	944
16. Fences and Friendships	950L	942
17. The House on Maple Street	950L	851
18. The Day I Saw the Ghost	960L	851
19. Lessons on the Ledge	970L	903
20. Cupcake Wars	980L	905
<i>M =</i>	913L	921

Note. Stories are organized by lowest to highest Lexile® score; *M* = Mean.

Figure 3 shows the discrepancy between students' Lexile® scores and the text Lexile® scores used for each lesson, illustrating that student participants read from grade level text throughout the entire study.

two or three high-level questions. Further, a question type (character/trait, event, idea) was only represented once in a question package. Each narrative story consisted of five or six question packages that were distributed throughout the text.

The second set of questions for each story consisted of low- to high-level (*text explicit* to *text* or *script implicit*) questioning sequences (i.e., relational and sequential value between questions). To develop these sequences, five to eight high-level questions that were written for the first set of questions were selected randomly while balancing for the number of question types (i.e., character/trait, event, idea) throughout the story. Next, one to three supporting low-level questions were written for each selected high-level question. Low-level or *text explicit* questions are those where the teacher is seeking literal, direct answers of factual information. These questions engage readers in the initial state of reading comprehension, the *construction* phase, where the readers learn the literal meaning of the text (*microstructure*) (Kintsch, 1988; 1998; 2004; 2013; 2018). Questions at this level are usually either “right or wrong” and involve recall of specific facts that are located directly in the text. For the purposes of this study, the low-level questions were written to increase the probability that students would have the low-level knowledge needed to more thoughtfully respond to the high-level question(s) within the sequence. Examples of low-level questions within a question package from a story in this study are: (1) What are the names of the triplets?, (2) What did the triplets’ mom suggest they do in the yard?, and (3) What picnic assignments were given to the triplets? The high-level question in this sequence is, “How does preparing meals and eating together improve relationships?” Each high-level question with its supporting low-level questions was organized into a question package that included a sequence of three to five total questions

(one to three low-level questions; one to two high-level questions). The second set of question packages (low- to high-level questions) were placed in the same five or six locations in each story as the first set of question packages (high-level questions only). An example of both questioning sequences developed for a narrative story is presented in Table 11 (see Appendix D).

Table 11

*Question Packages:
High-level Questions Only & Low- to High-level Questioning Sequence*

High-level questions only	Low- to high-level questioning sequence
<p>(H) How does the size, color, and weight of wings make a difference for things that can fly?</p> <p>(H) What are some reasons people enjoy spending time with family?</p>	<p>(L) How many butterflies flew past Armida's family?</p> <p>(L) What colors are the Monarch butterflies?</p> <p>(L) What part of the butterflies tickled the family's arms and legs?</p> <p>(H) How does the size, color, and weight of wings make a difference for things that can fly?</p>

Note. Use of the same high-level question for both sequences in bold. Example taken from *Butterflies Are Free*, question package 4. (L) = low-level question (*text explicit*); (H) = high-level question (*script implicit*).

Table 12 shows the percentage of character/trait, event, and idea questions in each story and Table 13 shows the total number of question packages, the total number of low- and high-level questions in each set, the average number of questions per package, and the average number of words between question packages for each story. When stories included only high-level questions, there were approximately twice as many high-level questions as when stories included low- to high-level questioning sequences.

Table 12

High-level Questions: Percentage of Question Type per Narrative Story

Narrative Story	<i>High-level questions only</i>			<i>Low- to high-level question sequences</i>		
	C	E	I	C	E	I
1. Bones for Christmas	25	33	42	33.3	33.3	33.3
2. Love Grows	33.3	33.3	33.3	33.3	33.3	33.3
3. Mr. Pancake Turkey	31	38	31	29	42	29
4. Picnic Food	31	38	31	29	42	29
5. Trout Fishing	36	28	36	33.3	33.3	33.3
6. Circumstantial Evidence	40	40	20	40	40	20
7. Kitchen Table	28	36	36	40	20	40
8. What Jo Did	28	36	36	33.3	33.3	33.3
9. Chandler's Secret Weapon	21	43	36	29	14	57
10. Blood for Chiaka	38	38	24	33.3	33.3	33.3
11. A Pet for Sugar	33.3	33.3	33.3	29	29	42
12. Lenny the Flying Inventor	33.3	33.3	33.3	38	38	24
13. Butterflies are Free	31	31	38	33.3	33.3	33.3
14. Amanda and Horace	30	40	30	20	40	40
15. Shadow and Carly	31	38	31	29	42	29
16. Fences and Friendships	36	36	28	29	42	29
17. The House on Maple Street	36	36	28	33.3	33.3	33.3
18. The Day I Saw the Ghost	30	40	30	33	50	17
19. Lessons on the Ledge	20	30	50	20	40	40
20. Cupcake Wars	31	38	31	29	29	42
Mean Percentage =	31	36	33	31	35	34

Note. C = Character/Trait Question; E = Event Question; I = Idea Question.

Table 13

Properties for High-level Questions Only and Low- to High-level Questioning Sequences

Narrative Story	Qpack	High-level Questions Only		Low- to High-level Questioning Sequences		Total Q		Q per Package (<i>M</i>)		Words (<i>M</i>)
		LL	HL	LL	HL	HO	LH	HO	LH	
		1. Bones	5	0	12	12	6	12	18	
2. Love Grows	6	0	12	11	6	12	17	2.0	2.8	158
3. Pancake	6	0	13	11	7	13	18	2.2	3.0	157
4. Picnic Food	6	0	13	12	7	13	19	2.2	3.2	156
5. Trout Fishing	5	0	11	11	6	11	17	2.2	3.4	178
6. Circumstantial	5	0	10	10	5	10	15	2.0	3.0	188
7. Kitchen Table	5	0	11	12	5	11	17	2.2	3.4	189
8. What Jo Did	6	0	14	12	8	14	20	2.3	3.3	149
9. Chandler's	6	0	14	12	7	14	19	2.3	3.2	147
10. Blood	6	0	13	12	6	13	18	2.2	3.0	158
11. Sugar	6	0	12	10	7	12	17	2.0	2.8	156
12. Lenny	5	0	12	9	8	12	17	2.4	3.4	187
13. Butterflies	6	0	13	13	6	13	19	2.2	3.2	155
14. Amanda	5	0	10	10	5	10	15	2.0	3.0	190
15. Shadow	6	0	13	12	7	13	19	2.2	3.2	157
16. Fences	5	0	11	9	7	11	16	2.2	3.2	188
17. House	5	0	11	10	6	11	16	2.2	3.2	170
18. Ghost	5	0	10	12	6	10	18	2.0	3.6	170
19. Lessons	5	0	10	11	5	10	16	2.0	3.2	181
20. Cupcake	6	0	13	12	7	13	19	2.2	3.2	151
<i>M</i> =	5.5	0	11.9	11.2	6.4	11.9	17.5	2.2	3.2	169

Note. Titles have been shortened. Qpack = Question Package; Q = Questions; LL = Low-level questions (*text explicit*); HL = High-level questions (*text or script implicit*); HO = High-level Questions Only; LH = Low- to High-level Questioning Sequence (*text explicit to text and script implicit*); *M* = Mean.

In summary, the average text Lexile® score for all 20 narrative stories was 913L and the average number of words per narrative story was 921 (see Table 11). Exactly 50 percent of the narrative stories had five total question packages and the remaining 50 percent had six total question packages. The average number of questions in each high-level questions only package was 2.2 and the average number of questions in each low- to high-level questioning sequence package was 3.2. Overall, when only-high level questions were employed, students responded to a mean of 11.9 high-level questions per story (range = 10-14 questions) and no low-level questions. When low- to high-level questions were employed, students responded to a mean of 6.4 high-level questions per story (range = 5-8 questions) and 11.2 low-level questions per story (range = 9-13 questions).

Measures: Dependent Variables

Two comprehension measures were used to evaluate student performance in this study. The first measure was the quantity (i.e., response quantity) of story ideas that students provided when responding to comprehension questions. The second measure was the accuracy (i.e., comprehension accuracy) of each story idea that students provided in their responses.

In addition to the two performance measures, a third measure was used to assess students' interest level for each narrative story. Finally, a social validity measure was administered to each student.

Student Performance

Comprehension Measures. Students responded verbally to three high-level assessment questions for each comprehension measure. One question referenced a character/trait, one question referenced an event, and one question referenced an idea. Table 14 lists the narrative stories and question categories that served as the basis for constructing the assessment questions for each story.

Table 14

Character/Trait, Event, and Idea References for Assessment Questions

Narrative Story	Character/Trait	Event	Idea
1. Bones for Christmas	Helping parents	Giving thoughtful gifts	Providing aid or rescue
2. Love Grows	Good sibling	Spending quality time	Trying something new
3. Mr. Pancake Turkey	Happiness	Selling possessions	Making mistakes
4. Picnic Food	Being nurtured	Selfless service	Needing food to live
5. Trout Fishing	Teasing others	New item didn't work	Proving yourself
6. Circumstantial Evidence	Follow example of parents	Staying up late	Being blamed though innocent
7. Kitchen Table	Anger	Making memories	Accepting change
8. What Jo Did	Known by nickname	Making excuses	Consistent practice
9. Chandler's Secret Weapon	Bravery	Hiking or camping	Haunted places
10. Blood for Chiaka	Providing counsel	Getting sick	Sacrifice
11. A Pet for Sugar	Getting parental permission	Receiving gifts	Ownership
12. Lenny the Flying Inventor	Hard working	Using physical strength	Being adventurous
13. Butterflies are Free	Family love and support	Delay in transportation	Learning to say no
14. Amanda and Horace	Stressed	Being lost	Dreams that seem real
15. Shadow and Carly	Being outgoing	Celebrating others' success	Letting go
16. Fences and Friendships	Being responsible	Getting to know someone	Lessons learned from camps
17. The House on Maple Street	Being friendly	Moving to a new home	Inventing and creating
18. The Day I Saw the Ghost	Curiosity	Doing things by yourself	Saving money
19. Lessons on the Ledge	Protector	Overcoming fear	Using equipment
20. Cupcake Wars	Getting help from parents	Accusing someone	Having a back-up plan

Each assessment question consisted of two parts that draw upon the *text explicit*, *text implicit*, and *script implicit* framework presented by Raphael and Pearson (1985). The first part focused on asking students to describe a personal experience (e.g., *Tell about a time you were a good friend*) (*script implicit* question; requires student to activate schema or background knowledge). The second part of each question directed students to link their personal experience back to the text (i.e., *How does that relate to the story we just read?*) (*text implicit* question; requires the reader to determine what story details support his or her answer to the previous question) (see Figure 4).

Assessment Question:

Tell about any clubs, organizations, or teams that you belong to. (P1) **How does that relate to the story we just read? (P2)**

Student Response:

Well, me and my friends made up a club, um, I forgot the name of it. But it was where we speak, like, a certain type of code. (P1) **And that relates to the story because the boy made up a recycling club at his school. (P2)**

Figure 4. Example of Assessment Question with corresponding Student Response: Part 1 (P1) and **Part 2 (P2)**.

The primary researcher assessed students individually and asked the various question types (character/trait, event, idea) in random order. All student responses were audio recorded and transcribed verbatim by the primary researcher and a research assistant. First, coders determined if students provided information about a personal experience (Part 1). Second, coders determined if students linked their personal experience back to the narrative story (Part 2) (see Appendix E). Importantly, only

information provided about the story (Part 2) was scored for the student performance measures, response quantity and comprehension accuracy.

Comprehension units. The story ideas that students provided in Part 2 were divided into *comprehension units*, or C-units, for scoring. A C-unit is “an independent clause with all subordinate clauses attached to it that cannot be further divided without the disappearance of its essential meaning” (Miller & Iglesias, 2012). Ultimately, it is a statement that contains a subject and a predicate and represents an idea or detail from the story.

Response quantity. First, student responses were coded to determine the quantity of C-units. C-units in a single sentence are separated by a coordinating conjunction (*and*, *but*, *or*, *yet*, *for*, *nor*, *so*) (see Appendix E for the full scoring rubric). For example, the following sentence contains two C-units: “She went to the zoo (1) *and* the lions were sleeping (2).” In this student response, four C-units are provided: “They went fishing early one morning (1). The dad caught the biggest fish (2), *but* the boy didn’t catch anything (3) *so* he was upset (4).” Scores for response quantity were presented as the total number of C-units for each narrative story.

Comprehension accuracy. Second, each C-unit was coded for accuracy. Accurate C-units are when the participant clearly provides details, information, or ideas that align with what happened in the story (see Appendix E for the full scoring rubric). The highest score for accuracy for each student’s response to a question was one point. To calculate the accuracy score for a response, coders divided the number of accurate C-units by the quantity of C-units in each response. For example, if a student provided five C-units when responding to a comprehension question, but only three of those C-units

were accurate, the accuracy score would be 0.6 (3/5) for that question. Since each assessment included three questions, the maximum score for comprehension accuracy was 3 on each comprehension assessment.

Student Interest

The purpose of the student interest measure was to evaluate if students' comprehension scores were higher on stories they preferred (or did not prefer) regardless of the experimental condition. At the end of the study, students individually rated each story: *Liked the Most*, *It was OK*, *Liked the Least*, or *I Don't Remember*. A preference score was calculated for each student for each experimental condition by dividing the number of stories rated as *Liked the Most* by the total number of stories rated. A non-preferred score was calculated for each student for each experimental condition by dividing the total number of stories rated as *Liked the Least* by the total number of stories rated.

Social Validity Measure

A social validity questionnaire was individually administered at the end of the study. This questionnaire prompted students to rate their learning and experience as a result of participating in the study. To minimize any pressure that students might have felt had the primary researcher collected this information, the reading specialist from the school district verbally administered this survey to all student participants (see Appendix F). Percentages were calculated from student ratings on a four-point scale (*No for all stories*, *No for most stories*, *Yes for most stories*, and *Yes for all stories*) and descriptive feedback was evaluated.

Treatment: Independent Variable

The independent variable in this study was the implementation of a low- to high-level questioning sequence within the reading lessons (i.e., moving from *text explicit* to *text* and *script implicit* questions). This was delivered in two phases. In the first phase, the low- to high-level questioning sequences were delivered *without linking prompts*. During the second phase, the low- to high-level questioning sequences were delivered *with linking prompts*. That is, the primary researcher added a *text implicit*, high-level supplemental question to the end of each low- to high-level questioning sequence to help students relate their response back to the narrative story. The supplemental, *text implicit* question was the same as part 2 of the questions on the comprehension assessment delivered after each reading lesson and consisted of: “How does that relate to this story?” (see Table 15).

Table 15

Question Packages for Treatment Conditions

Low- to High-Level Questioning Sequence <i>without linking prompt</i>	Low- to High-Level Questioning Sequence <i>with linking prompt</i>
(L) Why was Mrs. Dobson angry?	(L) Why was Mrs. Dobson angry?
(L) What were some other things Charlie was accused of doing?	(L) What were some other things Charlie was accused of doing?
(L) What did Tommy's mom say they needed to do with Charlie?	(L) What did Tommy's mom say they needed to do with Charlie?
(H) What are the reasons someone might deserve a second chance?	(H) What are the reasons someone might deserve a second chance?
	(LP) How does that relate to this story?

Note. Use of the linking prompt within the questioning sequence in bold. Example taken from first question package of *Circumstantial Evidence*. (L) = low-level question (*text explicit*); (H) = high-level question (*script implicit*); LP = linking prompt.

Experimental Design

The research questions for this study were addressed using a repeated measures design in which students ($n = 11$) were presented reading lessons with *high-level questions only* followed by reading lessons with low- to high-level questioning sequences *without or with linking prompts*. Our study, however, did not utilize a control group, resulting in the possibility of internal and external validity confounds (e.g., history, maturation, testing effects).

Reading lessons were counterbalanced across conditions to minimize order effects and to ensure that groups of students received the reading lessons in a different order. To do this, narrative stories were organized from lowest to highest text Lexile® score and then divided into four quartiles (each consisting of five narrative stories). Then a stratified random sample was created where the schedule for all 20 narrative stories for

each group of students was established by generating a random quartile sequence (e.g., 2-4-3-1, 4-2-1-3). This ensured that there was a balance of text difficulty across conditions (e.g., the most difficult narrative stories were not all in the *high-level questions only* condition. In the study, all 20 narrative stories were represented in the *high-level questions only* condition. Importantly, when students received a reading lesson for a particular narrative story in the *high-level questions only* condition, they did not receive the treatment lesson for that story. Only story 8, *What Jo Did*, appeared in both conditions. The schedule of reading lessons across groups is displayed in Table 16.

Table 16

Lesson Schedule Across Groups of Students

Group 1		Group 2		Group 3	
9	Chandler's	20	Cupcake Wars	8	What Jo Did
16	Fences	4	Picnic Food	11	Sugar
12	Lenny	13	Butterflies	5	Trout Fishing
1	Bones	10	Blood for Chiaka	18	Ghost
17	House	3	Mr. Pancake	6	Circumstantial
7	Kitchen Table	15	Shadow and Carly	14	Amanda and Horace
5	Trout Fishing	8	What Jo Did	2	Love Grows
11	Sugar	14	Amanda and Horace	17	House
3	Mr. Pancake	9	Chandler's	7	Kitchen Table
8	What Jo Did	16	Fences	19	Lessons
19	Lessons	2	Love Grows	20	Cupcake Wars
15	Shadow and Carly	18	Ghost	3	Mr. Pancake
14	Amanda and Horace	6	Circumstantial	1	Bones
6	Circumstantial	11	Sugar	10	Blood for Chiaka
4	Picnic Food	5	Trout Fishing	12	Lenny
20	Cupcake Wars	19	Lessons	16	Fences
10	Blood for Chiaka	17	House	15	Shadow and Carly
13	Butterflies	12	Lenny	4	Picnic Food
18	Ghost	7	Kitchen Table	9	Chandler's
2	Love Grows	1	Bones	13	Butterflies

Note. Most narrative story titles are shortened. Numbers next to story titles represent Lexile® difficulty, lowest to highest.

For the lowest-performing students (n=5) we utilized a multiple baseline design across students. The two study conditions were *high-level questions only* (baseline) and low- to high-level questioning sequences *with linking prompts* (treatment). The multiple baseline design provided for controls for history, length of time in baseline (maturation), and testing effects.

Procedures

High-Level Questions Only condition

Each reading lesson consisted of two sessions, Day 1 and Day 2. Due to the text Lexile® scores being higher than students' Lexile® scores (i.e., students read grade-level material), the students read the story twice to practice for decoding and fluency before responding to comprehension questions throughout the story. The rationale for having the students read the passage multiple times was to ensure that students could (1) simply read the words in the narrative story and (2) build fluency for processing the language of the text (NRP, 2000; Shanahan et al., 2012).

On Day 1, students read the narrative story aloud as a group and no comprehension questions or discussion took place. Students then read the same narrative story aloud for a second time and again, no comprehension questions or discussion took place.

On Day 2, the students read the story a third time and the primary researcher asked the questions within each question package throughout the reading lesson. Only high-level questions were asked during baseline lessons (see Appendix D). The primary researcher generically acknowledged each student's high-level response by saying, "Okay," or repeating their answer (e.g., Question: *What are some reasons grandmas and grandpas are loved so much?* Student: "*Because they're family.*" Primary researcher: "*Okay, because they're family.*") No error correction procedure was implemented for high-level questions. The primary researcher only provided praise that focused on classroom management or to maintain instructional pace and motivation throughout each

reading lesson. Appropriate responses included phrases like, “Okay,” “All right, “ “Good, let’s keep reading,” “Thank you for raising your hand,” and “Nice job being with me today.”

Comprehension Measure

Immediately following the Day 2 reading lesson, the primary researcher administered the comprehension assessment measure to each student individually. There was no time limit for assessment sessions. Overall, students took approximately 3-5 minutes to complete the measure. First, the researcher placed a copy of the three assessment questions (character/trait, event, idea) in front of the student. Second, the researcher provided scripted directions out loud for the comprehension task (see Appendix G). Next, the researcher pointed to the first question and read it out loud and the student provided an oral response. The session continued with the next two questions following this same procedure. The researcher reread a question if requested by the student and the researcher only said phrases like, “Do your best,” and “Thank you,” for feedback. No praise was provided during the assessment sessions and the primary researcher only responded by saying, “Okay,” before moving to the next question. Importantly, the researcher did not prompt the student for more information or clarification when the student finished speaking (e.g., “Anything else?”). Moreover, the researcher asked the three questions in a random order across students.

Low- to High-Level Question Sequences

Instructional sessions for the treatment condition were implemented in the same manner as the instructional sessions for the *high-level questions only* condition. However,

on Day 2, the primary researcher implemented the low- to high-level questioning sequences condition for each story. The primary researcher responded to students' high-level responses in the same manner as in the *high-level questions only* condition and did not ask any additional questions beyond those already scripted for each reading lesson. The primary researcher responded to correct low-level responses by repeating the student's answer to confirm understanding. When students responded incorrectly to low-level questions they were prompted to look back in the story to find the correct answer. The primary researcher then repeated the correct answer before moving on. For partially correct responses, the primary researcher confirmed what was correct and then prompted students to think further and/or look back in the story (e.g., "Yes, the pirates were looking for gold, but keep thinking. There was one more thing they were looking for the most. Look back in the story if you need to."). Again, the primary researcher repeated the correct answer before moving on.

The researcher administered the comprehension measure to each student following each Day 2 reading lesson. The comprehension measure was administered using the same procedures as those employed in the *high-level questions only* condition.

