The Effects of Conditioned Reinforcement for Reading on Reading Comprehension for 5th Graders

Colleen Cumiskey Moore

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017
ABSTRACT

The Effects of Conditioned Reinforcement for Reading on Reading Comprehension for 5th Graders

Colleen Cumiskey Moore

In three experiments, I tested the effects of the conditioned reinforcement for reading (R+Reading) on reading comprehension with 5th graders. In Experiment 1, I conducted a series of statistical analyses with data from 18 participants for one year. I administered 4 pre/post measurements for reading repertoires which included: 1) state-wide assessments, 2) district-wide assessments, 3) 20 min observational probes, and 4) preference probes. I utilized the standardized testing measurements to establish grade-level reading repertoires, while the additional two probes measured the reinforcement value of reading. Observational data were recorded in 10s whole-intervals; participants who were observed to read for 96 of the 120 intervals (80%) were considered to have R+Reading. The results demonstrated that R+Reading is significantly correlated with reading assessment outcomes. In Experiment 2, I implemented a two-year cross-sectional design with 33 participants, where I expanded the previous research to include probe trials for conditioned seeing (CS) and derivational responding (DR). Results of Experiment 2 indicated that increases in standardized testing scores were significantly correlated with R+Reading, and that CS and DR were prerequisite repertoires for the acquisition of R+Reading. In Experiment 3, I tested the effects of the peer-yoked contingency procedure on the reinforcement value of reading and assessed if increases in the reinforcement value of reading functioned to increase reading comprehension. Results indicated that increases in the reinforcement value of reading also was related to increases in reading comprehension.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>x</td>
</tr>
</tbody>
</table>