Student Interest

At the end of the study, students provided feedback on their interest level for each narrative story completed during the study. These sessions were conducted one-on-one with the research assistant using category cards and title cards. First, four category cards were placed in front of the student that included *Liked the Most*, *It was OK*, *Liked the Least*, and *I Don't Remember*. Next, title cards were shuffled and placed face down in front of the student. Each title card contained the title of a narrative story completed in

the study. The research assistant then read the directions for the interest survey aloud to the student (see Appendix H). The research assistant selected the top card from the title cards pile and handed the it to the student, read the title aloud, and then prompted the student to place the title card on top of one of the category cards that aligned with his or her rating for that narrative story. This process continued until the student selected an interest category for all applicable narrative stories completed during the study. The research assistant collected each pile and recorded the student's selections on a separate scoring sheet.

Social Validity

Students responded to items on a questionnaire to rate their learning and experience after completing the study. Similar to the student interest procedure, students met one-on-one with the research assistant and provided responses by pointing to rating cards (see Appendix F). For the majority of questions, students were asked to rate their experience across four categories: *No for all stories*, *No for most stories*, *Yes for most stories*, and *Yes for all stories*. The research assistant then read each question item aloud (e.g., "Reading out loud in the group made me nervous.") and the student was directed to point to the category that best captured his or her experience (see Appendix F). This process continued until all question items were answered. The research assistant recorded the student's selections on a separate scoring sheet.

Reliability

Point-by-point interobserver agreement (IOA) was calculated for response quantity and comprehension accuracy and expressed as a percentage (agreements divided

by number of agreements plus number of disagreements multiplied by 100). A second research assistant completed two 2-hour training sessions over two consecutive days before scoring transcriptions. The primary researcher scored 100 percent of the 184 transcriptions from the assessment sessions. The research assistant scored 25 percent of the transcriptions for quantity (46) that were randomly selected across all students (32.6% of the transcriptions (15) came from Group 1, 32.9 percent of the transcriptions (15) came from Group 2, and 34.8 percent of the transcriptions (16) came from Group 3). For comprehension accuracy, the research assistant scored 48% of the transcriptions coded for quantity (22). The research assistant was blind to the group, student, Lexile® difficulty, or condition when scoring transcriptions. To reduce the story comprehension demands on reliability scoring, the 46 transcriptions were rank ordered by text Lexile® (870L-980L) and divided into three scoring groups.

Reliability procedures. First, the second research assistant coded the transcriptions for quantity in the first scoring group. Second, each C-unit was coded as *accurate* or *inaccurate*. Importantly, transcriptions with disagreements between scorers and transcriptions with no C-units for quantity were not used for coding accuracy reliability. After scoring accuracy, a consensus meeting took place with the primary researcher to clarify scoring questions and to refine the scoring rubric before moving on to the next group of transcriptions. This procedure was repeated for the second and third reliability scoring groups (see Appendix E). Since several small adjustments were made in the scoring rubric during consensus meetings, the primary researcher examined and as needed recoded the remaining transcriptions (138) to align with the adjustments that were established for the scoring rubric. Importantly, adjustments were made in only nine of the

138 remaining transcriptions (6.5%). In eight transcriptions, the number of C-units was changed (range = 1-2 C-units per transcription). Only four transcriptions required accuracy adjustments of plus or minus one response. IOA scores by scoring group are summarized in Table 17. The overall mean percent agreement for quantity was 91.8% (range = 50-100%) and the mean percent agreement for accuracy was 87.1% (range 57-100%).

Table 17

IOA Results for Response Quantity and Comprehension Accuracy

	Response Quantity M (range)	Comprehension Accuracy M (range)
Group 1	94.3 (67-100%)	89.6 (66-100%)
Group 2	86.8 (50-100%)	88.2 (57-100%)
Group 3	94.1 (75-100%)	83.8 (60-100%)
Total	91.8 (50-100%)	87.1 (57-100%)

Treatment Fidelity

A fidelity checklist with six intervention components was developed to score the video recorded lessons to determine the extent to which the prescribed reading lesson was administered with fidelity (see Appendix I). At the conclusion of the study, fourteen video recordings (27%) from all the reading lessons (51) were randomly selected across groups of students. An equal number of baseline and intervention lessons were selected for each group of students, resulting in six (30%) lessons from Group 1, four (25%) lessons from Group 2, and four (27%) lessons from Group 3. The primary researcher and the research assistant independently scored all 14 fidelity sessions.

A component was scored “yes” if it occurred with fidelity and “no” if it did not occur with fidelity. The treatment fidelity score is presented as the mean percent agreement and was calculated by dividing the number of correctly implemented intervention components by the total number of possible components times 100. The overall treatment fidelity score for all scored sessions was 91.5% (range 83-100%). The few implementation inaccuracies consisted of adding or omitting a word when asking a question with one occurrence of asking a question package in the wrong location.

A treatment fidelity IOA score was also determined. An agreement was defined as both researchers marking “yes” for the same component or both researchers marking “no” for the same component. A disagreement was defined as one researcher marking “yes” and the other researcher marking “no” for the same component. A mean percent agreement was calculated for the number of agreements and disagreements and expressed as a percentage (agreements divided by number of agreements plus number of

disagreements multiplied by 100). The overall treatment fidelity IOA score was 97.6% (range 83-100%).

CHAPTER IV

RESULTS

The purpose of this study was to evaluate the effectiveness of a low- to high-level questioning sequence on low-performing fifth-grade students' text-based reading comprehension outcomes. Specifically, we examined the extent to which low- to high-level questioning sequences *without or with linking prompts* improved the students' response quantity and comprehension accuracy. In addition, we examined whether the treatment was effective with the lowest-performing students across reading groups and whether the treatment was effective with stories that students identified as high-interest and low-interest. Finally, through a social validity measure, we examined how students rated their experience.

Research Question 1:

To what extent do low- to high-level questioning sequences increase fifth-grade students' response quantity and comprehension accuracy on a post-reading curriculum-based reading comprehension measure?

Response Quantity Results

A one-way repeated-measures analysis of variance (ANOVA) was conducted to evaluate students' response quantity on comprehension assessments following reading lessons in which students responded to *high-level questions only* and following reading lessons in which students responded to low- to high-level questioning sequences *without or with linking prompts*. The within-subjects factor was the study condition (i.e., *high-level questions only* or low- to high-level questioning sequences *without or with linking prompts*), and the dependent variable was the mean number of C-units per

comprehension assessment during each condition. Descriptive data for response quantity, including means and standard deviations, are presented in Table 18. The results for the ANOVA indicated a significant effect, Wilks's $\Lambda = .562$, $F(1,10) = 7.799$, $p < .05$, multivariate $\eta^2 = .438$.

Overall, students provided more comprehension ideas to high-level questions when low- to high-level questioning sequences *without or with linking prompts* were delivered in the reading lesson. These findings support the hypothesis that students provided significantly more C-units on assessments that followed reading lessons with low- to high-level questioning sequences *without or with linking prompts* than on assessments that followed the reading lessons with *high-level questions only*. Further, the large effect size, $\eta^2 = .438$ indicates that 43.8 percent of variance was accounted for by the effect of the low- to high-level questioning sequences.

Table 18

Student Outcomes: Response Quantity and Comprehension Accuracy

	Response Quantity		Comprehension Accuracy	
	HL only	LL-HL	HL only	LL-HL
Group 1				
388L	2.21	2.33	1.46	1.25
528L	6.71	7.17	2.80	2.75
533L	6.14	9.67	2.37	2.79
Group 2				
000L	0.00	3.80	0.00	2.15
317L	0.50	5.83	0.10	2.56
478L	0.90	5.00	0.50	2.65
527L	4.86	7.17	2.07	2.52
Group 3				
283L	7.60	6.00	1.21	2.31
481L	4.89	6.20	1.75	2.90
577L	4.10	4.20	2.10	2.46
595L	4.56	5.00	2.42	2.80
M	3.86	5.67	1.53	2.47
SD	2.60	1.97	0.96	0.46
S _{MD}	0.78	0.59	0.29	0.14
VAR	6.76	3.86	0.93	0.21
Minimum	0.00	2.33	0.00	1.25
Maximum	7.60	9.67	2.80	2.90
Range	7.60	7.34	2.80	1.65

Note. Individual mean scores (*M*) are reported for each student. HL = *high-level questions only*; LL-HL = *low- to high-level questioning sequences without or with linking prompts*; SD = *standard deviation*; S_{MD} = *standard error of the mean*; VAR = *variance of the mean*.

Comprehension Accuracy Results

Similar to response quantity, an ANOVA was conducted to evaluate students' accuracy on comprehension assessments following reading lessons in which students responded to *high-level questions only* and following reading lessons in which students

responded to low- to high-level questioning sequences *without or with linking prompts*. The within-subjects factor was the study condition and the dependent variable was the mean comprehension accuracy score (maximum score of 3.0 per assessment) associated with each condition. Descriptive data for accuracy, including means and standard deviations, are presented in Table 18. The results for the ANOVA indicated a significant effect, Wilks's $\Lambda = .473$, $F(1,10) = 11.124$, $p < .05$, multivariate $\eta^2 = .527$.

Overall, students significantly increased their comprehension accuracy on assessments that followed reading lessons with the low- to high-level questioning sequences *without or with linking prompts*. Similar to the results for quantity, these findings support the hypothesis that students provided significantly more accurate C-units on assessments that followed reading lessons with the low- to high-level questioning sequences *without or with linking prompts* than on assessments that followed the reading lessons with *high-level questions only*. Similar to response quantity, the large effect size, $\eta^2 = .527$ indicates that 52.7 percent of variance was accounted for by the effect of the low- to high-level questioning sequence.

Research Question 1a:

With low-performing students, to what extent do low- to high-level questioning sequences without or with linking prompts increase response quantity and comprehension accuracy on a post-reading curriculum-based reading comprehension assessment?

Response Quantity & Comprehension Accuracy Results: Lowest Performers

The lowest-performing students in the study were identified in a two-step process using the assessment data collected prior to beginning the study (see Table 10). First, students with either no score or a score of *Below Proficient* on the *Student Assessment of*

Growth and Excellence (SAGE) for English-Language Arts were selected for inclusion. Second, students whose percentile rank was *Below Average* (< 22nd) or did not have a score on the Reading subtest from the *Iowa Tests of Basic Skills* (ITBS) and whose percentile rank was *Below Average* (< 50th) on the *Recalling Sentences* subtest from the *Clinical Evaluation of Language Fundamentals* (CELF-5) were selected as the lowest-performing students. Five students qualified; one student each from Groups 1 (388L) and 3 (283L), and three students from Group 2 (000L, 317L, 478L).

Results for response quantity and comprehension accuracy for the lowest-performing students are presented in Figures 5 and 6. The low- to high-level questioning sequences *without linking prompts* phase was only implemented with 388L. Subsequent students received only low- to high-level questioning sequences *with linking prompts* treatment phase.

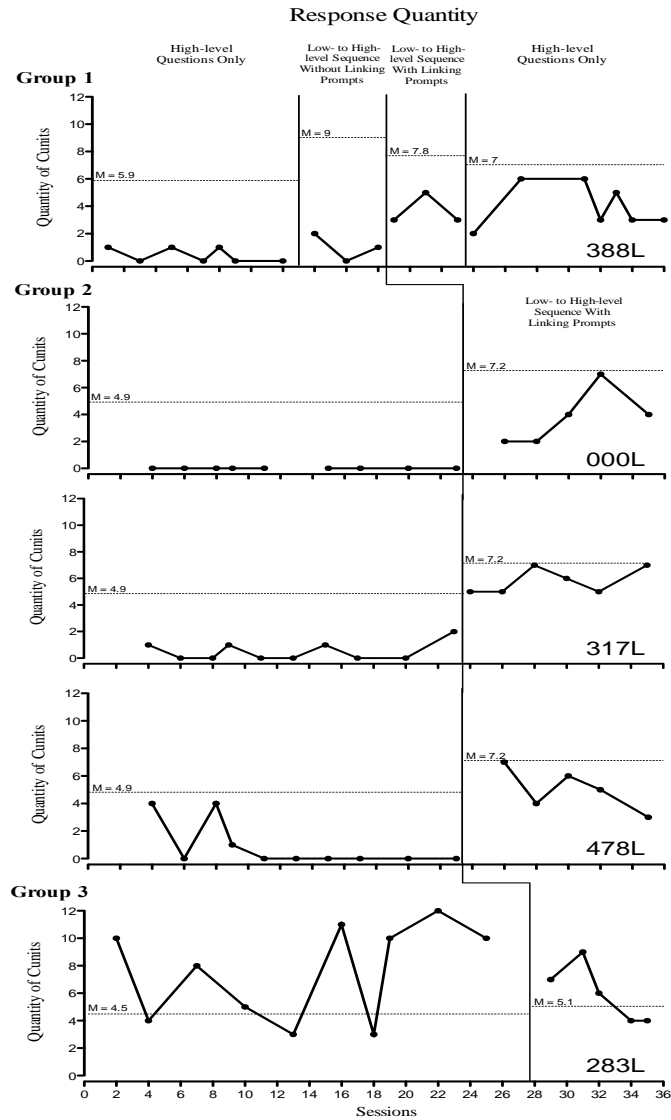


Figure 5. Response quantity results for the lowest-performing students.

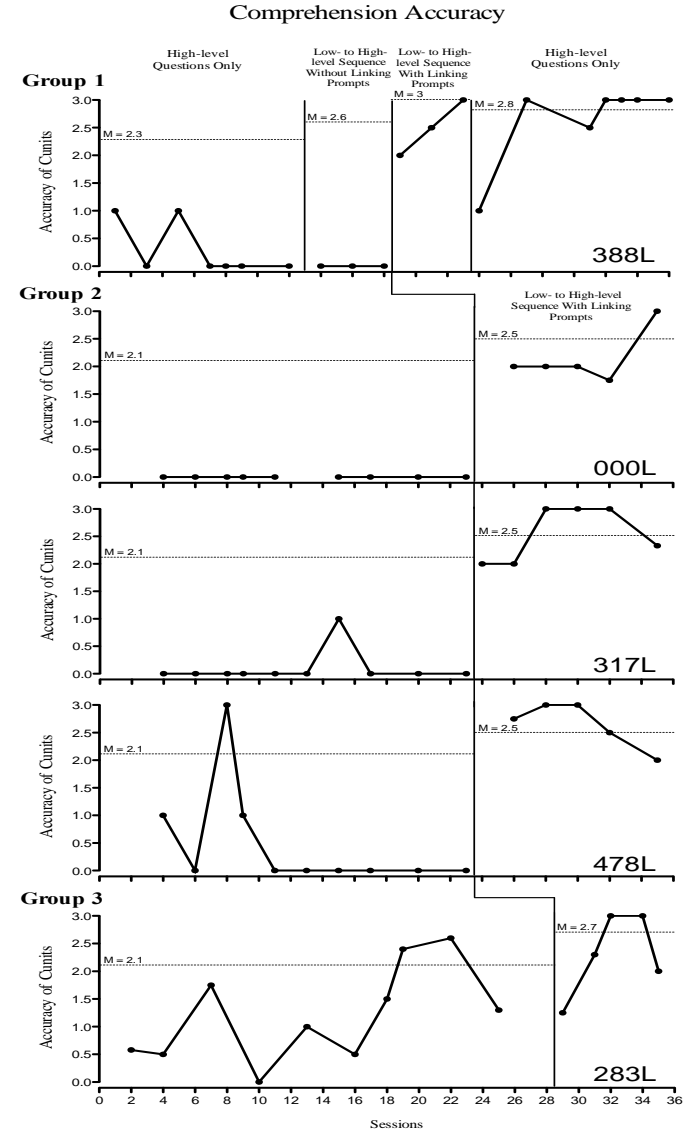


Figure 6. Comprehension accuracy results for the lowest-performing students.

388L. 388L was the only student from Group 1 who qualified as a low performer. For response quantity during the *high-level questions only* condition, 388L established steady state responding between zero and one C-units while the other two students in her group scored an average of 5.9 C-units per comprehension assessment. For comprehension accuracy, 388L's two highest scores were 1.0 while scores for the remaining sessions remained at zero, scoring well below the other two students in the group ($M = 2.3$).

Group 1 received both phases in the low- to high-level questioning sequences condition—*without linking prompts* and *with linking prompts*. In the first treatment phase, *without linking prompts*, 388L initially increased her response quantity to two C-units but then decreased her responding to baseline levels in the next two sessions. In contrast, the other two students in her group increased to an average of nine C-units per comprehension assessment. Of the three total C-units that 388L produced in this phase, none of them were accurate, and again she scored well below the other two students in her group for comprehension accuracy ($M = 2.6$). Since 388L did not improve her performance on the comprehension assessments, a modified treatment was applied to Group 1 that included linking prompts. In the *with linking prompts* phase, the students in Group 1 were provided the same low- to high-level questioning sequences as in the initial treatment phase and responded to a supplemental question at the end of sequence to help them relate their response back to the narrative story (“How does that relate to the story?”). 388L increased her response quantity to between three and five C-units per comprehension assessment. However, her performance continued to be below the average of the other two students in her group ($M = 7.8$). Further, 388L increased her accuracy

score to between 2.0 and 3.0 points, demonstrating a large effect for comprehension accuracy as a result of the treatment. The other two students maintained perfect accuracy scores at 3.0 per comprehension assessment in this phase.

Finally, the *high-level questions only* condition was reinstated and 388L generally maintained the same number of C-units per comprehension assessment as during assessments administered in the *with linking prompts* phase, producing between two and six C-units per story. The other two students in her group produced a mean of 7.0 C-units per comprehension assessment. For comprehension accuracy, 388L's score decreased to a 1.0 in the first session when only high-level questions were asked, but immediately increased and remained stable until the end of the study. Notably, the comprehension accuracy of the other two students in her group decreased slightly to an average of 2.5. This was the only condition throughout the entire study where 388L performed higher than the average of her peers in her group.

Overall, 388L's performance suggests that the addition of linking prompts to the low- to high-level questioning sequences in each narrative story was needed to improve response quantity and comprehension accuracy. In addition, she maintained her performance when the *high-level questions only* condition was reinstated.

000L, 317L, 478L. Three students from Group 2 qualified as low performers. Similar to 388L, all three students produced few C-units on comprehension assessments during the *high-level questions only* condition. 478L's pattern of responding differed slightly from the other two students in his group, initially producing a variable number of C-units before stabilizing his performance at zero C-units, matching the data patterns observed with his low-performing peers. 000L maintained a stable baseline performance

at zero C-units for the entire condition while 317L maintained steady state responding between zero and two C-units. Importantly, these three students maintained their performance level when the treatment was applied to Group 1. The remaining student in their group scored an average of 4.9 C-units during the *high-level questions only* condition. For comprehension accuracy, 000L and 478L's data patterns were nearly identical to their performances for response quantity where 000L maintained a stable trend with accuracy scores of zero and 478L had a variable performance early before producing stable responding with accuracy scores of zero for the remainder of the condition. The remaining student in the group, scored an average of 2.1 points per story on comprehension accuracy during the *high-level questions only* condition.

Due to the school year drawing to a close, only the *with linking prompts* treatment phase was applied to the students in Group 2 when delivering the low- to high-level questioning sequences. A clear level change in this phase was produced by all three students, suggesting that the treatment had an effect on the quantity of ideas that students remembered from the story. 000L increased his performance from zero C-units per story in the *high-level questions only* condition to between two and seven C-units per comprehension assessment in the *with linking prompts* phase. 317L increased his performance to between five and seven C-units and 478L scored between three to seven C-units in this phase. The remaining student in the group scored an average of 7.2 C-units per comprehension assessment. For comprehension accuracy, the three students' accuracy scores ranged from 2.0 to 3.0 points. The level changes for all three low-performing students suggest that the treatment had a strong effect on the accuracy of

ideas that students remembered from each story. The remaining student in the group averaged a score of 2.5 points for each comprehension assessment during this condition.

Overall, the intervention had a similar effect across the performances of the three lowest-performing students in Group 2, suggesting that the low- to high-level questioning sequences *with linking prompts* treatment improved the quantity of C-units on the comprehension assessments. Further, the treatment also produced improved accuracy of C-units on each comprehension assessment.

283L. 283L was the only student from Group 3 who qualified as a low performer. Like Group 2, only the *with linking prompts* treatment phase was applied to the students in Group 3. Overall, the treatment had no effect on the 283L's response quantity and comprehension accuracy throughout the study. Her response pattern was different than students from the other groups. 283L produced between three and twelve C-units during the *high-level questions only* condition. 283L's performance level continued to be variable when treatment was applied to Groups 1 and 2. This highly variable performance continued during the *with linking prompts* phase where she produced a decreasing trend within the range of the previous condition. The remaining three students in her group had a very small increase in the average number of C-units per story from the *high-level questions only* condition ($M = 4.5$) to the *with linking prompt* phase ($M = 5.1$). These results indicate that 283L produced an unpredictable number of ideas from the story regardless of study condition.

For comprehension accuracy, highly variable performance was evident during the *high-level questions only* condition (0.00 to 2.60) similar to her response quantity performance. This continued into the *with linking prompts* phase. However, it is worth

noting that her highest accuracy scores were produced during the *with linking prompts* phase where an increasing trend was demonstrated in the first three sessions before decreasing in a similar manner. In addition, 283L produced fewer C-units during this phase, but the accuracy of those C-units increased, suggesting that the *with linking prompts* phase may have influenced the comprehension accuracy of her responses on the comprehension assessments following each reading lesson. The remaining three students in her group increased their average comprehension accuracy scores from the *high-level questions only* condition ($M = 2.1$) to the *with linking prompts* condition ($M = 2.7$).

Overall, outcomes for response quantity and comprehension accuracy improved for four of the five lowest performers during the treatment in the *with linking prompts* phase. For the remaining student, comprehension accuracy improved in the *with linking prompts* phase.

Research Question 1b:

Given either high-interest or low-interest stories, to what extent do low- to high-level questioning sequences increase fifth-grade students' response quantity and comprehension accuracy?

Student Interest Results

At the end of the study, all students completed an interest survey to determine the extent to which they liked or did not like the narrative stories used in the reading lessons. These sessions were conducted one-on-one using category cards and title cards. Each student's interest ranking for all applicable narrative stories is presented in Table 19. The range of reading lessons completed throughout the study was 15-20. Students in Group 1 provided interest ratings for 20 stories. Students in Group 2 provided interest ratings for

16 stories and students in Group 3 provided interest ratings for 15 stories. All students read 11 of the narrative stories and the students in Groups 1 and 2 read the remaining nine narrative stories. All students except for 528L and 481L used the entire range of categories when rating stories.

Table 19

Interest Survey Results for each Student

Group 1			Group 2				Group 3						
388L	528L	533L	000L	317L	478L	527L	283L	481L	577L	595L			
<i>High-level questions only</i>			<i>High-level questions only</i>				<i>High-level questions only</i>						
Chandler's	✓	✓	✓	Cupcake	✓	+	✓	✓	What Jo Did	✓	--	X	X
Fences	+	✓	X	Picnic Food	--	X	✓	--	Sugar	+	✓	✓	--
Lenny	✓	✓	✓	Butterflies	--	X	✓	✓	Trout Fishing	--	+	--	✓
Bones	+	+	+	Blood	✓	✓	+	--	Ghost	✓	+	X	+
House	--	+	✓	Pancake	+	+	✓	+	Circumstantial	✓	--	X	X
Kitchen Table	+	+	✓	Shadow	X	✓	+	+	Amanda	X	+	X	X
Trout Fishing	X	+	--	What Jo Did	✓	+	✓	--	Love Grows	X	--	✓	X
<i>Low-to high-level questioning without linking prompt</i>			<i>Low-to high-level questioning without linking prompt</i>				<i>Low-to high-level questioning without linking prompt</i>						
Sugar	--	+	--	Amanda	✓	✓	X	--	House	✓	✓	✓	✓
Pancake	--	+	+	Chandler's	+	X	✓	+	Kitchen Table	✓	✓	✓	--
What Jo Did	X	✓	--	Fences	X	X	--	X	Lessons	--	✓	+	✓
<i>Low-to high-level questioning with linking prompt</i>			<i>Low-to high-level questioning with linking prompt</i>				<i>Low-to high-level questioning with linking prompt</i>						
Lessons	✓	✓	+	Love Grows	--	+	+	✓	Cupcake	+	+	+	+
Shadow	X	+	✓	Ghost	+	+	+	+	Pancake	+	✓	+	+
Amanda	✓	+	--	Circumstantial	X	✓	+	X	Bones	+	+	+	✓
<i>High-level questions only</i>			<i>High-level questions only</i>				<i>High-level questions only</i>						
Circumstantial	✓	+	X	Sugar	✓	✓	+	✓	Blood	--	+	✓	--
Picnic Food	X	+	✓	Trout Fishing	--	--	+	--	Lenny	--	✓	+	+
Cupcake	+	+	+	Lessons	+	X	--	✓					
Blood	+	✓	+										
Butterflies	+	✓	✓										
Ghost	+	+	--										
Love Grows	+	+	✓										

Note. Titles are shortened & treatment sessions shaded. Key: Liked the Most (+), It Was OK (✓), Liked the Least (--), I Don't Remember (X).

High-Interest Stories

Only stories that were ranked as *Liked the Most* were used in the high-interest analysis. Students needed to rate at least one story as *Liked the Most* during the *high-level questions only* condition and rate at least one story as *Liked the Most* during the treatment conditions to be included in the high interest analysis. Ten of eleven students met this criterion (388L did not rate any treatment stories as *Liked the Most*). For the *high-level questions only* condition, students selected an average of 50% (range = 20% to 69%) of their completed stories as high interest (see Table 20). Similarly, in the low- to high-level questioning sequences *without or with linking prompts* condition, students selected an average of 50% (range = 31% to 80%) of their completed stories as high-interest. Thus, students did not favor one condition over another in selecting high-interest stories.

Table 20

Percentage of High-interest Stories per Condition for each Student

	High-level questions only	Low- to high-level questioning sequences without or with linking prompts
	Percentage (#) of completed stories	Percentage (#) of completed stories
Group 1		
528L	69% (9)	31% (4)
533L	60% (3)	40% (2)
Group 2		
000L	50% (2)	50% (2)
317L	60% (3)	40% (2)
478L	33% (2)	67% (4)
527L	67% (2)	33% (1)
Group 3		
283L	25% (1)	75% (3)
481L	40% (2)	60% (3)
577L	20% (1)	80% (4)
595L	25% (1)	75% (3)
	<i>M</i> 50% (3.09)	50% (2.55)
	<i>SD</i> 24% (2.77)	24% (1.29)

Note. *M* = Mean; *SD* = Standard Deviation

Table 21 shows the mean response quantity and comprehension accuracy scores for high interest stories for each qualified student by study conditions.

Response quantity performance. Students produced a response quantity mean of 4.26 (*SD* = 2.89) C-units per comprehension assessment for high-interest stories in the *high-level questions only* condition. In the low- to high-level questioning sequences

without or with linking prompts condition, the response quantity mean increased to 6.80 (SD = 3.14) C-units.

Table 21

Response Quantity and Comprehension Accuracy Means and Standard Deviations for High-interest Stories

	Response Quantity		Comprehension Accuracy	
	High-level questions only	Low- to high-level questioning sequences without or with linking prompts	High-level questions only	Low- to high-level questioning sequences without or with linking prompts
528L	7.11	7.75	2.68	2.88
533L	7.33	13.50	2.83	2.63
000L	0.00	3.00	0.00	2.50
317L	0.67	5.00	0.33	2.00
478L	0.50	5.50	0.50	2.81
527L	5.50	10.00	2.25	1.83
283L	4.00	7.33	0.58	2.18
481L	4.50	6.67	1.00	3.00
577L	7.00	3.25	2.33	2.50
595L	6.00	6.00	3.00	3.00
<i>M</i>	4.26	6.80	1.55	2.53
<i>SD</i>	2.89	3.14	1.17	0.41

Note. *M* = Mean; *SD* = Standard Deviation

An ANOVA was conducted to evaluate students' response quantity on comprehension assessments following reading lessons with high-interest narrative stories. The within-subjects factor was the study condition (i.e., *high-level questions only* or low- to high-level questioning sequences *without or with linking prompts*), and the dependent variable was the mean number of C-units per comprehension assessment during each condition. The results for the ANOVA indicated a significant effect, Wilks's $\Lambda = .544$, $F(1,9) = 7.547$, $p < .05$, multivariate $\eta^2 = .456$. The large effect size, $\eta^2 = .456$ indicates that 45.6 percent of variance was accounted for by the effect of the low- to high-level questioning sequence.

Comprehension accuracy performance. Students produced a comprehension accuracy mean of 1.55 (SD = 1.17) per comprehension assessment for high-interest stories in the *high-level questions only* condition. In the low- to high-level questioning sequences *without or with linking prompts* condition, the comprehension accuracy mean increased to 2.53 (SD = 0.41).

Similar to response quantity, an ANOVA was conducted to evaluate students' accuracy on comprehension assessments following reading lessons with high-interest stories. The within-subjects factor was the study condition and the dependent variable was the mean comprehension accuracy score (maximum score of 3.0 per assessment) associated with each condition. The results for the ANOVA indicated a significant effect, Wilks's $\Lambda = .545$, $F(1,9) = 7.525$, $p < .05$, multivariate $\eta^2 = .455$. Similar to response quantity, the large effect size, $\eta^2 = .455$ indicates that 45.5 percent of variance was accounted for by the effect of the low- to high-level questioning sequence.

Low-Interest Stories

Only stories that were ranked as *Liked the Least* were used in the low-interest analysis. Students needed to rate at least one story as *Liked the Least* during the *high-level questions only* condition and rate at least one story as *Liked the Least* during the treatment conditions to be included in the low-interest analysis. Seven of eleven students met this criteria (528L, 317L, 481L, and 577L did not qualify). For the *high-level questions only* condition, students selected an average of 55% (range = 33% to 75%) of their completed stories as low-interest (see Table 22). Similarly, in the low- to high-level questioning sequences *without or with linking prompts* condition, students selected an average of 45% (range = 33% to 67%) of their completed stories as low interest. Students rated slightly fewer narrative stories as low-interest in the treatment condition.

Table 22

Percentage of Low-interest Stories per Condition for each Student

	High-level questions only	Low- to high-level questioning sequences without or with linking prompts
	Percentage (#) of completed stories	Percentage (#) of completed stories
Group 1		
388L	33% (1)	67% (2)
533L	40% (2)	60% (3)
Group 2		
000L	67% (2)	33% (1)
478L	50% (1)	50% (1)
527L	75% (3)	25% (1)
Group 3		
283L	50% (2)	50% (2)
595L	67% (2)	33% (1)
<i>M</i>	55% (1.86)	45% (1.57)
<i>SD</i>	15% (0.69)	15% (0.79)

Note. *M* = Mean; *SD* = Standard Deviation

Response quantity performance. Table 23 shows the mean response quantity and comprehension accuracy scores for low-interest stories for each qualified student by study conditions. Students produced a response quantity mean of 3.45 (*SD* = 3.55) C-units per comprehension assessment for low-interest stories in the *high-level questions only* condition. In the low- to high-level questioning sequences *without or with linking prompt* condition, the response quantity mean increased to 4.71 (*SD* = 2.81) C-units.

Table 23

Response Quantity and Comprehension Accuracy Means and Standard Deviations for Low-interest Stories

	Response Quantity		Comprehension Accuracy	
	High-level questions only	Low- to high-level questioning sequences without or with linking prompts	High-level questions only	Low- to high-level questioning sequences without or with linking prompts
388L	1.00	0.00	0.00	0.00
533L	7.00	8.00	2.75	2.83
000L	0.00	7.00	0.00	1.75
478L	0.00	3.00	0.00	2.00
527L	4.67	7.00	2.17	2.50
283L	9.00	4.00	1.53	2.50
595L	2.50	4.00	2.00	3.00
<i>M</i>	3.45	4.71	1.21	2.08
<i>SD</i>	3.55	2.81	1.18	1.02

Note. *M* = Mean; *SD* = Standard Deviation

Similar to the high-interest analysis, an ANOVA was also conducted to evaluate students' response quantity on comprehension assessments following reading lessons of low-interest narrative stories. The within-subjects factor was the study condition (i.e., *high-level questions only* or *low- to high-level questioning sequence without or with linking prompt*), and the dependent variable was the mean number of C-units per comprehension assessment during each condition. The results for the ANOVA indicated no significant effect, Wilks's $\Lambda = .880$, $F(1,6) = .820$, $p > .05$, multivariate $\eta^2 = .120$. The medium effect size, $\eta^2 = .120$ indicates that only 12.0 percent of variance was accounted for by the effect of the low- to high-level questioning sequence.

Comprehension accuracy performance. Students produced a comprehension accuracy mean of 1.21 (SD = 1.18) per comprehension assessment for low-interest stories in the *high-level questions only* condition. In the low- to high-level questioning sequences *without or with linking prompts* condition, the comprehension accuracy mean increased to 2.08 (SD = 1.02).

An ANOVA was conducted to evaluate students' accuracy on comprehension assessments following reading lessons of low-interest stories. The within-subjects factor was the study condition and the dependent variable was the mean comprehension accuracy score (maximum score of 3.0 per assessment) associated with each condition. The results for the ANOVA indicated a significant effect, Wilks's $\Lambda = .411$, $F(1,6) = 8.588$, $p < .05$, multivariate $\eta^2 = .589$. The large effect size, $\eta^2 = .589$ indicates that 58.9 percent of variance was accounted for by the effect of the treatment.

Overall, students' comprehension accuracy improved during treatment, regardless of their interest level in the story. These findings suggest that students may comprehend text more accurately when a low- to high-level questioning sequence is implemented, regardless of whether they liked or did not like the topic. Similarly, students increased the number of C-units for high-interest stories during the treatment condition, but not for low-interest stories. This finding suggests that students may provide more comprehension details from text they are interested in during the low- to high-level questioning sequences than when only provided high-level questions. In contrast, students may provide fewer comprehension details from text where they do not like the topic.

Research Question 2:

To what extent do student participants rate their overall experience in the study and its impact on their learning and reading ability?

Social Validity Results

A social validity questionnaire was administered to each student in order to gather information regarding the implementation of the study and students' perception of their reading ability as a result of participating in the study. For the majority of questions, students were asked to rate their experience across four categories: *No for all stories*, *No for most stories*, *Yes for most stories*, and *Yes for all stories*. Table 24 shows how students rated each item on the questionnaire. The numbers in bold represent the total number of student ratings for the item. The percentage below each bolded number indicates the proportion of students across all three groups who rated the item in that category.

The students generally liked the topics of the stories, which included topics such as playing sports, family relationships, solving mysteries, and helping animals. Also, the majority of students indicated that they liked reading the stories out loud in small groups and that they followed along while taking turns reading. Overall, the students did not feel that the stories were hard to read or understand and did not feel nervous when reading out loud. It is important to note that students had mixed ratings for leaving their classrooms to participate in the study.

Table 24

Social Validity Questionnaire Results

	YES All	YES Most	NO Most	NO All
Positive Valence Questions				
<i>I liked the topics of the stories, or what the stories were about.</i>		11 100%		
<i>I liked being excused from my classroom for the reading lessons.</i>	5 45%	1 9%		5 45%
<i>I enjoyed reading the stories out loud.</i>	5 45%	3 27%	2 18%	1 9%
<i>I liked reading stories in a small group.</i>	6 55%	5 45%		
<i>I carefully followed along while we took turns reading the stories out loud.</i>	5 45%	5 45%	1 9%	
Negative Valence Questions				
<i>Reading out loud in the group made me nervous.</i>		1 9%	4 36%	6 55%
<i>The stories were hard to read.</i>		2 18%	4 36%	5 45%
<i>The stories were hard to understand.</i>		1 9%	2 18%	8 73%
<i>It was hard to answer questions about my own experiences, opinions, and ideas from the story.</i>	1 9%	4 36%	6 55%	
<i>It was hard when I was asked to relate my own experiences, opinions, and ideas back to what was happening in the story.</i>	1 9%	4 36%	5 45%	1 9%

Note. n = 11 students. Questions for this table have been organized by valence type for the purpose of reporting results as students responded to positive and negative valence questions in random order when completing questionnaire.

When asked to identify which study condition they liked better, 73 percent of the students ($n = 8$) selected the *high-level questions only* condition. However, when asked to identify which study condition helped them remember the stories better, 82 percent of the students ($n = 9$) selected the low- to high-level questioning sequence condition. Finally, all students responded “yes” when asked if they felt like they were better readers after completing the reading lessons.

At the conclusion of the questionnaire, students were also asked to comment on anything else they wanted to say about participating in the study. Four of the eleven students did not have any further comments. For those who did respond, six students stated they felt like they were better readers and had more understanding of what happened in the stories. In addition, five students claimed that participating in the study helped them improve their overall reading ability. Three students stated that the study was “fun” and two students emphasized that they “liked participating a lot.” Finally, one student stated, “I loved the reading lessons and am glad I joined these reading lessons so I could get better at reading.”

CHAPTER IV

DISCUSSION

The purpose of this study was to empirically investigate student achievement as a result of using a low- to high-level questioning strategy *without or with linking prompts* during teacher-lead classroom discourse (see Lynch, 1991; Sinclair & Coulthard, 1975). When the questioning strategy *with linking prompts* was implemented, students' performance increased on two reading comprehension measures, response quantity and comprehension accuracy. This result was also found for four of the five lowest-performing students in the study.

The findings align with previous scholarship that strategically asking questions at different processing levels is one way to help students become proficient in understanding and constructing meaning from text, ultimately promoting deeper comprehension (Almasi, 2003; E. Kintsch, 2005; Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018). Further, understanding how to construct and deliver questions based on type (i.e., *text explicit*, or low-level, *text* or *script implicit*, or high-level) is an instructional skill teachers can incorporate into their pedagogy for building students' text-based reading comprehension (Bloom & Krathwohl, 1956; Gallagher & Aschner, 1963; Goodwin et al., 1983; Nassaji, 2003; Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018). Finally, data from this study support the use of the CI Model of Text Comprehension as a framework for how teachers can use text-based discourse to build comprehension in the classroom (Kintsch, 1988; 1998; 2004; 2013; 2018). That is, how *text explicit* comprehension can lead to improved performance on *text implicit* and *script implicit* comprehension (Raphael & Pearson, 1985).

An additional purpose of this study was to determine whether the implementation of low- to high-level questioning sequences *without or with linking prompts* was effective regardless of students' interest in the narrative story content. For high-interest stories, students significantly increased their response quantity and improved their comprehension accuracy during the low- to high-level questioning sequence *with linking prompts* condition. Similarly, for low-interest stories, students' outcomes for response quantity and comprehension accuracy both increased. In contrast, Belloni & Jongsma (1978) and Stevens (1980) found that students' comprehension increased on high-interest stories and did not improve on low-interest stories. Importantly, the questioning intervention in the present study was effective regardless of students' interest in the story.

Potential Confounds and Limitations

Questioning in the classroom has long been considered to be one of the foundational skills for good teaching and learning (Gall, 1970; Bulgren, 2011). As highlighted in our literature review, experimental research that addresses how teacher-lead questioning impacts student performance is scarce and generally methodologically weak. The experimental procedures and findings from this study contribute to the limited literature base on teacher questioning in the classroom. However, there are potential confounds and limitations within this study that should be addressed in future experiments. They include refining the elements of the low- to high-level questioning sequence (*text explicit to text* and *script implicit* questioning), modifying the student interest procedures, addressing students' opportunities to respond, controlling for potential bias in our researcher-developed comprehension measures, evaluating the

practicality of implementing our study procedures in classrooms, and acknowledging confounds with a single group as part of the research design.

Low- to high-level questioning sequence. The inclusion of the linking prompt (*How does that relate to this story?*) (i.e., *text implicit* question) as part of the low- to high-level questioning sequence was added to the study because the students in Group 1 did not respond to the questioning sequence without this *text implicit* linking prompt. As a result, it is not clear if students in Groups 2 and 3 would have responded to the low- to high-level questioning sequence without the scaffolding provided by the linking prompt. It is possible that students learned to respond correctly only when the linking prompt was present, making it unclear if they would have shown an improvement in comprehension if the linking prompt was not included in the assessment.

Student interest procedures. Although noteworthy, the results for student interest in this study should be interpreted cautiously. It is possible that too much time passed between reading the story and providing an interest ranking (*Like the Most, It Was OK, Liked the Least, I Don't Remember*). When asked about the first five stories that they read, four students indicated that they did not remember an average of two of the stories. In contrast, when asked about the last five stories they read, three students indicated that they did not remember an average of one of the stories. While a similar number of students did not remember stories introduced at the beginning and end of the study, students did not remember more stories from the beginning of the study than at the end of the study (see Table 25 in Appendix J).

In addition, analysis of students' high-interest rankings resulted in a bias toward stories introduced later in the study. For the first five stories, students identified an

average of 1.4 stories as *Liked the Most*. In contrast, for the last five stories, students ranked an average of 2.9 stories as *Liked the Most*. On average, the students identified twice as many high-interest stories at the end of the study than at the beginning of the study (see Table 25 in Appendix J). Thus, conducting the interest survey at the end of the study might have favored the intervention stories. Of course, it is also possible that the intervention helped students remember more stories.

Opportunities to respond. Asking questions at different cognitive levels is an effective instructional strategy that may lead to increases in students' opportunities to respond (Hattie, 2012). In this study, more questions were asked during intervention conditions (low- to high-level questioning sequence *without or with linking prompts*) than in baseline conditions (*high-level questions only*). This suggests that the improvements in student performance, particularly low-performing students, may have been a result of students having more opportunities to respond to more questions instead of the improvements being from the effects of the questioning sequence. In essence, one might argue that it may not be the addition of low-level questions that lead to stronger outcomes, rather, it may be due to the fact that there were more questions overall.

To address this issue, we sampled actual response opportunities for the lowest-performing students in all conditions and found that these students did not have substantially more opportunities to respond during baseline discussions than during intervention discussions (see Table 26 in Appendix K). All students were asked more implicit, high-level questions during the baseline condition than during the intervention condition. Therefore, we propose that simply asking more high-level questions does not produce the same outcomes as asking high-level questions and confirming text

knowledge (i.e., low-level knowledge). We can conclude that asking more high-level questions does not yield findings similar to asking a combination of high- and low-level questions. This conclusion is supported by Gall et al. (1978) who suggested that questioning patterns that include low-level questions (*text explicit*) lead to higher student achievement than questioning patterns that emphasize high-level questions only (*text and script implicit*). This conclusion also supports the CI Model of Text Comprehension where the process of comprehending text requires readers to establish a *microstructure* based on low-level (or *text explicit*) knowledge and to create a *textbase* by organizing that knowledge into higher-level units (*macrostructure*) (Raphael & Pearson, 1985; Kintsch, 1988; 1998; 2004; 2013; 2018). While it appears that it is critical for teachers to confirm students' *text explicit* knowledge (particularly low-performing students) throughout the discussion, it is not clear that low-level questions need to precede high-level questions. Our intervention focused on sequencing from low- to high-level questions, yet similar results might be gained by simply confirming students' *text explicit* knowledge, regardless of the order in which questions are asked. For example, Ward-Lonergan et al. (1998) mixed both low- and high-level questions (i.e., no linear direction), resulting in stronger outcomes for general education students but not for students with a language learning disability.