## CHAPTER I – INTRODUCTION AND REVIEW OF LITERATURE

- **Introduction** ............................................ 1
- **Review of Literature** .................................. 3
  - Overview .......................................................... 3
  - Conditioned Reflex ........................................... 3
  - Respondent Behavior ........................................ 4
  - Operant Behavior ............................................. 4
  - Overt and Covert Responding ............................... 5
- **Emergent Behavior** ...................................... 6
  - Stimulus Equivalence ...................................... 7
  - Relational Frame Theory ................................... 8
  - Naming Theory ................................................ 10
  - Verbal Behavior Development Theory .................... 10
  - Naming Joins Textual Responding .......................... 11
  - Joint Control ................................................ 13
  - Conditioned Seeing ......................................... 14
  - Reading ........................................................ 16
Measurements of Reading ................................................................. 17
Eye Tracking .................................................................................. 18
Comprehension ............................................................................. 19
Reporting ....................................................................................... 21
Research Models for Intervention Design ...................................... 23
Rationale ....................................................................................... 25
Research Questions ........................................................................ 27
CHAPTER II – EXPERIMENT 1 .......................................................... 28
Method .......................................................................................... 28
Participants .................................................................................. 28
Procedure and Measures ............................................................... 29
Statewide Standardized Testing: NJASK ....................................... 31
District Reading Assessment: DRA ............................................... 32
Observations of Reading Behaviors .............................................. 32
Participant Reading Preference ..................................................... 33
Inter-observer and Inter-scorer Agreement .................................... 34
Data Analysis ................................................................................ 34
Results .......................................................................................... 37
Spearman Analysis for Correlated Test Scores ............................. 38
One-way ANOVA Analysis for Categorical Value of R+ Reading ...... 38
Discussion ..................................................................................... 42
CHAPTER III – EXPERIMENT 2 .......................................................... 44
Method .......................................................................................... 44
Participants........................................................................................................44
Procedure ........................................................................................................44
Measures ..........................................................................................................45
Statewide Standardized Testing Scores: PARCC ...................................45
District-wide Reading Assessment Scores: DRA .........................46
Observations of Reading Behaviors.........................................................46
Conditioned Seeing & Derived Relation Probes ..................46
Inter-observer and Inter-scorer Agreement.........................................51
Data Analysis .................................................................................................52
Results...........................................................................................................52
Preliminary Findings..................................................................................52
Independent t-Test for Correlated Test Scores .............................53
ANOVA Analysis of Acquisition with 8 Categorical Variables .......55
Regression Models.......................................................................................57
Discussion ......................................................................................................60
CHAPTER IV – EXPERIMENT 3.....................................................................64
Method ..........................................................................................................64
Participants..................................................................................................64
Setting and Materials ...............................................................................65
Dependent Variables..................................................................................68
Observations of Reading Behaviors.................................................68
Woodcock Johnson III Diagnostic Reading Battery ..................69
Grays Silent Reading Tests.................................................................71
Appendix A. 20-min Reading Observation Data Sheet........................................119
Appendix B. Positive and Negative Exemplars of Conditioned Seeing.........120
Appendix C. Relational Analogies Probe ..................................................121
Appendix D. Positive Exemplars of Intervention Responses .....................122
Appendix E. Negative Exemplar with Recycle for Intervention Responses ....123
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Experimental Variable Coding and Definitions of Code Used in SPSS</td>
<td></td>
</tr>
<tr>
<td>2. Correlation Coefficients Between Reading Assessment Measures in 2013</td>
<td>36</td>
</tr>
<tr>
<td>3. Correlation Coefficients Between Reading Assessment Measures in 2014</td>
<td>38</td>
</tr>
<tr>
<td>4. Textual Stimuli and Novel Sentences Used in Experiment 2 and Experiment 3</td>
<td>38</td>
</tr>
<tr>
<td>5. Independent Group t-test Between Acquisition of Conditioned Reinforcement for Reading and Variable Data</td>
<td>48</td>
</tr>
<tr>
<td>6. Independent Group t-test Between Acquisition of Conditioned Seeing and Variable Data</td>
<td>54</td>
</tr>
<tr>
<td>7. Independent Group t-test Between Acquisition of Derivational Responding and Variable Data</td>
<td>55</td>
</tr>
<tr>
<td>8. Summary of Hierarchical Regression Analysis for Predicting Lexile Scores</td>
<td>59</td>
</tr>
<tr>
<td>9. Summary of Hierarchical Regression Analysis for Predicting ELA PARCC Scores</td>
<td>60</td>
</tr>
<tr>
<td>10. Participant Description for Experiment 3</td>
<td>65</td>
</tr>
<tr>
<td>11. Relational Analogy Stimuli and Types of Derived Relations Used in Experiment 3</td>
<td>73</td>
</tr>
<tr>
<td>12. Number of Correct Responses to Each Reading Assessment Pre and Post Intervention</td>
<td>92</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Naming Joins Textually Responding Diagram</td>
<td>15</td>
</tr>
<tr>
<td>2. Proposed Model of Skinnerian “Sentences”</td>
<td>16</td>
</tr>
<tr>
<td>3. Experimental Assessment Timeline</td>
<td>30</td>
</tr>
<tr>
<td>4. One-way ANOVA Analysis with Categories for R+ Reading and Mean NJASK ELA Score</td>
<td>40</td>
</tr>
<tr>
<td>5. One-way ANOVA Analysis with Categories for R+ Reading and Mean NJASK Reading Score</td>
<td>40</td>
</tr>
<tr>
<td>6. One-way ANOVA Analysis with Categories for R+ Reading and Mean June Lexile</td>
<td>41</td>
</tr>
<tr>
<td>7. One-way ANOVA Analysis with Categories for R+ Reading and Mean Selection</td>
<td>41</td>
</tr>
<tr>
<td>8. Procedure for Collecting Probe Data for Conditioned Seeing and Derivational Responding</td>
<td>50</td>
</tr>
<tr>
<td>9. One-way ANOVA Analysis Between the Acquisition of Reading Repertoires and Mean Lexile Scores</td>
<td>56</td>
</tr>
<tr>
<td>10. One-way ANOVA Analysis Between the Acquisition of Reading Repertoires and Mean ELA PARCC Scores</td>
<td>57</td>
</tr>
<tr>
<td>11. Intervention Materials</td>
<td>68</td>
</tr>
<tr>
<td>12. Peer-yoked Contingency Procedure</td>
<td>75</td>
</tr>
<tr>
<td>13. Experimental Sequence for Experiment 3</td>
<td>78</td>
</tr>
<tr>
<td>14. Observations of Reading Behaviors Pre and Post Intervention</td>
<td>81</td>
</tr>
<tr>
<td>15. Pie Graph Representation of Duration Pre and Post Intervention</td>
<td>82</td>
</tr>
</tbody>
</table>
16. The Grade Level Equivalencies to the Woodcock Johnson III Passage Comprehension Subtest ......................................................................................................................... 84
17. Increases in Grade Level Equivalencies to the Woodcock Johnson III Passage Reading Subtest Pre and Post Intervention ......................................................................................................................... 85
18. The Grade Level Equivalencies to Woodcock Johnson III Vocabulary Subtest .......... 84
19. Increases in Grade Level Equivalencies to the Woodcock Johnson III Reading Vocabulary Subtest Pre and Post Intervention ......................................................................................................................... 88
20. The Grade Level Equivalencies to the Grays Silent Reading Tests ........................................ 90
21. Increases in Grade Level Equivalencies to Grays Silent Reading Tests Pre and Post Intervention ......................................................................................................................... 91
22. A Comparison of Grade Level Equivalencies across Overt and Covert Reading Comprehension ............................................................................................................................................................................. 94
23. A Comparison of Overt and Covert Reading Repertoires for Derivational Relations ... 95
24. Number of Correct Yoked-Responses to Intervention Sessions for Participant A and Participant B ............................................................................................................................................................................ 97
25. Number of Pages Read During Intervention Sessions for Participant A and B ............ 97
26. Number of Correct Yoked-Responses to Intervention Sessions for Participant C and Participant D ............................................................................................................................................................................ 98
27. Number of Pages Read During Intervention Sessions for Participant C and D .......... 98
ACKNOWLEDGMENTS