Comprehension measures. Another limitation in our research lies in the measures used to assess students' reading comprehension and the measures used to identify students with poor comprehension. Unfortunately, there is a lack of standardized text-based reading comprehension measures to use in experimental research, and those that do exist may not be well-developed and may not yield consistent outcomes (Carlo et

al., 2004; Klingner, 2004; Shanahan, Kamil, & Tobin, 1982; Sáenz et al., 2005). For example, some researchers found that the effects of reading comprehension interventions are consistently lower on cloze assessments than on other traditional reading comprehension measures (e.g., short-answer questions, true/false questions) (Carlo et al., 2004; Shanahan et al., 1982). Many standardized assessments for reading comprehension often come from testing batteries, such as the *Comprehensive Reading Assessment Battery* (CRAB) and the *Woodcock-Johnson III* (WJIII) *Comprehension* subtests, where short-answer questions and cloze procedures are the most common assessments within these batteries (Almaguer, 2005; Sáenz et al., 2005; Woodcock, 1991). Further, the available standardized assessments can vary in what aspect of reading comprehension is being measured (e.g., word accuracy, vocabulary, inference) and tend to assess students broadly rather than pinpoint comprehension skills of the student (Nation & Snowling, 1997). Thus, we established a researcher-developed reading comprehension measure for this study to ensure that the outcome measures captured the effects of our questioning intervention. Our measures of response quantity and comprehension accuracy were developed and refined over the course of two pilot studies before the implementation of this study. Nonetheless, we recognize that the potential for bias in our research measures is a limitation and should be taken into consideration when interpreting our findings.

Importantly, there was not a single measure that clearly identified poor readers for our study. While MetaMetrics (2012) suggests that Lexile® scores are an effective tool for identifying good and poor comprehenders, we found that several student participants were not necessarily poor comprehenders based solely on Lexile® results. Therefore, we used multiple assessments (see Table 10) to identify potential student participants for this

study. The students identified as the lowest performers produced low scores on at least four of five of those assessments.

Practicality in the classroom. In this study, students read each story three times to limit the impact of poor decoding and reading fluency on their text-based reading comprehension (NRP, 2000; Shanahan et al., 2012). This approach requires more instructional time than is typically allotted in classrooms. However, a routine of this nature may need to be more prevalent in classrooms in order to provide struggling readers access to challenging text (e.g., grade level text), essential for building robust reading skills (Kuhn, Schwanenflugel, Morris, Morrow, & Woo, 2006; Shanahan et al., 2012). Brown (2015) presents a scaffolded weekly reading schedule that encourages teachers to read the same passage with students over three days to establish text fluency and build reading comprehension. The first day consists of teachers reading the story to students and asking basic comprehension questions (e.g., *Tell me a big idea from the story*) while students silently follow along. The second day involves “echo reading” where students read aloud in unison and respond to more complex questions from the teacher (*How is Tim developing as a character?*). The routine ends with partner reading on the third day where students ask their own comprehension questions to each other (*Do you think Tim did the right thing?*).

In essence, Brown utilized a high- to low-level questioning sequence mapped into the reading routine. That is, teachers asked students for big ideas (i.e., *text implicit* or high-level questions) on the first day while focusing on *text explicit* or low-level comprehension on the second day when students were more fluent with the text. Kuhn et al. (2006) investigated a similar scaffolded approach to improve reading fluency and

found that reading the same material across several days improved students' fluency on a standardized measure (*Gray Oral Reading Test* (4th ed.) [GORT-4]; Wiederholt & Bryant, 2001) (see also Stahl & Heubach, 2005).

Confounds with a single group. We acknowledge that we analyzed student performance in this study as a single cohort (using relevant statistical procedures) without the benefit of a control group. This approach opens our analysis to several potential confounds. First, we compared our quantity and accuracy data during baseline (*implicit*, high-level questions only) with quantity and accuracy data during treatment (low- to high-level questions). While students had different numbers of stories during baseline and treatment, we did not explicitly analyze whether trends for students changed when subgroups of students started treatment. Thus, it is not clear if the performance of students as a group maintained stability when a small subgroup began treatment. More importantly, since we did not include a control group, we could not show (in our pre/post statistical analysis) that another event did not occur at the same time as the treatment and produce the desired outcome.

In our analysis of the low performing students, we controlled for potential confounds that could account for the observed changes in performance by using a single-subject multiple baseline design across the lowest performing students. That is, intervention was first applied to the lowest performing student in Group 1. It was then sequentially applied to low performing students in Group 2 and finally to the lowest performing student in Group 3. It is clear, that the low performing students in Groups 1 and 2 who participated in the study improved their comprehension quantity and accuracy as a function of treatment. There is some question whether the low performing student in

Group 3 improved her performance as a function of treatment since improvements in the student's quality of C-units, and to some extent accuracy of C-units, correlated with implementation of the intervention for the lowest performing students in Group 1.

Future Research

This study acts as a springboard for researchers to continue to investigate teacher questioning in the classroom. There is little empirical research that addresses the effectiveness of strategic teacher questioning on student reading comprehension outcomes since Samson et al.'s (1987) meta-analysis. Specifically, there is little empirical research on how to sequence questions for building reading comprehension, with the exception of Bulgren and colleagues (2009; 2011; 2013) who implemented a low- to high-level questioning routine while using a graphic organizer. Based on the available teacher questioning research and our study, there are at least three critical areas that need further investigation: (1) establishing measures to assess text-based reading comprehension and to identify good and poor comprehenders, (2) refining elements of teacher questioning sequences, and (3) controlling for student interest.

First, we propose that researchers use multiple standardized and researcher-developed measures, as well as different types of reading comprehension measures in future research (see also Klingner, 2004). Only one measure of reading comprehension may be misleading as to the specific reading comprehension skills of students. Multiple assessments provide reading comprehension results in a broader context (e.g., generate a stable score over time) as well as capture the effects of the intervention.

Second, as mentioned previously, it is not clear if students in Groups 2 and 3 of this study would have responded to the low- to high-level questioning sequence without the text implicit *linking prompts*. Importantly, these linking prompts are the same as those used in the comprehension measure. It is possible that only the linking prompt is needed to aid students in this linking task, and therefore, the low- to high-level questioning sequence may not be required. Replications of this study are needed to establish if the linking prompt, the low- to high-level questioning sequence, or both strategies, effect student reading comprehension outcomes. It is also important to further explore questioning sequences (e.g., low- to high-level; high- to low-level) and/or combinations of questions teachers might ask during a reading lesson, as it is not clear that low-level questions need to precede high-level questions as was implemented in this study. In addition, investigating the effects of a low- to high-level questioning sequence in a scaffolded weekly reading schedule similar to Brown's (2015) routine may provide another instructional practice for building text-based reading comprehension. Finally, researchers might design experiments that focus on applying questioning strategies with other text structures (e.g., expository text).

Third, it might be useful to compare the effects of different procedures for assessing student interest in the stories. For example, administering the interest survey prior to the onset of a new study condition would capture student preferences in a timelier manner. Another option would be to standardize the administration of the interest survey across groups of students by having all students rank stories earlier and consistently during the study (e.g., after every third or fourth story), regardless of condition. Other procedural modifications might include reading story titles and abstracts similar to the

procedure Belloni and Jongsma (1978) employed, as compared to only reading story titles as was implemented in this study.

More investigations are needed that examine how student interest interacts with interventions designed to build text-based reading comprehension. In this study, while the intervention was effective with both high-interest and low-interest stories, larger effects were observed on high-interest stories. This suggests that students' interest in a story does impact their reading comprehension. Additional research is needed to gain a deeper understanding of this relationship.

Clearly, research for identifying good and poor comprehenders, measuring reading comprehension outcomes, and developing interventions to improve text-based reading comprehension is in its infancy (NRP, 2000; Carlson et al., 2014; McMaster et al., 2012). Teacher-lead strategies (e.g., questioning levels) is one way to approach reading comprehension instruction. (Raphael & Pearson, 1985; Reutzel, 2014; Reutzel et al., 2005; Kintsch, 1988; 1998; 2004; 2013. 2018). While raising numerous questions, this research provides a foundation for building a meaningful program of research that will ultimately produce interventions that help students engage in higher order thinking skills and improve their text-based reading comprehension.

REFERENCES

**Indicates qualifying studies for literature review*

- Afflerbach, P., Blachowicz, C. L., Boyd, C. D., Cheyney, W., Juel, C., Kame'enui, E. J., & Sebesta, S. (2008). Scott Foresman Reading Street. *Glenview, IL: Pearson Education.*
- Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*, 273-298.
- Alexander, P. A., & Jetton, T. L. (1996). The role of importance and interest in the processing of text. *Educational Psychology Review*, 8(1), 89-121.
- Almasi, J. (2003). *Teaching strategic processes in reading*. New York: Guilford.
- Almaguer, I. (2005). Effects of Dyad Reading Instruction on the Reading Achievement of Hispanic Third-Grade English Language Learners. *Bilingual Research Journal*, 29(3), 509-526.
- Anderson, R.C., & Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. *Handbook of reading research*, 1, 255-291.
- August, D., Francis, D. J., Hsu, H. -Y. A., & Snow, C. E. (2006). *Assessing reading comprehension in bilinguals*. *The Elementary School Journal*, 107, 221–238.
- Baldwin, R. S., Peleg-Bruckner, Z., & McClintock, A. H. (1985). Effects of topic interest and prior knowledge on reading comprehension. *Reading research quarterly*, 497-504.
- Belloni, L. F., & Jongsma, E. A. (1978). The effects of interest on reading comprehension of low-achieving students. *Journal of Reading*, 22(2), 106-109.
- Bintz, W. P. (2000). Using freewriting to assess reading comprehension. *Reading Horizons*, 40(3), 205.
- Bloom, B. S., & Karthwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook 1: Cognitive domain, New York: Longman.
- Borich, G. D. (1980). A needs assessment model for conducting follow-up studies. *Journal of teacher education*, 31(3), 39-42.

- Brown, K. J. (2015, June). *Text Wars: An Unnecessary, Wasteful Polemic*. Invited talk/keynote. Presented to the University of Utah Reading Clinic, College of Education, University of Utah, Salt Lake City, UT. Retrieved from <http://www.uurc.utah.edu/Educators/Presentations/UURC2015-Brown-TextWarsAnUnnecessaryWastefulPolemic.pdf>
- Bulgren, J. A., Lenz, B. K., Marquis, J., Schumaker, J. B., & Deshler, D. D. (2002). *The effects of the use of the question exploration routine on student performance in secondary content classrooms*. (Research Report #10). U.S. Department of Education, Office of Special Education Programs.
- *Bulgren, J. A., Marquis, J. G., Lenz, B. K., Schumaker, J. B., & Deshler, D. D. (2009). Effectiveness of question exploration to enhance students' written expression of content knowledge and comprehension. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 25(4), 271-289.
- *Bulgren, J. A., Marquis, J. G., Lenz, B. K., Deshler, D. D., & Schumaker, J. B. (2011). The effectiveness of a question-exploration routine for enhancing the content learning of secondary students. *Journal of Educational Psychology*, 103(3), 578-593.
- * Bulgren, J. A., Marquis, J. G., Deshler, D. D., Lenz, B. K., & Schumaker, J. B. (2013). The use and effectiveness of a question exploration routine in secondary-level english language arts classrooms. *Learning Disabilities Research & Practice*, 28(4), 156-169.
- Buschman, L. (2001). Using student interviews to guide classroom instruction: An action research project. *Teaching Children Mathematics*, 8(4), 222-227.
- Carlo, M.S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., ... & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39(2), 188-215.
- Carlson, S. E., Seipel, B., & McMaster, K. (2014). Development of a new reading comprehension assessment: Identifying comprehension differences among readers. *Learning and Individual Differences*, 32, 40-53.
- Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, 24(4), 565-591.
- Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures?. *Applied Cognitive Psychology*, 23(6), 747-759.

- Campbell, D. T., Stanley, J. C., & Gage, N. L. (1966). *Experimental and quasi experimental designs for research*. Boston: Houghton Mifflin.
- Caram, C. A., & Davis, P. B. (2005). Inviting student engagement with questioning. *Kappa Delta Pi Record*, 42(1), 19-23.
- Carlson, S. E., van den Broek, P., McMaster, K., Rapp, D. N., Bohn-Gettler, C. M., Kendeou, P., & White, M. J. (2014). Effects of comprehension skills on inference generation during reading. *International Journal of Disability, Development, and Education*, 61(3), 258-274.
- Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, speech, and hearing services in schools*, 32(1), 38-50.
- Cherry, K. (2019, June 26). The role of a schema in psychology. *VeryWell Mind*. Retrieved from <https://www.verywellmind.com/what-is-a-schema-2795873>
- Cleghorn, P. (2002) *Thinking through philosophy*. Blackburn, England: Educational Printing Services.
- Daniel, M. H. (2010). *Reliability of AIMSweb reading curriculum-based measurement (R-CBM) (oral reading fluency)*. New York, NY: Pearson.
- Dantonio, M., & Beisenherz, P. C. (2001). *Learning to question, questioning to learn: Developing effective teacher questioning practices*. Allyn & Bacon.
- Dillon, J. T. (1981). Duration of Response to Teacher Questions. *Contemporary Educational Psychology*, 6, 1-11.
- Duke, N. K., Pearson, P.D., Strachan, S. L., & Billman, A. K. (2011). Essential elements of fostering and teaching reading comprehension. *What research has to say about reading instruction*, 4, 286-314.
- Ellis, K. (1993). Teacher questioning behavior and student learning: What research says to teachers. Paper presented at the Convention of the Western States Communication Association, Albuquerque, NM.
- Ellis, E. S., & Worthington, L. A. (1994). *Research synthesis of effective teaching principles and the design of quality tools for educators*. (Tech Rep. No. 5). Eugene, OR: National Center to Improve the Tools of Educators, University of Oregon.
- Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research*, 40, 707-721.

- Gall, M. D., Ward, B. A., Berliner, D. C., Cahen, L. S., Winne, P. H., Elashoff, J. D., & Stanton, G. C. (1978). Effects of questioning techniques and recitation on student learning. *American Educational Research Journal*, 15(2), 175-199.
- Gallagher, J. J., & Aschner, M. J. (1963). A preliminary report on analyses of classroom interaction. *Merrill-Palmer Quarterly of Behavior and Development*, 9(3), 183-194.
- Gambrell, L. B. (1996). Elementary Students' Motivation to Read. Reading Research Report No. 52.
- Gambrell, L. B. (2011). Seven rules of engagement: What's most important to know about motivation to read. *The Reading Teacher*, 65(3), 172-178.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Council for Exceptional Children*, 71(2), 149-164.
- Gholson, B., & Craig, S. D. (2006). Promoting constructive activities that support vicarious learning during computer-based instruction. *Educational Psychology Review*, 18(2), 119-139.
- Goodwin, S. S., Sharp, G. W., Cloutier, E. F., Diamond, N. A., & Dalgaard, K. A. (1983). *Effective classroom questioning*. Office of Instructional and Management Services, Urbana, IL: University of Illinois.
- Graesser, A. C. (2007). An introduction to strategic reading comprehension. *Reading comprehension strategies: Theories, interventions, and technologies*, 2579, 3-26.
- Graesser, A. C., Person, N. K., & Hu, X. (2002). Improving comprehension through discourse processing. *New directions for teaching and learning*, 2002(89), 33-44.
- Gray, S. (2004). Word Learning by Preschoolers With Specific Language Impairment: Predictors and Poor Learners. *Journal of Speech, Language, and Hearing Research*, 47(5), 1117-1132.
- Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *The Journal of Educational Research*, 99(4), 232-246.
- Haggard, M. R. (1986). The vocabulary self-collection strategy: Using student interest and world knowledge to enhance vocabulary growth. *Journal of Reading*, 29(7), 634-642.

- Hattie, J. A. C. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review*, 1(2), 69-82.
- Hunkins, F. P. (1968). The influence of analysis and evaluation questions on achievement in sixth grade social studies. *Educational Leadership*, 1, 326-332.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional children*, 71(2), 165-179.
- Jitendra, A. K., Burgess, C., Gajria, M. (2011). Cognitive strategy instruction for improving expository text comprehension of students with learning disabilities: The quality of evidence. *Council For Exceptional Children*, 77(2), 135-159.
- 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.
- Kintsch, E. (2005). Comprehension theory as a guide for the design of thoughtful questions. *Topics in Language Disorders*, 25(1), 51-64.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review*, 95(2), 163.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University press.
- Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. *Theoretical models and processes of reading*, 5, 1270-1328.
- Kintsch, W. (2013). Meaning in context. In *Handbook of latent semantic analysis* (pp. 101-118). Psychology Press.
- Kintsch, W. (2018). Revisiting the construction–integration model of text comprehension and its implications for instruction. In *Theoretical Models and Processes of Literacy* (pp. 194-219). Routledge.
- Klingner, J. K. (2004). Assessing reading comprehension. *Assessment for effective intervention*, 29(4), 59-70.

- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What works clearinghouse*.
- Kuhn, M.R., Schwanenflugel, P.J., Morris, R.D., Morrow, L.M., & Woo, D. et al. (2006). Teaching children to become fluent and automatic readers. *Journal of Literacy Research*, 38, 357-387.
- Lennon, C., & Burdick, H. (2004). The Lexile Framework as an approach for reading measurement and success. *MetaMetrics. Inc., Durham*.
- *Lenz, B. K., Adams, G. L., Bulgren, J. A., Pouliot, N., & Laroux, M. (2007). Effects of curriculum maps and guiding questions on the test performance of adolescents with learning disabilities. *Learning Disability Quarterly*, 30(4), 235-244.
- Levin, J., & Nolan, J. (2004). *Principles of classroom management: A professional decision-making model* (4th ed.). Boston: Pearson.
- Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K., & Busick, M.D. (2012). *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms*. Washington, D.C.: *National Center for Special Education Research*.
- Lynch, T. (1991). Questioning roles in the classroom. *ELT Journal*, 45, 201–210.
- Magliano, J. P., Millis, K. K., Development Team, R. S. A. T., Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning*, 6, 131–154.
- Mangano, N. G., & Benton, S. L. (1984). Comparison of question-response-feedback interactions during basal reader instruction. *The Journal of Educational Research*, 119-126.
- MetaMetrics, Inc. (2012). *The Lexile Framework for Reading*. Retrieved from <https://lexile.com/parents-students/measuring-growth-lexile-measures/college-and-career-trajectory/>
- McMaster, K. L., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., ... & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, 22(1), 100-111.
- Miller, J. & Iglesias, A. (2012). *Systematic Analysis of Language Transcripts (SALT), Research Version 2012 [Computer Software]*. Middleton, WI: SALT Software, LLC.

- Mills, S. R., Rice, C. T., Berliner, D. C., & Rosseau, E. W. (1980). The correspondence between teacher questions and student answers in classroom discourse. *The Journal of Experimental Educational*, 48(3), 194-204.
- Moyer, P. S., & Milewicz, E. (2002). Learning to question: Categories of questioning used by preservice teachers during diagnostic mathematics interviews. *Journal of Mathematics Teacher Education*, 5, 293-315.
- Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, 48(2), 439-458.
- Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, 87(2), 261-276.
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67(3), 359-370.
- National Governors Association Center for Best Practices (NGA), Council of Chief State School Officers (2012). *Common Core Standards, 117*. National Governors Association for Best Practices, Council of Chief State School: Washington, D.C.
- National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development, National Institutes of Health.
- O'Flynn, K. (2016). Students' Reading Interests Impact on Reading Comprehension Abilities.
- Parker, D. H. (2004). Reading Laboratory, 2A. Chicago, IL: Science Research Associates (SRA)/McGraw Hill.
- Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.) New York: International Universities Press. (Original work published 1936)
- Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, 105, 243-255.
- Pressley, M. (2001). Comprehension instruction: What makes sense now, what might make sense soon. *Reading Online*, 5(2).

- Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J. L., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *The Elementary School Journal*, 92(5), 513-555.
- RAND Reading Study Group (RRSG) (2002). Toward an R&D program in reading comprehension. Santa Monica, CA: RAND.
http://www.rand.org/pubs/monograph_reports/2005/MR1465.pdf
- Raphael, T. E., & Pearson, P. D. (1985). Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal*, 22(2), 217-235.
- Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal*, 30(3), 523-553.
- Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, 51(2), 237-245.
- Renninger, K. A., & Hidi, S. (2002). Student interest and achievement: Developmental issues raised by a case study. *Development of achievement motivation*, 173-195.
- Reutzel, D. R., & Cooter Jr, R.B. (2012). *Teaching children to read: The teacher makes the difference*. Boston: Pearson.
- Reutzel, D. R. (2014). *Why Theoretical Knowledge of Comprehension is Important* [PowerPoint slides]. Emma Eccles Jones College of Education, Utah State University, Logan, Utah.
- Reutzel, D. R., Smith, J. A., & Fawson, P. C. (2005). An evaluation of two approaches for teaching reading comprehension strategies in the primary years using science information texts. *Early Childhood Research Quarterly*, 20(3), 276-305.
- 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.
- Roth, W. M. (1996). Teacher questioning in an open-inquiry learning environment: Interactions of context, content, and student responses. *Journal of Research in Science Teaching*, 33(7), 709-736.
- Sáenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-assisted learning strategies for English language learners with learning disabilities. *Exceptional Children*, 71(3), 231-247.

- Samson, G. E., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. *Journal of Educational Research, 80*, 290-295.
- Sauer, K. (2012). *The Impact of Student Interest and Instructor Effectiveness on Student Performance*. (Master's thesis). Retrieved from Fisher Digital Publication, Paper 243.
- Schiefele, U. (1996). Topic interest, text representation, and quality of experience. *Contemporary educational psychology, 21*(1), 3-18.
- Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and individual differences, 8*(2), 141-160.
- *Schirmer, B. R., & Woolsey, M. L. (1997). Effect of teacher questions on the reading comprehension of deaf children. *Journal of Deaf Studies and Deaf Education, 2*, 47-56.
- Scholastic Inc. (2007). *Scholastic Reading Inventory technical guide*. New York: Author. <https://www.hmhco.com/products/assessment-solutions/assets/pdfs/sri/SRI-TechGuide.pdf>
- Seymour, K. (2017, January 23). Schema theory and reading comprehension. Retrieved from <https://wehavekids.com/education/Reading-Comprehension-Theory>
- Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 22*(2), 229-255.
- Shanahan, T., Fisher, D., & Frey, N. (2012). The challenge of challenging text. *Educational Leadership*.
- Sinclair, J. McH., & Coulthard, R. M. (1975). *Towards an Analysis of Discourse*. Oxford: Oxford University Press.
- Sindelar, P. T., Bursuck, W. D., & Halle, J. W. (1986). The effects of two variations of teacher questioning on student performance. *Education & Treatment Of Children, 9*(1), 56-66.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation.
- Stahl S. A., Heubach K. (2005). Fluency-oriented reading instruction. *Journal of Literacy Research, 37*, 25-60.
- Stevens, K. (1980). The effect of topic interest on the reading comprehension of higher ability students. *The Journal of Educational Research, 73*(6), 365-368.

- Stronge, J. H. (2010). *Evaluating what good teachers do: Eight research-based standard for assessing teacher excellence*. Larchmont, NY: Eye On Education.
- Subramaniam, P. R. (2009). Motivational Effects or Interest on Student Engagement and Learning in Physical Education: A Review.
- Taboada, A., & Guthrie, J. T. (2006). Contributions of student questioning and prior knowledge to construction of knowledge from reading information text. *Journal of literacy research*, 38(1), 1-35.
- Tienken, C. H., Goldberg, S., & Dirocco, D. (2009). Questioning the questions. *Kappa Delta Pi Record*, 46(1), 39-43.
- Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children*, 71(2), 181-194.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The Stability of Primary Language Disorder Four Years After Kindergarten Diagnosis. *Journal of Speech, Language, and Hearing Research*, 46(6), 1283-1296.
- *Topping, K. J., & Trickey, S. (2007). Impact of philosophical enquiry on school students' interactive behavior. *Thinking Skills and Creativity*, 2, 73-84.
- *Ward-Lonergan, J. M., Liles, B. Z., & Anderson, A. M. (1998). Listening comprehension and recall abilities in adolescents with language learning disabilities and without disabilities for social studies lectures. *Journal of Communication Disorders*, 31, 1-32.
- Wassermann, S. (1991). The art of the question. *Childhood Education*, 67(4), 257-259.
- Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests* (4th ed.). Austin, TX: PRO-ED.
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of educational psychology*, 89(3), 420.
- Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading: Domain specificity and instructional influences. *The Journal of Educational Research*, 97(6), 299-310.
- Wiig E. H., Semel E., Secord W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.

- Wilén, W. W. (1991). *Questioning skills, for teachers. What research says to the teacher.* (ISBN 0-8106-1089-2). Washington, DC: National Education Association of the United States. (EDRS Document Reproduction No. ED332983)
- Wilén, W., & Clegg, A. (1986). Effective questions and questioning: A research review. *Theory and Research in Social Education, 14*, 153-161.
- Wilkinson, I. A., & Son, E. H. (2010). 16 A Dialogic Turn in Research on Learning and Teaching to Comprehend. *Handbook of Reading Research, Volume IV*, 359.
- Wimer, J. W., Ridenour, C. S., Worthy Thomas, K., & Place, A. W. (2001). Higher order teacher questioning of boys and girls in elementary mathematics classrooms. *The Journal of Educational Research, 95*(2), 84-92.
- Winne, P. H. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research, 49*, 13-50.
- *Wolf, M. K., Crosson, A. C., & Resnick, L. B. (2005). Classroom talk for rigorous reading comprehension. *Reading Psychology, 26*(1), 27-53.
- Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery—Revised, English and Spanish Forms: Examiner's manual.* Itasca, IL: Riverside.
- Worthy, J. (2002). The Intermediate Grades: What Makes Intermediate-Grade Students Want to Read? *The Reading Teacher, 55*(6), 568-569.
Retrieved from <http://www.jstor.org/stable/20205098>

APPENDICES

APPENDIX A

Quality Indicators Coding Sheet: Correlational Studies

Correlational Studies QUALITY INDICATORS (22)

Quality Indicators	Present		Notes
	1	0	
Rationale, Participants, & Setting			
1. Plausible rationale (Review of Literature, Purpose Statement and/or research questions).			
2. Student participants are described with sufficient detail and the process for selecting participants is described with replicable precision.			
3. Sufficient information was provided for teachers/researchers implementing treatment (e.g., years of experience, certification).			
4. Critical features of the physical setting are described with sufficient precision to allow replication.			
Rationale, Participants, & Setting Total:	out of 4		
Measurement			
1. Dependent variables are described with operational precision and with a procedure that generates a quantifiable index.			
2. Score reliability coefficients are reported for all measured variables, bases on analysis of data from the study.			
3. If reliability coefficients are inducted from a prior study or test manual for the inferences made in the study, evidence that scores are valid is provided.			
4. The influences of score reliability and validity on study interpretations are considered in reasonable detail.			
Measurement Total:	out of 4		
Practical & Clinical Significance			
1. One or more effect size statistics is reported for each study outcome and the effect statistic used is clearly identified.			
2. Authors interpret study effect sizes by directly comparing study effects with those reported in related prior studies.			
3. Authors explicitly consider study design and effect size statistic limitations as part of effect interpretation.			
Practical & Clinical Significance:	out of 3		
Potential Analysis Errors			
1. Interpretations of weights from the general linear model (GLM) (e.g., regression) includes examinations of structure coefficients.			
2. Interval data are not converted to nominal scale unless justified.			
3. Univariate methods (e.g., mean, standard deviation, bar charts) are not used in the presence of multiple outcome variables			
4. Univariate methods are not used post hoc to multivariate tests.			
5. Rationale is provided that assumptions of statistical methods used are sufficiently well-met for interpreting results.			

Potential Analysis Errors Total:		out of 5	
Confidence Intervals			
1. Confidence intervals are reported for the sample statistics (e.g., mean, correlation coefficients) in the study.			
2. Confidence intervals are reported for study effect sizes.			
Confidence Intervals Total:		out of 2	
Data Analysis			
1. Were data analysis techniques linked to research question(s) and appropriate for the study?			
2. Were data documented on attrition rates?			
3. Were findings statistically significant?			
4. If findings were not statistically significant, was the statistical power adequate?			
Data Analysis Total:		out of 4	
OVERALL TOTAL:			
		out of 22	

Note. Indicators based on criteria proposed by Gersten et al. (2005), and Jitendra et al (2011), and Thompson, Diamond, McWilliam, Snyder, & Snyder (2005).

APPENDIX B

Quality Indicators Coding Sheet: Experimental Studies

Experimental Studies
QUALITY INDICATORS (18)

Quality Indicators	Indicator Rating		Notes
	Present (1)	Not Present (0)	
Introduction and Description of Participants			
1. Plausible rationale (Review of Literature, Purpose Statement and/or research question)			
2. Was sufficient information provided for student participants (e.g., academic performance and/or disability/difficulty, age, race, gender, IQ, SES)? The process for selecting participants is described with replicable precision?			
3. Was equivalence of groups established across conditions (1 for random or quasi-experimental, 0 for no random assignment); Did the authors indicate the equivalence of groups in the study; (one group did not start out at an advantage)? *If no Control Group or Comparison Group, score 0.			
4. Was sufficient information provided for teacher/interventionists provided (e.g., years of experience, certification, age, gender, etc.)?			
Intervention & Comparison Conditions			
5. Independent Variable: Was the intervention clearly described and implemented as intended?			
6. Was there a description of treatment for comparison groups?			
7. Were measures defined? If researcher-developed measures were used, was there a description of how it was developed?			
8. For researcher measures, was there a description of reliability? If standardized measures (e.g., Cronbach's) was reliability reported ($\leq 80\%$)?			

	Indicator Rating		
9. Were outcomes evaluated with blinding (e.g., blinding of scorers, examiners, parents, family members, etc.)?			
10. Was there a description and measurement of fidelity?			
Outcome Measures & Data Analysis			
11. Did the outcome measures align with the intervention AND demonstrate that generalizable skills have been successfully taught?			
12. Were outcomes for capturing the intervention's effect measured at the appropriate time (within 2 weeks of intervention)?			
13. Were data analysis techniques linked to research question(s) and appropriate for the study (rationale for analysis and support for the unit of analysis)?			
14. Did the authors discuss variables that could have distorted the findings (e.g., history, instrumentation, other threats to internal validity)?			
15. Were data documented on attrition rates?			
16. Were findings statistically significant?			
17. If no, if the findings were not statistically significant, was the statistical power adequate? (Were there enough subjects?)			
18. Did the authors not only include inferential statistics but also effect size calculations (i.e., were effect sizes reported)?			
TOTAL			

Note. Indicators based on criteria proposed by Gersten et al. (2005) and Jitendra et al (2011).

APPENDIX C

Letter of Informed Consent



Page 1 of 4
USU IRB Approval: Feb. 21, 2015
Approval Terminates: 02/20/2016
Protocol #6517
IRB Password Protected per IRB Administrator

INFORMED CONSENT

The Effects of Using a Low- to High-level Questioning Strategy on Students' Text-based Reading Comprehension

Introduction/ Purpose Shannon Harris and Dr. Benjamin Lignugaris/Kraft (primary researchers) from the Department of Special Education and Rehabilitation at Utah State University are conducting a research study to find out more about teacher questioning strategies in the classroom and how implementation of these strategies could help students improve their text-based reading comprehension. Your permission is requested to allow your child to participate in this research due to the results of his/her comprehending text, below grade level according to the Scholastic Reading Inventory (SRI) Lexile® score that is assessed in Canyons School District. There will be approximately 9 to 12 total participants in this research.

Procedures If you agree to have your child participate in this research study, the following will occur:

1. Your child will be given the Clinical Evaluation of Language Fundamentals (CELF-5; Fifth Edition) Recalling Sentences subtest. The anticipated assessment time for this subtest is five minutes.
2. Your child will read short narrative stories aloud with a researcher as part of his/her reading instruction in his/her elementary school. Each story will be read three times.
3. On the third reading of the story, the researcher will ask your child to verbally respond to comprehension questions throughout the story.
4. After reading the story, your child will provide verbal responses to three comprehension questions to assess his/her understanding of the story.
5. All of the reading lessons will be video recorded to ensure accuracy of observations.
6. At the end of the study, your child will be asked to complete a questionnaire about his/her experience in the study.
7. It is anticipated that your child will engage in up to 20 reading lessons over the course of 10-12 weeks.

New Findings During the course of this research study, you will be informed of any significant new findings (either good or bad), such as changes in the risks or benefits resulting from participation in the research, or new alternatives to participation that might cause you to change your mind about having your student continue in the study. If new information is obtained that is relevant or useful to you, or if the procedures and/or methods change at any time throughout this study, your permission to allow your student to participate in this study will be obtained again.

Risks There is minimal risk in participating in this research. However, your child may be asked to engage in more reading tasks than his/her peers in other small reading groups. This may single your child out from his/her peers. To minimize this risk, the reading lessons will be scheduled in collaboration with your child's teacher over the course of the study. The lessons from this study are aimed to compliment and support your child's reading progress as he or she continues to engage in the ongoing regular reading curriculum. Further, breaks will be given in between multiple readings of each story and there will be no pressure of time limits for your child to complete any of the reading tasks.



Page 2 of 4
USU IRB Approval: Feb. 21, 2015
Approval Terminates: 02/20/2016
Protocol #6517
IRB Password Protected per IRB Administrator

INFORMED CONSENT

The Effects of Using a Low- to High-level Questioning Strategy on Students' Text-based Reading Comprehension

The primary researchers in this study will be responsible to report any illegal activity or potential abuse discovered during the research process.

Benefits There may or may not be any direct benefit to your child from these procedures. Your child may benefit from the questioning strategies that were designed to help increase students' ability to understand a story and improve their text-based reading comprehension. In addition, the researchers may learn more about effective teacher questioning strategies when working with students who are comprehending text below grade level. This knowledge may contribute to the research for determining the best teaching methods to help students in the area of reading, especially for building students' reading comprehension.

Explanation & offer to answer questions Shannon Harris has explained this research study to you and answered your questions. If you have other questions or research-related problems, you may reach Shannon Harris at (435) 797-8674 or shannonharris8@gmail.com. Dr. Lignugaris/Kraft may be reached at (435) 797-2382 or ben.lig@usu.edu.

Voluntary nature of participation and right to withdraw without consequence Participation in research is entirely voluntary. You may refuse to have your child participate in this research or you may withdraw your child at any time without consequence or loss of benefits. Your child may also withdraw from this study at any time without consequence. Children who engage in noncompliant behaviors, will be removed from the study by the researchers; parents will be notified.

Confidentiality The research team will have access to your child's test scores (e.g., SRI, AIMSweb, SAGE, ITBS) that are available to his/her teachers in Canyons School District. Research records will be kept confidential, consistent with federal and state regulations. Only the primary researchers and Monica Lewis, the Achievement Coach for Canyons School District, will have access to this information for assigning students to reading groups and to measure and monitor their progress throughout the study. All test scores and records will be kept in a locked file cabinet in a locked room to maintain confidentiality and also on password-protected computers used by the research team. To protect the privacy of your child, his/her name will be replaced with a code number on all data collected for this study. The code will be stored in a locked file cabinet at USU. Personal, identifiable information will be kept for one year in order to finish analyzing data and finalize results. Then the code along with any other personal, identifiable information will be destroyed.

Further, each reading lesson will be video recorded using flipcams or iPads. These records will be stored using an encrypted and password-protected application. The password will only be available to members of the research team. Access will be monitored regularly to ensure that no unauthorized individuals have tried to access the video content.



Page 3 of 4
USU IRB Approval: Feb. 21, 2015
Approval Terminates: 02/20/2016
Protocol #6517
IRB Password Protected per IRB Administrator

INFORMED CONSENT

The Effects of Using a Low- to High-level Questioning Strategy on Students' Text-based Reading Comprehension

Flipcams or iPads will be set up at an angle that directly captures the teacher, but only captures the backs of the students. If students' faces are captured directly on camera, their faces will be blurred using video editing software prior to storing videos.

Other than video records, all data will be stored only in the lead researcher's password-protected computer at Utah State University in password-protected files. All standards for confidentiality required by the American Psychological Association (APA) will be followed.

IRB Approval Statement The Institutional Review Board for the protection of human participants at Utah State University has approved this research study. If you have any questions or concerns about your rights or a research-related injury and would like to contact someone other than the research team, you may contact the IRB Administrator at (435) 797-0567 or email irb@usu.edu to obtain information or to offer input.

Copy of Parent Permission You have been given two copies of this Parent Permission document. Please sign both copies and keep one copy for your files.

Investigator Statement "I certify that the research study has been explained to the individual, by me or my research staff, and that the individual understands the nature and purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered."

Signature of Researchers:

Shannon K Harris
Student Researcher
(435) 797-2381
shannonharris8@gmail.com

Benjamin Lignugaris/Kraft
Principal Investigator


Utah State University
 Department of Special Education & Rehabilitation
 2865 Old Main Hill
 Logan UT 84322-2865
 Telephone: (435) 797-3243



Page 4 of 4
 USU IRB Approval: Feb. 21, 2015
 Approval Terminates: 02/20/2016
 Protocol #6517
 IRB Password Protected per IRB Administrator

INFORMED CONSENT
*The Effects of Using a Low- to High-level Questioning Strategy on Students’
 Text-based Reading Comprehension*

Signature of Parent / Guardian By signing below, I agree for my child to participate in this study. I also give my consent to have my child be video recorded under the protocols described in this document.

Parent(s)/Guardian Signature

Date

Child/Youth Assent: I understand that my parent or guardian is aware of this research study and that they have given permission for me to participate. I understand that it is up to me to participate even if they say yes. If I do not want to be in this study, I do not have to and no one will be upset if I don’t want to participate or if I change my mind later and want to stop. I can ask any questions that I have about this study now or later. By signing below, I agree to participate.

Name (please print)

Date

Child/Youth Signature

Date

APPENDIX D

Sample Reading Lesson Materials

Narrative Story: What Jo Did

What Jo Did: *High-level Questions Only* (BASELINE)

BASELINE LESSON

Lexile Score	880L
Word Count	896
Question Packages	6
High-Level	14
Character	4 (28%)
Event	5 (36%)
Idea	5 (36%)
Low-Level	0
Total Questions	14
Average # questions per package	2.3
Average # words between packages	149

Notes:

Joanna loved to play basketball. She especially loved the sound the ball made as it fell through the net. She practiced every day, jumping high enough to touch the backboard. Joanna’s parents had no idea how high a basketball rim should be. They hung it on the side of their roof, which was a whopping sixteen feet high.

Joanna saw rims on TV and figured they looked about the same height as hers—she had no idea they were only ten feet high. **[83 words]**

Question Package 1:

(H) Why do lots of people like to play basketball? (*IDEA(1): What people like/interests*)

(H) Describe why you think it would be good or bad to practice basketball with a rim that is not the standard height. (*EVENT(1): Joanna practiced on a hoop that was too high*)

She also didn’t realize that most people couldn’t jump up and touch the backboard because she hadn’t ever played with anyone else. But her parents marveled at how high she jumped and how she could make baskets. Her father was especially proud because he couldn’t even touch the bottom of the net.

One day Joanna, her hair bundled up under her baseball cap, was dribbling her basketball on the way to the store. A young boy dressed in sneakers, shorts, and a basketball jersey came by.

“Hey, we need one more to play a game. You in?” he asked.

“Sure, why not?” she responded. As Joanna approached the other boys, she remembered she had her hat on. “They probably think I’m a boy,” she thought. “Might as well enjoy the ride.”

The boys picked teams, and since Joanna was smaller than everyone else, she got picked last. **[147 words]**

Question Package 2:

(H) Why do you think Joanna could jump so high? (*EVENT(2): Joanna could jump high and touch the backboard*)

(H) How would being small be an advantage or disadvantage when playing basketball? (*CHARACTER(1): physical traits*)

(H) How do you feel when your parents or someone else is proud of you? (*IDEA(2): being proud of someone*)

It didn't bother her, though, because she had never played with anyone before and was just happy to be there.

"Hey kid, what's your name?" asked one of the boys.

"Uhh...Jo. My name is Jo," Joanna said nervously.

"All right, Jo, you pick up T.J. over there, see. Don't let him score. He can jump pretty high, you know!" [59 words]

Question Package 3:

(H) Describe the benefits for people who have played on the same sports team together. (*EVENT(3): Joanna had never played basketball with anyone before*)

(H) Do you think Joanna is an honest or a dishonest person? Explain why. (*CHARACTER(2): Trait = dishonesty*)

Jo moved around, just trying to get a feel for playing with other people. She had never even passed the ball or received a pass herself. Playing with others took getting used to, but in no time she was passing the ball. The only thing that puzzled her was why the hoop was so low.

Soon, T.J. took a jump shot and Jo came out of nowhere, jumped into the air, and swatted his shot into the next court.

"Wow, did you see that? Did you see how high he jumped?" one boy said, his mouth wide open. "I've never seen anybody jump that high."

"Hey, I got fouled, and besides, it wasn't that high," said T.J., but his face was so red that he couldn't hide his embarrassment. [129 words]

Question Package 4:

(H) Why do people sometimes act differently when they are trying to fit in with a new group? (*IDEA(3): Joanna didn't hid or dumb down her skills in an attempt to fit in*)

(H) Describe what a person could look like or behave like when they are embarrassed. (*EVENT(4): T.J. was embarrassed*)

"Oh, it's just something I picked up. I practice a lot with my dad," Jo added. The game continued, and Jo was passed the ball more often. The boys encouraged her to shoot more, and when she did, they were amazed how the ball arced in the air like a rainbow before falling straight through the hoop, without touching the rim. As the game

progressed, Jo felt hot, but she knew she couldn't take her hat off, or else she'd be found out.

Jo blocked a few more shots and then one of the boys asked her if she could dunk the ball.

"Dunk? What's that?" Jo asked. This was a word she had never heard before.

"A dunk. You know—a slam, a jam, to throw it down. You jump up and put the ball in the rim while holding on to it."

"You guys, can we finish this game? It's getting dark and my mom wants me home soon," said T.J., still upset that Jo was getting all the attention and that his shot was blocked.

"Hold your horses, T.J.," said one of the boys. "I wanna see Jo dunk."