I would like to begin by thanking my family, who have encouraged my education since I was a child. Your support has shaped me into the person I am today. Dad, thank you for the long hours, the late discussions, and being my lifelong foundation that I can always count on. Mick, your ability to make me laugh and to focus on the brighter qualities of life has been essential to conquering any day, good or bad. Dan, I would have never begun, or succeeded, in this path without you. I can only hope to be to you as you are to me.

To my second family, those who I have worked with, you have motivated me just by sitting around a table for weeks on end, discussing both the mundane and the important. Without you, my thoughts would never have emerged into anything more than passing fancies. Sarah, Ginger, Caroline, Alyssa, Brittany, Angela, Brittany, and Rachel, you all helped me establish the core of my research interests by constantly stretching my thoughts into new possibilities. This research would have never been completed without all of you. Reggie Spilotras, every single time you stepped into the classroom you asked about the students and research first, but never failed to ask about me second. Your thoughtfulness was crucial to my first year in the doctoral program and I have you to thank for a successful year.

Finally, to those who have educated me: every moment you spent speaking with me inspired me to learn more. This applies to my Newark Academy teachers, such as Neil Stourton, my Rutgers University professors, such as Stacy Klein, Darryl Ellsion, and Karen Stubaus, and all the faculty at Teachers College. Thank you, Dr. Jahromi, for expanding my love of research into new possibilities; I had never opened a statistics book before your class and your guidance was instrumental looking at data differently. Dr. Delgado, your advice over the years
has greatly influenced my path towards becoming a respected professional in the field; you model excellence and I hope I can emulate it. Dr. Dudek, your constant support and encouragement helped foster the scientist in me, even when I didn’t know it was there. Dr. Greer, there are not enough words to express my appreciation. You have changed my life and I can never thank you enough.
DEDICATION