"Well, I'll try," Jo said, curious herself to see if she could dunk. She started at half-court, dribbled the ball, and headed straight for the rim. She remembered how high her basket was and realized that this one was much lower. As she got to the free throw line, she lifted her left leg and went flying into the air, until she was so high she was looking down on the hoop. Then she put the ball in the rim with both hands. She was up there for a while before she felt her hands on the rim, the ball going through, and her feet touching the ground. When she landed, all of the boys' mouths were hanging open, and for a moment they were speechless. As the boys stared at her, Jo looked down at the ground and saw her hat lying there. She froze.

"So, like...you're a girl?" said one of the boys.

"I can't believe it you guys, we've been playing basketball with a girl," T.J. said with disgust.

"Hey, she may be a girl, but I'd play on her team anytime," said one of the boys and he gave Jo a high-five.

After that, they congratulated Jo and introduced themselves. They even came up with a nickname for her: Jumpin' Jo. [407 words]

Question Package 5:

(H) How does confidence help you try new things? (*CHARACTER(3): Joanna was confident/she tried new things*)

(H) What are some ways that people can make new friends? (*EVENT(5): Joanna made new friends*)

(H) Why do you think sports between boys and girls can be such a big deal or issue? (*IDEA(4): gender fairness in sports*)

In the end, T.J. walked up to her and apologized.

"Sorry Jo," he said. "I've just never played against a girl before. Especially a girl as good as you. I've never seen anyone who can jump like that! You should come and play with us again sometime. But next time, leave your hat at home." Jo smiled.

"Thanks guys. It's more fun to play basketball with you instead of by myself."

Question Package 6:

(H) Tell why you would like or would not like to play basketball? (*IDEA(5): choosing activities you like to do*)

(H) What lessons did Joanna learned from this experience? (*CHARACTER(4): Joanna's experience broadened her concepts of playing basketball with others, she was friendly, etc.*)

What Jo Did: *Low- to High-level Questions (TREATMENT)*

INTERVENTION LESSON

Lexile® Score	880L
Word Count	896
Question Packages/ Sequences	6
High-Level	8
Character	3 (37%)
Event	2 (25%)
Idea	3 (37%)
Low-Level	12
Total Questions	20
Average # questions per package	3.3
Average # words between packages	149

Notes:

Joanna loved to play basketball. She especially loved the sound the ball made as it fell through the net. She practiced every day, jumping high enough to touch the backboard. Joanna’s parents had no idea how high a basketball rim should be. They hung it on the side of their roof, which was a whopping sixteen feet high.

Joanna saw rims on TV and figured they looked about the same height as hers—she had no idea they were only ten feet high. **[83 words]**

Question Package 1:

(L) So what is the standard height that a basketball rim should be?

(L) How high was the rim at Joanna’s house?

(H) Describe why you think it would be good or bad to practice basketball with a rim that is not the standard height. (EVENT(1): Joanna practiced on a hoop that was too high)

She also didn’t realize that most people couldn’t jump up and touch the backboard because she hadn’t ever played with anyone else. But her parents marveled at how high she jumped and how she could make baskets. Her father was especially proud because he couldn’t even touch the bottom of the net.

One day Joanna, her hair bundled up under her baseball cap, was dribbling her basketball on the way to the store. A young boy dressed in sneakers, shorts, and a basketball jersey came by.

“Hey, we need one more to play a game. You in?” he asked.

“Sure, why not?” she responded. As Joanna approached the other boys, she remembered she had her hat on. “They probably think I’m a boy,” she thought. “Might as well enjoy the ride.”

The boys picked teams, and since Joanna was smaller than everyone else, she got picked last. **[147 words]**

Question Package 2:**(L) Who got picked last for teams?****(L) What is Joanna’s size compared to the other boys?****(H) How would being small be an advantage or disadvantage when playing basketball? (CHARACTER(1): physical traits)**

It didn’t bother her, though, because she had never played with anyone before and was just happy to be there.

“Hey kid, what’s your name?” asked one of the boys.

“Uhh...Jo. My name is Jo,” Joanna said nervously.

“All right, Jo, you pick up T.J. over there, see. Don’t let him score. He can jump pretty high, you know!” [59 words]

Question Package 3:**(L) What did Joanna say her name was?****(H) What are some reasons that people might lie about something?****(CHARACTER(2): Joanna lied)****(H) Do you think Joanna is an honest or a dishonest person? Explain why.****(CHARACTER(3): Trait = dishonesty)**

Jo moved around, just trying to get a feel for playing with other people. She had never even passed the ball or received a pass herself. Playing with others took getting used to, but in no time she was passing the ball. The only thing that puzzled her was why the hoop was so low.

Soon, T.J. took a jump shot and Jo came out of nowhere, jumped into the air, and swatted his shot into the next court.

“Wow, did you see that? Did you see how high he jumped?” one boy said, his mouth wide open. “I’ve never seen anybody jump that high.”

“Hey, I got fouled, and besides, it wasn’t that high,” said T.J., but his face was so red that he couldn’t hide his embarrassment. [129 words]

Question Package 4:**(L) Which boy shot the basketball?****(L) What did Joanna do to the ball when T.J. shot it?****(L) What things happened that help you know that T.J. was embarrassed?****(H) Describe what a person could look like or behave like when they are embarrassed. (EVENT(2): T.J. was embarrassed)**

“Oh, it’s just something I picked up. I practice a lot with my dad,” Jo added. The game continued, and Jo was passed the ball more often. The boys encouraged her to shoot more, and when she did, they were amazed how the ball arced in the air like a rainbow before falling straight through the hoop, without touching the rim. As the game

progressed, Jo felt hot, but she knew she couldn't take her hat off, or else she'd be found out.

Jo blocked a few more shots and then one of the boys asked her if she could dunk the ball.

"Dunk? What's that?" Jo asked. This was a word she had never heard before.

"A dunk. You know—a slam, a jam, to throw it down. You jump up and put the ball in the rim while holding on to it."

"You guys, can we finish this game? It's getting dark and my mom wants me home soon," said T.J., still upset that Jo was getting all the attention and that his shot was blocked.

"Hold your horses, T.J.," said one of the boys. "I wanna see Jo dunk."

"Well, I'll try," Jo said, curious herself to see if she could dunk. She started at half-court, dribbled the ball, and headed straight for the rim. She remembered how high her basket was and realized that this one was much lower. As she got to the free throw line, she lifted her left leg and went flying into the air, until she was so high she was looking down on the hoop. Then she put the ball in the rim with both hands. She was up there for a while before she felt her hands on the rim, the ball going through, and her feet touching the ground. When she landed, all of the boys' mouths were hanging open, and for a moment they were speechless. As the boys stared at her, Jo looked down at the ground and saw her hat lying there. She froze.

"So, like...you're a girl?" said one of the boys.

"I can't believe it you guys, we've been playing basketball with a girl," T.J. said with disgust.

"Hey, she may be a girl, but I'd play on her team anytime," said one of the boys and he gave Jo a high-five.

After that, they congratulated Jo and introduced themselves. They even came up with a nickname for her: Jumpin' Jo. [407 words]

Question Package 5:

(L) What did the boys discover about Joanna?

(L) How did everyone but T.J. react when they discovered Joanna was a girl?

(H) Why do you think sports between boys and girls can be such a big deal or issue?
(*IDEA(1): gender fairness in sports*)

In the end, T.J. walked up to her and apologized.

"Sorry Jo," he said. "I've just never played against a girl before. Especially a girl as good as you. I've never seen anyone who can jump like that! You should come and play with us again sometime. But next time, leave your hat at home." Jo smiled.

"Thanks guys. It's more fun to play basketball with you instead of by myself."
[71 words]

Question Package 6:

(L) What sport did Joanna love to play?

(L) What were some of the things Joanna loved about basketball?

(H) Tell why you would like or would not like to play basketball? (*IDEA(2): choosing activities you like to do*)

(H) What other good things can people learn from playing sports? (*IDEA(3): life lessons/skills from playing sports*)

What Jo Did

Joanna loved to play basketball. She especially loved the sound the ball made as it fell through the net. She practiced every day, jumping high enough to touch the backboard. Joanna's parents had no idea how high a basketball rim should be, they hung it on the side of their roof, which was a whopping sixteen feet high. Joanna saw rims on TV and figured they looked about the same height as hers—she had no idea they were only ten feet high.

She also didn't realize that most people couldn't jump up and touch the backboard because she hadn't ever played with anyone else. But her parents marveled at how high she jumped and how she could make baskets. Her father was especially proud because he couldn't even touch the bottom of the net.

One day Joanna, her hair bundled up under her baseball cap, was dribbling her basketball on the way to the store. A young boy dressed in sneakers, shorts, and a basketball jersey came by.

"Hey, we need one more to play a game. You in?" he asked.

"Sure, why not?" she responded. As Joanna approached the other boys, she remembered she had her hat on. "They probably think I'm a boy," she thought. "Might as well enjoy the ride."

The boys picked teams, and since Joanna was smaller than everyone else, she got picked last. It didn't bother her, though, because she had never played with anyone before and was just happy to be there.

"Hey kid, what's your name?" asked a one of the boys.

"Uhh...Jo. My name is Jo," Joanna said nervously.

"All right, Jo, you pick up T.J. over there, see. Don't let him score. He can jump pretty high, you know!"

Jo moved around, just trying to get a feel for playing with other people. She had never even passed the ball or received a pass herself. Playing with others took getting used to, but in no time she was passing the ball. The only thing that puzzled her was why the hoop was so low.

Soon, T.J. took a jump shot and Jo came out of nowhere, jumped into the air, and swatted his shot into the next court.

"Wow, did you see that? Did you see how high he jumped?" one boy said, his mouth wide open. "I've never seen anybody jump that high."

"Hey, I got fouled, and besides, it wasn't that high," said T.J., but his face was so red that he couldn't hide his embarrassment.

"Oh, it's just something I picked up. I practice a lot with my dad," Jo added. The game continued, and Jo was passed the ball more often. The boys encouraged her to

shoot more, and when she did, they were amazed how the ball arced in the air like a rainbow before falling straight through the hoop, without touching the rim. As the game progressed, Jo felt hot, but she knew she couldn't take her hat off, or else she'd be found out.

Jo blocked a few more shots and then one of the boys asked her if she could dunk the ball.

"Dunk? What's that?" Jo asked. This was a word she had never heard before.

"A dunk. You know—a slam, a jam, to throw it down. You jump up and put the ball in the rim while holding on to it."

"You guys, can we finish this game? It's getting dark and my mom wants me home soon," said T.J., still upset that Jo was getting all the attention and that his shot was blocked.

"Hold your horses, T.J.," said one of the boys. "I wanna see Jo dunk."

"Well, I'll try," Jo said, curious herself to see if she could dunk. She started at half-court, dribbled the ball, and headed straight for the rim. She remembered how high her basket was and realized that this one was much lower. As she got to the free throw line, she lifted her left leg and went flying into the air, until she was so high she was looking down on the hoop. Then she put the ball in the rim with both hands. She was up there for a while before she felt her hands on the rim, the ball going through, and her feet touching the ground. When she landed, all of the boys' mouths were hanging open, and for a moment they were speechless. As the boys stared at her, Jo looked down at the ground and saw her hat lying there. She froze.

"So, like...you're a girl?" said one of the boys.

"I can't believe it you guys, we've been playing basketball with a girl," T.J. said with disgust.

"Hey, she may be a girl, but I'd play on her team anytime," said one of the boys and he gave Jo a high-five.

After that, they congratulated Jo and introduced themselves. They even came up with a nickname for her: Jumpin' Jo. In the end, T.J. walked up to her and apologized.

"Sorry Jo," he said. "I've just never played against a girl before. Especially a girl as good as you. I've never seen anyone who can jump like that! You should come and play with us again sometime. But next time, leave your hat at home." Jo smiled.

"Thanks guys. It's more fun to play basketball with you instead of by myself."

Assessment Questions: What Jo Did
Student Copy

What Jo Did

C

Tell the story of how you got your name or a nickname. How does that relate to the story we just read?

E

Describe a time when you made excuses. How does that relate to the story we just read?

I

Tell about something you've done or that you still do that takes a lot of practice. How does that relate to the story we just read?

Appendix E

Scoring Rubric for Reading Comprehension Measures

Text-based Reading Comprehension Scoring Rubric: C-units

4 Elements of Scoring Participant Responses:

1. Question Types	2. Response	3. Scoring	4. Link
1. Character/ Character Trait 2. Event 3. Idea	Part 1 (P1): Participant relates the question to a personal experience. Part 2 (P2): Participant provides details from the text/story.	1. Part 1 Accuracy: a. P1 aligns with question. 2. Part 2 Accuracy: a. Individual Cunits are determined. b. Each Cunit is scored for accuracy. (A) = 1 point (I) = 0 points	The participant relates his or her personal experience back to the text/story. 1. Clear = 1 point 2. Weak = 0 points 3. None = 0 points

1. Question Types: Definitions

1. Character/character trait: references something about the character (e.g., the main character is a police officer) or character trait (e.g., dishonesty, kindness, helps out)

2. Event: references an activity that happened in the story (e.g., went on a family picnic)

3. Idea: references an overall idea or the theme of the story (e.g., friendship, helping others, giving someone a second chance)

2. Response: Part 1 and Part 2

Every participant response will be broken down into two parts: Part 1 cues the participant to give a personal experience that aligns with the question and Part 2 cues the participant to link his or her personal experience back to the text/story. Figure 1 shows an example of what constitutes Part 1 and Part 2 (bolded) of the response.

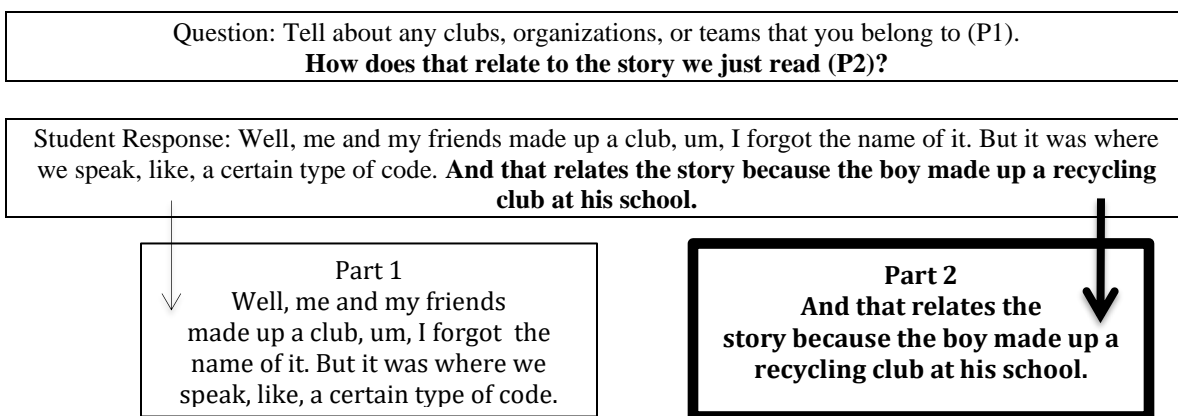


Figure 1. Example of Student Response: Part 1 (P1) and Part 2 (P2)

3. Scoring: Part 1 (P1)

Accurate (A): P1 is accurate when the participant provides a personal experience that aligns with the question and is on topic with what was asked. ***Part 1 is only to be given a score of Accurate (A) or Inaccurate (I).**

Example:

Question: Tell about a time you did something that wasn't easy.

Accurate Response: I gave a speech in an assembly.

Inaccurate Response: I like going out for ice cream after my soccer games.

*Sometimes the participant genuinely does not have a personal experience that aligns with the question. The participant is not penalized for this and the following responses [or similar responses] are to be scored as Accurate (A):

- a. I don't belong to any clubs or organizations.
- b. I don't belong to any clubs because they cost money.
- c. I've never cried because I was happy.
- d. I've never done that before.

*In contrast, the following statements [or similar responses] are to be scored as Inaccurate (I) because it is unclear if the participant did have a personal experience that aligned with the question and/or if the participant lacked the effort to respond to Part 1.

- a. I don't remember.
- b. I don't remember a time.
- c. I can't think of a time.
- d. I don't know.

*Scoring: Link Phrase

Link Phrase: Phrases such as, "And it relates to the story because," or "And it's like the story because," do not count as P1 or P2 and are not to be scored. Also, a student is not penalized if he or she does not include a Link Phrase in a response.

3. Scoring: Part 2 (P2)

Determining Cunits: Each participant will respond to three assessment questions (character/character trait, event, idea). From their entire response, only Part 2 will be scored for Cunits, as this is the part of the response where the participant references the text/story. Scorers will first record the total number of Cunits in each response (i.e., Quantity) and will then score all Cunits as either Accurate (A) or Inaccurate (I).

To break P2 into Cunits:

The formal definition of a Cunit is "an independent clause with its modifiers" (cite SALT software). A Cunit includes one main clause with all subordinate clauses attached to it and cannot be further divided without the disappearance of its essential meaning.

Definition of a clause: A clause, whether it is the main clause or a subordinate clause, is a statement containing both a subject and a predicate. Grammatically, a subject is a noun phrase (*a pronoun also counts for the noun phrase) and a predicate is a verb phrase.

- *Main clauses can stand by themselves and count as one Cunit (independent).
- *Subordinate clauses depend on the main clause to make sense and cannot stand alone or be separated (dependent).

Prepositional phrases (PP) (e.g., *in the house, around the corner, up the tree*) do not count as separate Cunits; rather, they are to be included with the preceding clause.

Example: **The boy made up a recycling club** *at his school* (1).
 [The PP *at his school* cannot stand alone; it is connected to the clause preceding it and the entire response counts as one Cunit.]

Coordinating Conjunctions:

Coordinating Conjunctions (independent) words that signal or cue a new independent clause/Cunit (i.e., connects independent clauses of the sentence). When scoring P2, only coordinating conjunctions can separate Cunits. The coordinating conjunctions used to score participant responses for this study are:

and but or yet for nor so

1. The mom in the story was really stressed (1) for days (2).
2. He had the package ready to mail (1) but left it on the kitchen counter (2).
3. The decorator couldn't decide between the leather (1) or cloth couches (2).
4. She didn't feel hungry (1) yet she wanted something to eat at the football game (2).
5. He was tired of waiting for his friends (1) so he hailed a taxi (2) and went to the concert himself (3).

*There are instances when a coordinating conjunction is part of a phrase, idiom, or definition and the entire phrase is one idea (i.e., it is not possible to break up the phrase without losing meaning). For scoring P2, the coordinating conjunctions of these types of phrases DO NOT signal or cue a new Cunit. These phrases are fixed phrases (i.e., cannot substitute new words). It must be clear that these types of phrases have a contextual meaning on their own, otherwise the phrase will most likely be a binomial phrase (see below).

Examples:

raining cats and dogs
costs an arm and a leg
 left me **high and dry**
all or nothing attitude
 has the **ball and chain**

the research and development division
rise and shine
 looked like **skin and bones**
rock and roll music
 attend the **meet and greet**

hide and seek	stayed at the bed and breakfast
went above and beyond	all the whistles and bells
shall divide and conquer	at the dog and pony show
the hit and run crime	heard you loud and clear
odds and ends around the house	last time once and for all
warm and fuzzy speech	sweet and sour sauce
fun to say trick-or-treat	wash and wear clothing

1. My parents took a weekend vacation (1) stayed at a **bed and breakfast** (2) then went to the beach (3).
2. I play **hide and seek** every recess (1).
3. My favorite music is **rock and roll** music (1) country (2).

*In contrast, binomial phrases DO signal or cue a new Cunit. This type of phrase is a pair or grouping of words often used together as an expression, usually conjoined by the words *and* or *or*. Although these phrases might seem to be a phrase, idiom or definition (see above), it IS possible to break up the phrase to capture specific details from the story. These phrases, although common, are not fixed. For example, even though a peanut butter and jelly sandwich is widely common, one could also make a peanut butter and honey sandwich or a peanut butter and pickle sandwich, which constitutes two details or Cunits within the phrase.

Examples:

delicious bacon (1) and eggs (2)	carried his bait (1) and tackle (2)
was big (1) and tall (2)	served bread (1) and butter (2)
wore a coat (1) and tie (2)	with your eyes (1) and ears (2)
ordered fish (1) and chips (2)	peanut butter (1) and jelly sandwich (2)
bought new socks (1) and shoes (2)	the bride (1) and groom (2)
ladies (1) and gentleman (2)	mom (1) and dad (2)
felt safe (1) and secure (2)	installed the washer (1) and dryer (2)

1. The teacher told the boys (1) and girls (2) to line up for lunch (3) and walk quietly (4).
2. She was bound (1) and determined (2) to succeed.
3. There were many pros (1) and cons (2) about the decision.

*Any paired examples with repetition are to scored as one idea.

Examples:	go, go go	such and such
	again and again	higher and higher

1. The mom in the story was really stressed (1) for **days and days** (2).
2. The balloon lifted **higher and higher** into the air (1) then popped (2) we bought another one (3).

Subordinating Conjunctions:

Subordinating Conjunctions (dependent) establish the relationship between a dependent clause and the rest of the sentence, turning the entire clause into something

that depends on the rest of the sentence for its meaning. Therefore, subordinating conjunctions DO NOT signal or cue a new Cunit. They are used to introduce a dependent clause and to connect it to the independent clause in the sentence. The subordinating conjunctions are:

after	although	as	as if	as long as	
as though	*because	before	even if	even though	if
if only	in order that	now that	once	rather than	since
*so that	than	that	though	till	unless
until	when	whenever	where	whereas	
wherever	while				

***do not confuse the subordinating conjunction “because” as a word that signals or cues a new Cunit (i.e., “because” is not a coordinating conjunction).**

***In addition, do not confuse the subordinating conjunction “so that” to signal or cue a new Cunit. This may be confusing due to the word “so” (coordinating conjunction) being part of the phrase.**

Accurate: **Henry was angry with his mother** because he didn’t get to buy a toy (1).

Inaccurate: **Henry was angry with his mother** (1) because he didn’t get to buy a toy (2).

Examples:

1. When I was learning how to play baseball, my coach went over the rules **again and again** until we stopped making so many mistakes (1).
2. My brother thinks he is all cool now that he has his driver’s license (1).
3. They trusted each other since the girl in the story saved him from falling when they went rock climbing (1).
4. The animals learned how to get along whenever they were in the same pasture even though they hadn’t been trained yet (1).
5. One time when I was three years old I went swimming (1) and almost drowned because I jumped in the deep end, (2) but my mom didn’t see me until the lifeguard jumped in (3) and got me (4).
6. She went first so that the others would follow her example (1).

Understood Pronouns:

When there is an understood pronoun (the subject of the sentence or a name can be substituted in), the clause counts as a new Cunit because the understood pronoun is part of an independent clause.

Accurate Example: He picked the garbage up (1) and threw it away (2).

[**He** picked the garbage up (1) and [**he**] threw it away (2)].

*The understood pronoun is the reason the clause “threw it away” is independent; otherwise, without the understood pronoun, the clause “threw it away” cannot stand alone (i.e., be an independent clause).

Accurate Example: **They** have an assembly (1) and **[they]** give prizes to people (2) **who** clean up the most around the school (3).

*who = understood pronoun

Accurate Example: **They** have an assembly (1) and **[they]** give prizes to people that have the cleanest lockers around the school (2).

*that = subordinating conjunction

Accurate Example: The student body officers have an assembly (1) and **[they]** give prizes to the teachers if they have clean desks (2).

*if = subordinating conjunction

*Sometimes a participant might respond to P2 and then return to P1. Any return to his or her personal experience IS NOT to be scored for Cunits; rather, this most likely occurs because the participant is making the link between personal experience and the story.

Examples:

1. The instrument I like is the violin (P1). That relates to the story we just read because Susan wanted to play the banjo (1) and I want to play the violin (P1).

**There is only 1 Cunit in this response.*

2. One time I helped my uncle build a bookcase out of old barn wood. It was fun and I still have it in my room (P1). It's like the story because the uncle helped Bill with his gear (1) and with fishing (2) and my uncle helped me a lot too (P1).

**There are 2 Cunits in this response.*

One Cunit Examples:

(main clauses are in bold; PP are in italics; subordinate clauses are underlined)

1. **Emily laughed at the playful squirrel** (1). [main clause]
2. **Emily laughed at the playful squirrel** *in the park* (1). [main clause plus PP]
3. **Emily laughed at the playful squirrel** when it rolled over (1). [main clause plus a subordinate clause]
3. **Emily laughed at the playful squirrel** because it rolled over (1). [main clause plus a subordinate clause]
4. **He invented a toaster** that sprayed toast *with melted butter* (1). [main clause plus a subordinate clause with a PP]
5. **He invented a toaster** that sprayed toast while playing music (1). [main clause plus two subordinate clauses]

Multiple Cunit Examples: (separated by Coordinating Conjunctions)

(main clauses are in bold; PP are in italics; subordinate clauses are underlined)

1. **Emily laughed at the playful squirrel** (1) and **took a picture** (2).
2. **Emily laughed at the playful squirrel** *in the park* (1) and fed it a walnut (2).
3. **Emily laughed at the playful squirrel** when it rolled over (1) but ended up startling it (2) and it ran up a tree (3).

4. **He invented a toaster that sprayed toast with melted butter** (1) so he got rich (2) and bought a big house (3) and a new car after he paid all his bills (4) and gave some money to his friends (5).

3. Scoring: Part 2 (P2): Accuracy

Accurate (A): P2 is accurate when the participant provides a response that aligns with what happened in the story or information from the story.

0 points = Participant does not attempt P2

1 point = Each individual Cunit

(A) = Accurate Cunit

(I) = Inaccurate Cunit

*The participant must be accurate in the knowledge of details from the story. For example, in the story *A World of Good*, Uncle Matt joined the Peace Corps. If the participant referred to it as the Marine Corps, that Cunit would be scored as Inaccurate (I). However, if the student mispronounces a word but captures the information from the text accurately, then the Cunit is to be scored as Accurate (A) (e.g., the participant says “suburbian” instead of “suburbia.”).

*If there is inference when the student discusses the text in P2, then the corresponding Cunit is inaccurate.

*If the student speaks in general terms and/or a hypothetical situation (even if the idea could be accurately derived from the story) then the corresponding Cunits are to be scored as Inaccurate (I). For example, in the story *Career Crisis*, the participant created a hypothetical situation with the response, “Maybe on a job someday someone doesn’t understand what to do,” and this is Inaccurate (I) because the participant needed to reference characters, events, or ideas that are in the story.

Examples:

Question: Tell about a time you did something that wasn’t easy. How does that relate to the story we just read?

Accurate: Uncle Matt had to work really hard for food (1) and water (2). *This really did happen in the story.*

Inaccurate: Uncle Matt went on a safari while he was in Africa (1) and it was hot (2) and he got a sunburn (3). *(Only Cunits 1 and 2 are accurate; Cunit 3 is Inaccurate because that event did not happen in the story).*

Inaccurate: Uncle Matt worked really hard for food (1) and water (2), but he probably felt good to help the people (3) and the work didn’t bother him (4). *(Cunits 1 and 2 are Accurate, but Cunits 3 and 4 are Inaccurate due to inference made by the participant)*

Calculating Cunits:

Cunits are calculated into two categories: Quantity and Accuracy. The maximum score for each assessment question is 1 point and therefore the maximum score a participant can achieve overall is 3 points.

To calculate Quantity of Cunits: Sum the total number of Cunits for each individual question type.

To calculate Accuracy of Cunits: For each type of assessment question, divide the total number of Cunits (i.e., denominator) by the number of Accurate Cunits. This is the score out of 1 point for that individual question.

FINAL SCORING:

Quantity: Sum the number of Cunits from all three assessment questions (see score of 9 below).

Accuracy: Sum the Accuracy of Cunits score from all three assessment questions (see score of 2.46 below).

Example:

Question	Question Type	Participant Response	Cunits	
			Quantity	Accuracy
1	Character/Character Trait		3	$2/3 = .66$
2	Event		1	$1/1 = 1$
3	Idea		5	$4/5 = .80$
TOTAL:			9	2.46

4. Scoring: Link

Clear = 1 point

Weak = 0 points

None = 0 points

Clear Link:

1. Common Terms
2. Movement Back and Forth
3. A Separate Sentence

A *Clear* link is scored when the student uses key phrases like, “It relates to the story because ___,” and links P1 and P2 together using **common terms** (“I’m good at gymnastics and that relates to the story because Grandma Betty is good at playing the banjo.”), **movement back and forth between parts** (I’m good at gymnastics (P1) and that relates to the story because Grandma Betty is good at playing the banjo (P2), and I want to do harder tricks and move up a level (P1) and Grandma Betty wanted to get better to perform more, so she practiced harder songs (P2).”), or a **separate sentence**

(“My mom can do a backflip and so can my dad. And it relates to the story because Grandma Betty could play the banjo. So that goes together because they both are something interesting that they can do.”).

Weak Link:

A *Weak* link is scored when the student uses key phrases like, “This relates to the story because ___,” but the scorer has to infer the similarities between the two question parts. Further, a weak link is scored when the student leaves out the words “I,” or “Me,” and tends to focus on the character in the story only.

*A *Weak* link is also scored when a Link Phrase is present (e.g., And it relates to the story because...) and includes accurate information, but the link may not exist or is not clearly articulated (i.e., jumbled ideas).

* A *Weak* link is also scored when the student provides a vague or generic statement in the attempt to link P1 and P2.

Examples:

1. Math is hard. Like at first I couldn’t do long division. And he had to do hard stuff like install the water pump and help the people grow their own food. And yeah, that’s how it’s related. (vague or generic statement)

2. I was embarrassed once when I went to hand in my worksheet to the homework box and I tripped on a backpack and fell and everybody laughed. And T.J. didn’t like that a girl was better than he was. And it was pretty much all about that. (vague or generic statement)

3. Well, I didn’t like playing the piano at first but I kept practicing and now I like it kind of like in the story (vague or generic statement)

No Link or None:

If no attempt is made to link the question parts, the response will be scored as *No Link* or *None*.

Linking Examples:

Clear 1. The instrument I like is the violin (A). That relates to the story we just read because Susan wanted to play the banjo (1) and I want to play the violin (P1).

*In this example, the link is *Clear* due to **common terms** (“Susan wanted to play” and “I want to play”). NOTE: There is only 1 Cunit in this response, as the participant returned to Part 1 (personal experience) when linking.

Clear 2. **He had to do hard stuff, like install the water pump and help the people grow their own food.** And I had to go in after school to Mrs. Brunner’s room to get extra help. **And he helped teach the people how to grow food** like she helped me with the steps of long division. **And the people were always nice to**

him and Mrs. Brunner was always nice to me. **So he did hard stuff** like I did long division.

*In this example, the link is *Clear* due to **movement back and forth**.

Clear 3. I think I'm really good at soccer. And it relates to the story because Emily was really good at rock climbing. So that goes together because they are something that we are both good at.

*In this example, the link is *Clear* due to **a separate sentence**.

Weak 4. One time I was helpful when I helped my neighbor rake his leaves. He's really old and has a big yard. And it relates to the story because the kids in the story had fun when they spent the money they earned from doing jobs.

*In this example, the link is *Weak* because the participant did not link the idea of "being helpful" to what the kids did in the story to "be helpful." The participant only talked about the money they earned from helping.

Weak 5. I felt brave one time when I jumped off a really high diving board. I was really scared but my friend went first and so I felt like I could do it too, you know? And it's like the story because the girl, she was rock climbing and was up really high and had to wear all the equipment and just do it and I got really good at jumping off the high dive.

*In this example, the link is *Weak* because the participant had jumbled ideas and did not make a clear link using either common terms, movement back and forth, or a separate sentence.

Potential Scoring Instances:

1. It is possible that a participant will provide accurate details for P2 from the text/story but those details don't align with the assessment question. In these circumstances, the Cunits are to be scored as Accurate (A) but the link is *Weak*.

Example:

Question: Tell about how you got your name or a nickname. How does that relate to the story we just read?

Response: I got my nickname Lundy Undy because my name is London and my dad always called me that. And it relates to the story because the girl in the story dunked the basketball (1) and was better than T.J. (2) and they thought she was a boy (3).

*All the Cunits in P2 are Accurate details from the story and should be scored as 3/3 = 1; however, the link should be scored as *Weak* because she did not provide details from the story that aligned with the assessment question or her personal experience.

2. It is possible that P1 will be Inaccurate, but that P2 will align with the assessment question. In these circumstances, the Cunits are to be scored as Accurate (A) but the link is *Weak*.

3. Sometimes a participant may provide a vague Link Phrase but because it has a reference point, the Cunit is Accurate (A) (see example below).

Example a: ...but it had this cliff (P1) kind of like what they're explaining (P2).
(vague statement, yet the participant had a clear reference to the cliff the personal experience. This would be scored to have 1 Cunit that is Accurate. However, the link would be **Weak** due to a lack of common terms, movement back and forth, or a separate sentence).

Example b: ...but I've never done that (P1) kind of like what they're explaining (P2).
(vague statement and the participant did not provide a clear reference. This would be scored to have 1 Cunit that is Inaccurate and the link would be **Weak** because the statement, "kind of like what they're explaining" cannot stand alone without a reference point.

4. It is common for participants to use "filler words" like in the following examples (see below). These words and phrases could new Cunit, but would NOT count toward Quantity of Cunits and should not be scored (i.e., disregard Cunits made up of filler words or phrases).

- a. and stuff
- b. and yeah
- c. I guess
- d. and that's all I have to say

5. REPEATS and SELF-CORRECTS (SC): Do not "double count" any self-correct phrases or repeated phrases in the response. If a student does give a SC, then score the final answer or what was stated to correct a prior statement.

Repeat Example: They went to the zoo (1) and, well, ~~they went to the zoo~~ and the lions were sleeping (2).

SC Example: ~~Jo decided to go to the store to buy a new, um,~~ she went to the store for her mom (1).

SC Example: Well, **like a month ago** ~~I did a report on,~~ I did an autobiography on Rosa Parks (1), and **I worked on that for a month (repeat, no Cunit).**

SC Example: And then it relates to the story because the farmer guy gave the dog away, **well found the dog (SC)** (1), and they came and picked it up (2).

Repeat and SC: He **picked up the garbage** because he went to the picnic (1) and there was trash all over (2) and ~~he picked it up~~ (3) and threw it away (3).

More Accurate Scoring Examples: Cunits

1. ...and it relates to the story because there's a team of recycling clubs (1) and, ~~um~~, and they want to work together to, ~~um, together to~~, work something they all can do together (2).
2. They boy (1) and his dad (2) and his uncle went fishing early one morning (3). The dad caught the biggest fish (4), but ~~the boy~~, the boy didn't catch anything all day (5).
3. ...and it relates to the story because when Marcie, ~~or whoever he was~~, **the dragonfly** (SC), when he landed on her nose it scared her (1).

Notes from IOA Consensus Meetings: Groups of Transcriptions

Group 1:

- If Part 1 (P1) is Inaccurate, there can still be a Clear Link (mostly due to common terms).
- If Part 1 (P1) is "I don't remember a time," then the link is Weak; however, if the statement is "I don't remember a time when I _____," (i.e., student adds more detail), then P1 is still Inaccurate but there can be a clear link (mostly due to common terms).
- For synonymous common terms (e.g., built and made), make sure the "idea" is the same; Example of Weak link: *I gave* vs. *Bones was given* (even though common terms appear synonymous, the meaning is different).
- Example: "Natalie and Josh, well, Josh said _____." There is no Cunit between Natalie and Josh because the student self-corrected with "...well, Josh said...".
- Weak Link: Inference: Example: "I like spending quality time...and Natalie and Josh go letterboxing every Saturday." This is a Weak Link because the scorer has to infer/inference, even though common terms could be related.
- PRONOUNS: Across the board, misuse of pronouns does NOT make the Cunit Inaccurate.
- Anytime a student moves from P2 back to P1 at the end of his or her response, do not count as a Cunit whenever there is a personal experience.
- "So that"...make sure to look at the potential Cunit preceding "so that"—is it is complete subject and predicate?
- Clear reference point (for Link) would be an Accurate Cunit when Cut is "like in the story." Example: My mom helped me like in the story.
- Inference: "the thing" stated in place of Trout Attract: This is an Inaccurate Cunit because the scorer has to infer too much.
- Trout catcher "thingamabobber" is an Inaccurate Cunit

- “For something”; Example: “He got accused for something.” Would *something* be Accurate or Inaccurate? These are Accurate (stick with subordinate conjunctions). More examples: For something, for her, for fun, for today.
- Inference: “stayed up the whole night”...these events in the story happened at midnight...is it an Accurate interpretation that the student said, “they stayed up all night?” Continue to mark these as Accurate.

Group 2:

- For Part 1 (P1): Hypothetical situations or philosophy or generic instances are Inaccurate (watch for the word “if”)
- A period does not necessarily indicate the onset of a new Cunit.
- Do not be distracted by punctuation (transcribers’ interpretations vary with punctuation).
- Statements like, “Then that made it,” is a Cunit, but Inaccurate because there is no clear reference point and scorers had to infer too much.
- For P1 and P2 Cunits, any hypothetical situation or big idea as a response is Inaccurate.
- When the scorer implies either an “and” or a “like”; Student: “How brave he was to, like, go in the mountain,” “How brave he was to do that and go in the mountain.” (No understood “and”; that = go into the mountain); Be careful to apply an understood *like* or *and*; step back and think, “What is the overall thought?”
- FOR: this is a subordinating conjunction, so count all “for” in the study—do not count them toward prepositional phrases.
- “...and I don’t know how it relates to the story,” counts as one Cunit and is Inaccurate.
- Be careful of students saying “so”—if a student self-corrects from a coordinating conjunction to a subordinate conjunction, then you count the subordinate conjunction...the last thing the students says in the self-correct is what counts.
- Self-corrects vs. repeats (operational definitions)
- If students say multiple coordinating conjunctions (and subordinating conjunctions) in a series, you only count the last one he or she said before moving on in the Cunit count: Example: because and so (only count “so”).
- For future studies: Do students have to quote characters from the story EXACTLY, or is it Accurate if the student captures the idea of what was said or what happened? (For Letterboxing story, student said, “Sure, I’ll go...” even though the girl did not say that in the story; From Amanda and Horace: “Was that real or was that fake?”). Big Idea: Students do not need to quote verbatim. If they capture the big idea, then the Cunit is Accurate.
- When a student does attempt to quote a character from the story: Does the entire quote count as one Cunit, regardless if there are multiple onsets of new Cunits within that quote? OR should we break down the student’s quote into separate Cunits? For current scoring, the entire quote counts as ONE Cunit even if the student has several sentences or multiple Cunits within the quote. The quote captures “one idea or Cunit” from the story.

Group 3:

- Several transcription errors accounted for IOA disagreements
- When two conjunctions are together, use the last one (“and” “because”)
- When student self-corrects, take the second part of the answer, even in Part 1: “I never had that happen before. Well, that happened with one of my cousins. I think. Sort of.”
- “For”: sometimes “for” acts as a coordinating conjunction and sometimes acts as a preposition to initial a prepositional phrase; for this study, we decided to stick with “for” as a coordinating conjunction always; remains an issue if replicating this study.
- Discussion: Accuracy of Cunits: Does the student have to align his or her interpretation with the question/even referenced by the researcher? Or can the student have another Accurate interpretation that fits: Example: “He was curious about the new comic book.” (instead of being curious about the grave, etc.); for this item, we scored as Accurate.
- Students are not penalized for incorrect pronouns (e.g., he/she).
- If the student has movement from P1 to P2 and back to P1, then if any of the P1 parts are Inaccurate, then the entire P1 is Inaccurate.
- P2 needs to be scored as the student laid out in response; be careful when scoring not to rearrange and recreate what the student was trying to say; this is important when students repeat phrases.

Appendix F
Social Validity Questionnaire
Group 1

GROUP 1: Social Validity Questionnaire: Teacher Script and Recording Sheets

Student: _____

Directions: "I will read some questions out loud to you about the reading lessons you had with Miss Harris. Please be honest in your answers and do your best."

(Lay out tan cards) *"These cards tell about the questions Miss Harris asked when you read stories.*

(Point to a.) *This card describes the lessons when Miss Harris asked questions about your own experiences, opinions, and ideas from the story. I will read the examples while you follow along:*

- *Why is laughter important?*

OR

- *Would you have gone into the cave by yourself? Explain your answer.*

OR

- *How does recycling improve your community?*

(Point to b.) *This card describes the Lessons when Miss Harris asked you about the details in the story PLUS questions about your own experiences, opinions, and ideas from the story AND THEN asked you how those related back to the story. I will read the examples while you follow along:*

- *What is the name of the sister in the story?*

OR

- *How does what you said relate back to this story?*

ITEMS

1. "Point to the card of the reading lesson you liked better."

(Circle student's response).

a.

b.

2. "Point to the card of the reading lesson that helped you remember the stories better." (Circle student's response).

a.

b.

(Remove cards)

(Place story list in front of student)

3. Look at the following list of stories that you read with Miss Harris. Point to the three stories you remember the most details about (Circle the three stories the student pointed to).

Amanda and Horace

Bones for Christmas

Chandler's Secret Weapon

Lessons on the Ledge

Shadow and Carly

Fences and Friendships

***Place Student Questions 4-15 in front of student and the Rating Scale.**

***Record each student's response as they answer each item.**

4. I liked the topics of the stories, or what the stories were about.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

5. I liked being excused from my classroom for the reading lessons.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

6. I enjoyed reading the stories out loud.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

7. I liked reading stories in a small group.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

8. Reading out loud in the group made me nervous.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

9. The stories were hard to read.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

10. The stories were hard to understand.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

11. I carefully followed along while we took turns reading the stories out loud.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

12. It was hard to answer Miss Harris' questions about my own experiences, opinions, and ideas from the story.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

13. It was hard when Miss Harris asked me to relate my own experiences, opinions, and ideas back to what was happening in the story.

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories

14. "This is the last question. You just need to respond with a "yes" or a "no." I feel like I am a better reader after completing the reading lessons with Miss Harris.

_____ Yes

_____ No

15. (use iPad to record)

Please comment on anything else you'd like to say about participating in the reading lessons with Miss Harris.

GROUP 1

Student Materials: Social Validity Questionnaire

1.

a. Questions about your own experiences, opinions, and ideas from the story.

Examples:

- Why is laughter important?
- Would you have gone into the cave by yourself? Explain your answer.
- How does recycling improve your community?

2.

b. Questions about the details in the story PLUS questions about your own experiences, opinions, and ideas from the story AND THEN asked you how those related back to the story.

Examples:

- What is the name of the sister in the story?
- How does what you said relate back to this story?

Question 3: List of Stories

Amanda and Horace
Bones for Christmas
Chandler's Secret Weapon
Lessons on the Ledge
Shadow and Carly
Fences and Friendships

Rating Scale for students to point to: Questions 6-15

No for ALL stories	No for most stories	Yes for most stories	Yes for ALL stories
------------------------------	-------------------------------	--------------------------------	-------------------------------

6. I liked the topics of the stories, or what the stories were about.
7. I liked being excused from my classroom for the reading lessons.
8. I enjoyed reading the stories out loud.
9. I liked reading stories in a small group.
10. Reading out loud in the group made me nervous.
11. The stories were hard to read.
12. The stories were hard to understand.
13. I carefully followed along while we took turns reading the stories out loud.
14. It was hard to answer Miss Harris' questions about my own experiences, opinions, and ideas from the story.
15. It was hard when Miss Harris asked me to relate my own experiences, opinions, and ideas back to what was happening in the story.
16. I feel like I am a better reader after completing the reading lessons with Miss Harris.
17. Please comment on anything else you'd like to say about participating in the reading lessons with Miss Harris.