To those, who throughout my life, have helped me explore my curiosities.
Reading is a language developmental milestone that requires continual expansion of verbal repertoires in order to acquire more complex human verbal behavior (Greer & Longano, 2010). Lyon and Chhabra (2004) stated that students who struggle with reading will have difficulty mastering academic content, succeeding in school, and achieving career potentials. These difficulties are being reported nationwide. According to the International Association for the Evaluation of Educational Achievement (IEA), the United States has not made significant gains in reading since 2001; thus putting the U.S. farther and farther behind in its peers (Thompson et al., 2012). Similar reports are also found from organizations such as The National Assessment of Educational Progress, where results indicated that reading scores at each of the five selected percentiles in 2015 were not statistically different than those found in 2013 (The National Assessment of Educational Progress, 2015).

The Progress in International Reading Literacy Study (PIRLS), conducted by the Lynch School of Education, sought to examine possible factors that affect reading score outcomes. Researchers found that factors included: home resources for learning, parents “liking” reading, parental educational expectations, preprimary education, school location, and economic background. While none of these influences are unreported (Hart & Risely, 2003; Hersch, 2011; Snow et al., 1998), PIRLS did provide additional data to support factors such as “liking” reading can influence the school environment. After conducting statistical analyses with a sample size of over 19,000,000 students in 39 countries, researchers found that three variables, “Liking to
Read,” “Motivated to Read,” and “Confident in Reading,” were all significantly correlated with higher reading test scores (PIRLS, 2011).

A year earlier, Petscher (2010) conducted a meta-analysis of the relationship between attitudes in reading and achievement in reading in 32 research studies. The study sought to provide a statistical summary of the observed variability of previously reported effect sizes. Results indicated that the mean strength of the relationship between reading attitudes and achievement was overall moderate ($Z_r = .32$); however the relationship was stronger for students in elementary school ($Z_r = .44$) in comparison with middle school students ($Z_r = .24$) (Petscher, 2010). Petscher also noted that the “reading attitudes” were commonly measured as a secondary function to a larger research interest. Further research was suggested that would require an increased precision of measurement and an effective approach for improving reading attitudes (Petscher, 2010).

In the current study, I conducted two statistical analyses with a duration measure for conditioned reinforcement for reading and then tested an intervention procedure that established acquisition of the conditioned reinforcement for reading in elementary-aged students. I propose that conditioned reinforcement for reading can be observably measured through sustained, extensive interest in reading a text (Greer & Ross, 2008). I propose that the acquisition of conditioned reinforcement for reading will be correlated directly with reading test scores, similar to the results found in Petscher (2010). The subsequent review of literature will establish the importance of conditioned reinforcement for reading and provide direct evidence for procedural designs.
Review of Literature

The primary focus of this study is increasing the reinforcement value of a specific behavior, in our case, reading. In order to accurately define this increase, we must first discuss the process through which reinforcement values can change and how we can apply these principles to reading. The following review of literature will discuss: 1) a brief overview of the concepts applicable to covert and overt responding, 2) the establishment of emergent behavior and its applications to reading behaviors, 3) current measurements of reading that provide evidence for experimental designs, and 4) previous research designs that provide models for intervention.