Appendix G

Example: Comprehension Assessment with Scripted Directions

Scripted Directions: (given orally at the beginning of each Assessment Session)
(Place sheet of questions in front of student) I'm going to ask you three questions (point to all three questions) about the story we just read. The questions are here in front of you and you can ask me to repeat the questions as many times as you need, but I cannot help you answer the questions. Remember, there are two parts for each question, so remember to answer both parts. Do your best.

*Assessment Questions: The House on Maple Street
Student Copy*

The House on Maple Street

C

Describe a time one of your neighbors was friendly to you.
How does that relate to the story we just read?

E

Tell about a time when you or someone you know moved into a new house. How does that relate to the story we just read?

I

Tell about something you tried to invent or make once.
How does that relate to the story we just read?

Appendix H
Student Interest Survey

Student Interest Survey

(Lay out the following cards with these four categories: 1. Liked the Most, 2. It was OK, 3. Liked the Least, 4. I Don't Remember)

Liked the Most

It Was OK

Liked the Least

I Don't Remembe r

You will use these categories to tell me your opinion about the stories you read with Miss Harris. They are (point to each as you read aloud) Liked the Most, It was OK, Liked the Least, and I Don't Remember.

(Present stack of Story Title Cards)

The stories you read with Miss Harris are listed on these cards. I will hand you a card one at a time and read the title to you. You will place the card on of these categories here (point). If it is a story you really liked, place the card here (point). If you thought the story was OK, place the card here (point). If it is a story you didn't like, place the card here (point). If you don't remember the story, place the card here (point).

Record the category given by the student using the following symbols:

Write a plus sign (+) next to the stories liked the most.

Write a check mark (✓) next to the stories that were OK.

Write a minus sign (-) next to the stories liked the least.

Write an (X) next to the stories that the student did not remember.

FOR EACH CARD:

1. Flip a card over, read the title, and hand the card to the student.
 2. Ask: "Do you remember this story?"
 3. If NO, then say, "Place the card in the I Don't Remember pile."
If YES, then say, "Place the card in the pile that tells how you felt about this story."
 4. Record the student's responses below.
- *After you record the responses, the cards must be returned to this same order to be presented to the rest of the students.*

_____Shadow and Carly

_____Trout Fishing

_____Lessons on the Ledge

- _____ Chandler's Secret Weapon
- _____ Bones for Christmas
- _____ Picnic Food
- _____ What Jo Did
- _____ Blood for Chiaka
- _____ Circumstantial Evidence
- _____ Love Grows
- _____ Amanda and Horace
- _____ Cupcake Wars
- _____ Lenny the Flying Inventor
- _____ Mr. Pancake Turkey
- _____ Butterflies are Free
- _____ Fences and Friendships
- _____ The Day I Saw the Ghost
- _____ Kitchen Table
- _____ A Pet for Sugar
- _____ The House on Maple Street

Liked
the
Most

It
was
OK

Liked
the
Least

I
Don't
Remember

A Pet for Sugar	Amanda and Horace	Blood for Chiaka	Bones for Christmas
Butterflies are Free	Chandler's Secret Weapon	Circumstantial Evidence	Cupcake Wars
Fences and Friendships	Kitchen Table	Lenny the Flying Inventor	Lessons on the Ledge
Love Grows	Mr. Pancake Turkey	Picnic Food	Shadow and Carly
The Day I Saw the Ghost	The House on Maple Street	Trout Fishing	What Jo Did

Appendix I
Treatment Fidelity Checklist

Fidelity Checklist

Leg _____ Story _____ Condition _____

Yes/No	Item	Notes
	Instructor stopped to ask Question Packages at the correct locations in the story as outlined in the script.	
	Instructor asked the correct Question Package when stopping to ask questions.	
	Instructor asked each question within Question Packages as outlined in the script (i.e., asked in order; no adlib or additions/omissions for each question)	
	Instructor's repeats/paraphrases of student responses were to confirm what was said and/or clarify the response, not to praise or prompt a response to a new question.	
	<p>Instructor only provided praise focused on classroom management or to maintain instructional pace and motivation.</p> <p>Allowed:</p> <ul style="list-style-type: none"> - Responding to student's answers with phrases like, "Ok," "All right," "Good, let's keep reading," "Excellent," "Thank you." - Providing praise that targeted management throughout the lesson, such as "Thank you for raising your hand," "Good reading today," "Thank you for following along," "You're doing great," "Thanks for being with me." 	
	<p>The instructor only provided error correction for decoding words, repeating the question to students, or directing students to text on low-level questions only.</p> <p>Allowed:</p> <ul style="list-style-type: none"> - Instructor can assist students with reading words (i.e., telling student the word or pronunciation) but cannot provide a model/test/delayed test error correction sequence. - Instructor can repeat questions as needed to students during each Question Package. - If needed, the instructor can encourage students to look back to the text to refine their answers. - Instructor can provide scaffolds (e.g., look back to the text, tell students they are close and to try again) to lead students to the correct answer, especially for low-level questions during the low- to high-level sequences (e.g. "Think back to...") 	

Appendix J
Student Interest Analysis

Table 25

Student Interest Analysis: Recentness Effects Across All Students

Students	First 5 Narrative Stories				Last 5 Narrative Stories			
	<i>Liked the Most</i>	<i>It Was OK</i>	<i>Liked the Least</i>	<i>I Don't Remem- ber</i>	<i>Liked the Most</i>	<i>It Was OK</i>	<i>Liked the Least</i>	<i>I Don't Remem- ber</i>
388L	2	2	1		5			
528L	2	3			3	2		
533L	1	3		1	2	2	1	
000L	1	2	2		2	1	1	1
317L	2	1		2	2	1	1	1
478L	1	4			4	1		
527L	1	2	2		1	2	1	1
283L	1	3	1		3		2	
481L	2	1	2		3	2		
577L		1	1	3	4	1		
595L	1	1	1	2	3	1	1	
Rank total:	14	23	10	8	32	13	7	3
# of students:	10	11	7	4	11	9	6	3
Avg # of stories:	1.4	2.1	1.4	2.0	2.9	1.4	1.2	1.0

Appendix K

Opportunities to Respond: Lowest Performing Students

Table 26

Opportunities to Respond: Lowest-Performing Students

	Scripted Questions	Actual OTR	Ratio: (OTR:Q)
388L (Group 1)			
High-level only (Story 1)	12	10	.83:1
High-level only (Story 2)	11	12	1.1:1
High-level only (Story 3)	13	10	.77:1
<i>M</i>	12	10.7	.90:1
L-H with linking prompt (Story 1)	20	13	.65:1
L-H with linking prompt (Story 2)	25	9	.36:1
<i>M</i>	22.5	11	.51:1
000L (Group 2)			
High-level only (Story 1)	13	11	.85:1
High-level only (Story 2)	13	9	.69:1
<i>M</i>	13	10	.78:1
L-H with linking prompt (Story 1)	21	14	.67:1
L-H with linking prompt (Story 2)	23	12	.52:1
<i>M</i>	22	13	.60:1
317L (Group 2)			
High-level only (Story 1)	13	13	1.0:1
High-level only (Story 2)	13	14	1.1:1
<i>M</i>	13	13.5	1.1:1
L-H with linking prompt (Story 1)	21	13	.62:1
L-H with linking prompt (Story 2)	23	16	.70:1
<i>M</i>	22	14.5	.66:1
478L (Group 2)			
High-level only (Story 1)	13	8	.62:1
High-level only (Story 2)	13	5	.38:1
<i>M</i>	13	6.5	.50:1
L-H with linking prompt (Story 1)	21	7	.33:1
L-H with linking prompt (Story 2)	23	10	.43:1
<i>M</i>	22	8.5	.38:1
283L (Group 3)			
High-level only (Story 1)	11	5	.46:1
High-level only (Story 2)	11	7	.63:1
<i>M</i>	11	6	.55:1
L-H with linking prompt (Story 1)	24	7	.29:1
L-H with linking prompt (Story 2)	25	14	.56:1
<i>M</i>	24.5	10.5	.43:1

Note. Sample stories selected at random. L-H = Low- to high-level questions; *M* = Mean; OTR = Opportunities to Respond.

CURRICULUM VITAE

Shannon Harris Brown

Curriculum Vitae
shannonharris8@gmail.com

EDUCATION

Ph.D. Disability Disciplines,

Utah State University, Logan, Utah

Specialization: Special Education

Emphasis: Multidisciplinary Language and Literacy

Dissertation: Teacher questions in the classroom: The effects of using a low- to high-level questioning sequence on the text-based reading comprehension outcomes of low-performing students

Advisor: Dr. Benjamin Lignugaris/Kraft

M.S. Special Education

Utah State University, Logan, Utah

Thesis: The effects of teaching prefix meaning and a strategy to derive word meaning on a prefix vocabulary test and sentence comprehension test for middle school students with learning disabilities

Advisor: Dr. Benjamin Lignugaris/Kraft

Graduation Date: December 2010

B.S. Dual Special Education & Elementary Education

Utah State University, Logan, Utah

Emphasis: Mild/Moderate Special Education

Honors: Cum Laude Graduate

Dean's List: Spring 2001, Spring, 2002, Summer 2002

Recipient: College of Education; Tuition Waiver Scholarship, 2002-2003

Recipient: ISU/USU Transfer Scholarship, 2000-2001

Graduation Date: December 2002

Associate Arts & Sciences in General Studies

Ricks College (BYU-Idaho), Rexburg, Idaho

Emphasis: Health and Education

Graduation Date: April 2000

RESEARCH AND EDUCATIONAL INTERESTS

- Evidence-based language & literacy practices for children and adolescents
- Evidence-based instruction and behavior strategies for students with mild/moderate disabilities
- Implementation of evidence-based practices and their effects on student outcomes
- Special education teacher and leadership preparation; teacher pedagogy

- Didactic instruction/teacher training
- Field experiences for preparing special education teachers to apply pedagogical skills and evidence-based instructional methods
- Individualizing instruction and delivery for English-language Learners (ELL)
- Classroom management
- Applied Behavior Analysis
- Evaluation of academic outcomes from instructional interventions
- Direct Instruction
- Data analysis; Data-based decision making in the classroom
- Preservice teacher supervision, follow up, and evaluation
- Scientific research methods; single-subject research design

PROFESSIONAL WORK HISTORY

- 2015-2016 Temporary Assistant Professor, Department of Special Education and Rehabilitation, Instructor for two undergraduate courses on-campus and distance education
Utah State University, Logan, Utah
*Awarded Teacher of the Year
- 9/2014 Teaching Assistant, Department of Special Education and Rehabilitation
Utah State University, Logan, Utah
- 2010-2013 Supervisor for Preservice Special Education Teachers' Reading Practicum
Utah State University, Logan, Utah
- 2007-2009 Special Education Teacher & Case Manager
Star Valley Middle School, Afton, Wyoming; Grades 7-8
- 2007-2009 District Direct Instruction Mentor (part-time)
Lincoln County School District No. 2
Afton, Wyoming
- 2005-2007 Special Education and Elementary Education Teacher
Cache Valley Learning Center (Bear River Charter School) Logan, Utah;
Grades 6-8
- 7/2006 Summer Program Teacher
Cache Valley Learning Center (Bear River Charter School) Logan, Utah;
Children ages 5-12
- 2006-2006 Supervisor for Preservice Special Education Teachers' Reading Practicum
Off-campus through Utah State University
Davis School District, Farmington, Utah

- 6/2005-8/2005 Education Program Specialist: Research to Practice Division
Office of Special Education Programs (OSEP)
United States Department of Education
Washington, D.C.
- 2004-2005 Special Education Teacher & Case Manager
E.G. King Elementary School
Layton, Utah; Grades K-6
- 2004-2005 Supervisor for Preservice Special Education Teachers' Reading Practicum
Utah State University, Logan, Utah
- 2003-2004 Elementary Education Teacher
Cache Valley Learning Center (Bear River Charter School) Logan, Utah;
Grades 6-8
- 8/2003-12/2003 Supervisor for Preservice Special Education Teachers' Reading
Practicum
Utah State University, Logan, Utah
- 12/2002-6/2003 Elementary Education Teacher
Cache Valley Learning Center (Bear River Charter School) Logan,
Utah; Grades 2-3, 6
- 8/2002-12/2002 Student Supervisor for USU Undergraduate Special Education
Students' Reading Practicum
Utah State University, Logan, Utah

PUBLICATIONS

- Lignugaris/Kraft, B. & **Harris, S.** (2014). Teacher preparation: Principles of effective pedagogy. In P. Sindelar, E. D. McRay, M. T. Brownell & B. Lignugaris/Kraft (Eds.). *Handbook of research on special education teacher preparation*. New York: Routledge, Taylor, & Francis.
- Harris, S. H.**, Lignugaris, B., & Forbush, D. E. (2016). The effects of morphemic vocabulary instruction on prefix vocabulary and sentence comprehension for middle school students with learning disabilities. *Education and Treatment of Children*, 39(3), 301-337.
- Harris, S. K.**, & Duran, L. (in preparation). Reading comprehension for English language learners: A literature review.

Kunnavatana, S. S., Bloom, S. E., Samaha, A. L., Lignugaris/Kraft, B., Dayton, E., & **Harris, S. K.** (2013). Using a modified pyramidal training model to teach special education teachers to conduct trial-based functional analyses. *Teacher Education and Special Education, 36*, 267-285.

Walbeck, D., Garff T., Menlove, S., **Harris, S.**, & Menlove, R. (2003, March). *What preservice and first year teachers need to know to survive and thrive*. 2003 American Council on Rural Special Education (ACRES) Annual Conference, Salt Lake City, Utah.

PRESENTATIONS

Dawson, M.R., **Harris, S.**, & Lignugaris/Kraft, B. (2014, November). *Using virtual environments to develop teaching skills: Training and generalized effects*. Oral Presentation, Teacher Education Division (TED) of The Council for Exceptional Children (CEC), Indianapolis, IN.

Harris, S. (2014, June). *The effects of didactic instruction on the rate of preservice teachers' low- and high-level questions with students with disabilities*. Poster Presentation, Utah Multi-Tiered System of Supports (UMTSS) Conference, Layton, Utah.

Harris, S. (2014, April). *The effects of didactic instruction on the rate of preservice teachers' low- and high-level questions with students with disabilities*. Oral Presentation, Graduate Symposium, Utah State University.

*Awarded 1st Runner Up

Harris, S. (2014, April). *The effects of didactic instruction on the rate of preservice teachers' low- and high-level questions with students with disabilities*. Poster Presentation, Graduation Symposium, Utah State University.

*Selected to participate in poster session for USU's Inclusive Excellence Research Event

Kunnavatana, S. S., Bloom, S. E., Samaha, A. L., **Harris, S. K.**, & Dayton, E. (2012, November). *Addressing problem behavior in classrooms: Training educators to conduct trial-based functional analyses*. Interactive Paper Session presented at the 35th Annual Teacher Education Division (TED) Conference, Grand Rapids, MI.

Harris, S. (2011, July). *The Effects of Morphemic Vocabulary Instruction on a Prefix Vocabulary Test and a Sentence Comprehension Test*. Poster Presentation, United States Office of Special Education Programs (OSEP) Project Directors' Conference, Washington, D.C.

Harris, S. (2011, June). *Using morphemic vocabulary instruction to improve word knowledge and sentence comprehension for middle school students with learning disabilities*. Oral Presentation, Effective Practices Conference, Utah State University.

Harris, S. (2011, April). *Morphemic vocabulary instruction for middle school students with learning disabilities*. Poster Presentation, Graduate Symposium, Utah State University.

*Awarded First Place

Harris, S. (2002, May). *Self-managed behavioral system stimulates student success*. Poster Presentation, Applied Behavioral Analysis Showcase, Utah State University.

COURSES TAUGHT

- Spr 2016 Main Instructor as Temporary Assistant Professor (Distance Program)
 SPED 5310: Teaching Reading and Language to Students with Disabilities
 SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
 Utah State University, Logan, Utah
- Spr 2016 Guest Lecturer
 Utah State University's ATP (Alternative Teacher Preparation) program
 Instructional Scaffolding and Error Correction Strategies
 Granite Education Center, Salt Lake City, Utah
 Professor: Dr. Melanie Dawson
- Fall 2015 Main Instructor as Temporary Assistant Professor (On-campus)
 SPED 5310: Teaching Reading and Language to Students with Disabilities
 SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
 Utah State University, Logan, Utah
- Fall 2014 Teaching Assistant for Data-based Decision Making Training
 SPED 5310: Teaching Reading and Language to Students with Disabilities
 SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
 Utah State University, Logan, Utah
 Professor: Dr. Timothy Slocum
- Spr 2013 Support Coach for Distance Students' Data-based Decision Making
 Records (DBDR)
 SPED 5310: Teaching Reading and Language to Students with Disabilities
 SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
 Utah State University, Logan, Utah
 Professor: Dr. Nancy Glomb

- Fall 2012 Main Instructor as a doctoral student
SPED 5310: Teaching Reading and Language to Students with Disabilities
SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
Utah State University, Logan, Utah
Professor: Dr. Timothy Slocum
- Spr 2012 Guest Lecturer
SPED 5310: Teaching Reading and Language to Students with Disabilities
SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
Utah State University, Logan, Utah
Professor: Dr. David Forbush
- Fall 2011 Teaching Assistant
SPED 5310: Teaching Reading and Language to Students with Disabilities
SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
Utah State University, Logan, Utah
Professor: Dr. Timothy Slocum
- Fall 2010 Teaching Assistant
SPED 5310: Teaching Reading and Language to Students with Disabilities
SPED 5510: Curriculum for Students with Severe Disabilities (Part 1)
Utah State University, Logan, Utah
Professor: Dr. Timothy Slocum
- March 2008 District Workshop for Effective Instructional Strategies: Part 2
Training for Special Education Paraeducators
Lincoln County School District No. 2
Afton, Wyoming
- Oct 2007 District Workshop for Effective Instructional Strategies: Part 1
Training for Special Education Paraeducators
Lincoln County School District No. 2
Afton, Wyoming
- Feb 2006 Trainer for *Step Up To Writing* Curriculum
Cache Valley Learning Center (Bear River Charter School)
Logan, Utah

RESEARCH EXPERIENCE

- Current: **Dissertation:** *Teacher Questions in the Classroom: The Effects of Using a Low- to High-level Questioning Sequence on the Text-based Reading Comprehension Outcomes of Low-Performing Students*
 Evaluated the effectiveness of a low- to high-level questioning sequence *without or with linking prompts* on the reading comprehension outcomes of fifth-grade students who evidenced poor reading comprehension. A secondary analysis was used to determine whether the low- to high-level questioning sequence was effective regardless of students' interest in the narrative story content.
- 2014-2015 **Pilot Study 1 for Dissertation**
 Evaluated whether preservice teachers could be trained to implement questioning strategies in a reading lesson. The setting was TLE TeachLivE™ Lab, a mixed-reality teaching environment supporting teacher practice in pedagogy and content.
- Pilot Study 2 for Dissertation**
 All methods and procedures from Pilot Study 1 were replicated; additions included revised didactic instruction and a researcher-developed measure for reading comprehension. The setting was TLE TeachLivE™ Lab.
- 2011 **Study 2 following Thesis:**
 Procedures from thesis were replicated and an instructional component to teach students how to apply the meaning of the prefixed word to sentences was added.
- 2009 **Thesis:** *The effects of teaching prefix meaning and a strategy to derive word meaning on a prefix vocabulary test and sentence comprehension test for middle school students with learning disabilities*
 Examined morphemic prefix instruction as a means to enhance word knowledge and reading comprehension for students with learning disabilities.

GRANT EXPERIENCE

Feb 2013 As a doctoral student, I collaborated with grant committee members in the Department of Special Education and Rehabilitation to propose a new strand within the existing Utah State University Disability Disciplines Doctoral Program (through the *U.S. Department of Education, Office of Special Education Programs*). The new strand focused on Evidence-based practice within multi-tier systems of support (EBP-MTSS).

Main responsibilities included gathering syllabi of courses taught for the Disability Disciplines Doctoral Program, compiling all pieces of the grant application (e.g., appendices, letters, proposal), editing and reformatting, and attending all work meetings with grant committee members.

CERTIFICATIONS, PROFESSIONAL MEMBERSHIPS, AND/OR LEADERSHIP

- State of Utah Teacher Licensure in Special Education; Emphasis: Mild/Moderate
- State of Utah Teacher Licensure in Elementary Education
- State of Wyoming Teacher Licensure in Special Education
- Council for Exception Children (CEC)
- Association for Direct Instruction (ADI)
- Attended the Individuals with Disabilities Education Act (IDEA) Reauthorization Public Forum in Salt Lake City, Utah, on October 10, 2001, and again at Gallaudet University in Washington, D.C., July, 2005.
- Volunteer, Common Ground: Outdoor Recreation for Individuals with Disabilities, Logan, Utah
- Volunteer, Special Olympics, Utah State University
- Student Leader: Selected by USU Vice President for commitment to enhancing the quality of student life at Utah State University, 2001
- Certified in First Aid and CPR: American Red Cross