Overview

**Conditioned reflex.** In order to identify the principles included in conditioned reinforcement, we must acknowledge the initial research of Ivan Pavlov (1927). Ivan Pavlov was awarded the Nobel Prize in medicine for his research with canines, where he paired an unconditioned stimulus, food, with a neutral stimulus, a tone (Keller & Schoenfeld, 1950). In these series of studies, Pavlov measured the salivary reflex of the dog in order to determine if the bell tone would elicit the same behavioral response as the presence of the the primary reinforcer. Later considered to be “classical conditioning,” Pavlov developed an experimental method which studied the acquisition of new stimulus-response connections (Keller & Schoenfeld, 1950). Pavlov stated the new positive conditioned stimulus becomes “firmly established” through repeated repetitions of reinforcement (Pavlov, 1924). Subsequent studies also found that a "delayed" conditioned reflex may also be established if the presentation of the tone was followed by progressively later presentations of the unconditioned stimuli (Keller & Schoenfeld, 1950).
**Respondent behavior.** Pavlov’s results have been considered to be one of the first instances of research that measured the relationship between a stimulus and an organism’s response. Wolpe and Plaud (1997) claimed that Pavlov’s research has shown to be a “systematic basic learning paradigm” that was crucial for the foundation of behavior therapy. Alternatively called Pavlovian conditioning, the unconditioned stimulus is always observable and the reflex response is emitted automatically (Keller & Schoenfeld, 1950). The stimulus-elicited behavior is described as a respondent behavior, where the respondent behavior is the specific instance of the conditioned behavior to the stimuli. In his article “Two Types of Conditioned Reflex and a Pseudo Type,” Skinner distinguished his work from classical conditioning, explaining that there are two distinct types of conditioning. Skinner (1935) differentiated the two paradigms by stating, “Type 1 does not prepare for the reinforcing stimulus, it produces it.” This difference becomes essential in later discussions about how Skinner (1938) defined conditioned seeing as a “conditioned reflex” as opposed to an operant behavior. When applied to reading, this could shed light on why conditioned seeing occurs as a result of reading a textual stimulus.

**Operant behavior.** In his book *Behavior of Organisms* (1938), Skinner later clarified this sentiment by stating there were two distinct types of conditioning: Type S and Type R. Skinner (1938) labeled Type S conditioning as having similar traits to Pavlovian classical conditioning, while Type R was defined as *operant conditioning*, where if the occurrence of the operant was reinforced, the strength of the response increased. This voluntary response operates or interacts with the environment to produce a reinforcing stimulus. Keller and Schoenfeld (1950) theorized that a large proportion of an organism's behavior is emitted rather than elicited and although the initial behavior may have been spontaneous, the operant can quickly become associated with the stimulus. Skinner (1938) also proposed that behavior could be analyzed by
observing the chains of discriminative operants. He explained that each link of the operant chain can set the occasion for the subsequent operant and can, in turn, become a conditioned reinforcer for the operant that preceded it (1938). The operant behavior is then associated with the stimuli and the associated stimulus “sets the occasion” as a discriminative stimulus (Ormrod, 1999). As an organism’s behavior is shaped by the environment, stimuli that are similar to the discriminative stimulus can be generalized and generalized stimuli that retain these differences can be used to distinguish it from other stimuli (Carlson & Buskist, 1997).

**Overt and covert responding.** Within this framework, we can start addressing how respondent behavior and operant behavior have mutual effects on reading. Stolurow and Walker (1962) tested the effects of overt and covert responding during learning tasks on 56 subjects. In their initial explanation, the experimenters stated that there is a strong debate over which learning strategies are most effective; the intension was to see if overt responses, which were thought to be aligned with stimulus-response (S-R) theories such as those of Skinner and Thorndike, were more effective than covert responses, which were similar to Pavlov’s stimulus-stimulus responses (S-S) (Spence, 1951). The data indicated that the response modality did not have significant differences for verbal tasks, meaning that those participants who emitted overt or covert responses had equal criterion scores. However, the mean durations to complete the program were significantly different as the covert response group required less time than the overt response group. Because of this, the experimenters proposed that covert instruction was more efficient for learning practices. Stolurow and Walker (1962) also stated that only those from the covert learning group failed to attend the final retention test session. Experimenters suggested that this difference may link overt responding as more reinforcing than covert responding.
As we now circle back to implications for reading interventions, traditional methods of teaching reading repertoires covertly with silent reading and comprehension questions may not be effective for all students (Moxley, 1982). Pragmatist, John Dewey raised this concern in 1897 when he wrote “My Pedagogic Creed” and claimed that a lack of overt responses is a neglect that is “the cause of a large part of the waste of time and strength in school work.” He continued to state that when a student is placed into a “passive, receptive, or absorbing attitude” the learning environment becomes a place where children are not permitted to follow their active nature (Dewey, 1897). Eighty-five years later, Moxley (1982) modified this assertion as he proposed students develop covert repertoires after responding overtly in a discussion about the material (p. 13). He maintained, “the initial overt responses, however clumsy, and as unnecessary as they seem, were nevertheless essential for obtaining this result” (p. 14). If reading is “oversimplified” as Moxley (1982) proposes, how can we design an effective intervention that rotates through multiple topographies?

**Emergent Behavior**

Greer and Longano (2010) stated that generalizations and distinctions between stimuli become increasingly important as higher order verbal operants are being developed. Generalized operant response classes established by a history of reinforcement across exemplars can allow untaught of relations to emerge (Gomez, Lopez, Martin, Barnes-Holmes, & Barnes-Holmes, 2007). The emergence of untaught relations has been researched extensively. Stimulus Equivalence was a theory initially proposed by Sidman (1971), which sought to illustrate the emergence of an untaught relation between two stimuli. After extended research was conducted, RFT theorists re-examined Sidman’s Stimulus Equivalence’s terms and altered them to describe key components of language (Barnes-Holmes, Barnes-Holmes, Smeets, Cullen, & Leader, 2004).
These terms continue to be used in both the theory of RFT and other behavioral fields (Greer, Yuan, & Gautreax, 2005). The following will briefly discuss the influential research related to emergent relations.

**Stimulus Equivalence.** The term Stimulus Equivalence was used by Sidman (1971) to describe the relationship found during a match-to-sample procedure, where the subject emitted untaught responses to indirectly taught material. Further studies were conducted to ascertain the validity of the findings. Sidman and Tailby (1982) set out to define the relationships they found; they cited three terms that described all possible relations between stimuli. These include reflexivity, symmetry, and transitivity. Sidman continued to test his stimulus equivalence theory by experimenting with the roles of the listener and speaker. During the experiment, the participant was presented with three stimuli; after teaching the student to match the vocal utterance of “car” and the textual response for car, the participant emitted the untaught response for matching the textual response for car and a picture of a car (Sidman, 1986). Sidman (1986) found that there was a direct link between the speaker and the listener. Sidman (1994) also discussed that a stimulus can “evoke” either a non-verbal or verbal response.

Sidman and Tailby (1982) defined the three terms of Stimulus Equivalence as relations of reflexivity, symmetry, and transitivity. Reflexivity was defined as two related stimuli being equivalent. Therefore, Stimulus A was equivalent to Stimulus A. Symmetry referred to the relationship where if Stimulus A was equivalent to Stimulus B, then B was equivalent to A. The final relationship defined was transitivity, when if Stimulus A was equivalent to Stimulus B, and Stimulus B was equivalent to Stimulus C, then Stimulus A was equivalent to Stimulus C (Sidman & Tailby, 1982). These definitions were continued to be tested with additional studies (Place, 1995; Sidman, 1986). Many of the studies included a version of a match-to-sample,
through which untaught relations were shown to emerge from the mastery of equivalencies (Horne & Lowe, 1996; Sidman, 1971; Sidman, 1973; Sidman 1990).

**Relational Frame Theory.** Sidman’s Stimulus Equivalence served to function as a basis for what was later described as Relational Frame Theory (RFT). Hayes and Hayes (1989) analyzed and studied these stimulus/stimulus relations and re-identified the relations with three new terms: mutual entailment, combinatorial entailment, and the transformation of stimulus function. While some these terms shared similar components to Sidman’s Stimulus Equivalence, Hayes and Hayes (1989) sought to redefine equivalence within the behavioral treatment of language (Barnes-Holmes, Barnes-Holmes, Cullivan; 2000). These terms were studied primarily through a procedure called multiple exemplar instruction, where untaught responses were derived through teaching these relations (Barnes & Holmes, 1991; Barnes-Holmes & Barnes-Holmes, 2000; Barnes, Healy, & Hayes, 2000; and Barnes-Holmes, Barnes-Holmes, Smeets, & Roche, 2001). Additional studies were also conducted in order to test the limitations of stimulus equivalence. Lipkents and Hayes (1994) found that there were two types of organisms that could not be taught equivalence: children under 18 months and non-humans.

Through the empirical studies a theory was formed. Relational Frame Theory (RFT) defined relations of language as having two distinct applications. Language was either learned through the direct contingencies of reinforcement or arbitrary applied relational responding (Barnes-Holmes, Barnes-Holmes, & Cullen, 2000). Arbitrary applied relational responding was defined by the “relation” to the other stimulus (Barnes-Holmes, Barnes-Holmes, & Cullen, 2000). These two categories were developed through the research and close analysis of Skinner’s (1957) treatment of language. Barnes-Holmes, Barnes-Holmes, and Cullen (2000) defined the three RFT terms as mutual entailment, combinatorial entailment, and the transformation of
stimulus function. Mutual entailment was defined as the relation shown when the stimuli A=B and B=A. Combinatorial entailment was defined as the relation shown when A=B and B=C then A=C. The final term of the transformation of stimulus function occurred when an untaught stimulus acquired the equivalence of another stimulus that has no similar physical features and has never been directly taught (Barnes-Holmes, Barnes-Holmes, & Cullen, 2000).

RFT theorists continued to define other vocal operants, classifying the emission of a vocal response of a single word without frames of relation as non-verbal, while emission of vocal response with frames of relation as verbal (Barnes-Holmes, Barnes-Holmes, & Cullen, 2000). Blackledge (2003) reaffirmed some of components of Stimulus Equivalence, when he stated that the listener and speaker were “interlocked” with each other in a frame of relation. Certain stimuli had a relation as they share a common feature; this relation was defined as a frame of coordination. Other frames included: opposition, distinction, comparison, spatial, temporal, hierarchical, causal, and deictic (Barnes-Holmes, Hayes, & Dymond, 2001). Holmes, Smeets, Cullen, and Leader (2004) stated that the key difference between the two theories lies in the fact that RFT uses language as the key component. The researchers stated that each stimulus was directly related to environmental influences, and Stimulus Equivalence did not express these possible alterations (Barnes-Holmes, Barnes-Holmes, Smeets, Cullen, & Leader, 2004).

Most important to the current study, in their article “Teaching Derived Relational Responding to Young Children,” Barnes-Holmes, Barnes-Holmes and McHugh (2004) stated that recently, behavioral researchers have embraced what has been called a post-Skinnerian account of verbal behavior; this asserts that human language and cognition are equal and operate similarly. The researchers believed that arbitrarily applicable relational responding was a core process involved in human language and cognitive abilities (Barnes-Holmes, Barnes-Holmes &
McHugh, 2004). Stewart, Barnes-Holmes, Hayes, and Lipkens, (2001) add to this notion as they described that language and cognition have relating of relations and that these relating of relational networks link to other relational networks; complex relational skills are essential to development of analogical reasoning. Barnes-Holmes, Barnes-Holmes and McHugh (2004) argued that one of core assumptions of RFT is that verbal relational skills are the basis of a range of cognitive abilities and that the verbal relational skills correlate with educational achievement, cognitive skills, and the development of flexibility.

**Naming theory.** Horne and Lowe (1996) identified Naming as the verbal behavior that emerges from stimulus equivalence. Horne and Lowe (1996) focused on the contingencies that generate Naming as the coordination of listener and speaker repertoires that are necessary for equivalence to emerge. Naming was considered to have bidirectionality of speaker and listener responses, where if a child heard someone tact a stimulus as a listener, the child could then later produce the tact for the stimulus as a speaker (Horne & Lowe, 1996; Horne, Lowe, & Randle, 2004; Lowe & Horne, 1996; Lowe, Horne, & Hugh, 2005; Lowe, Horne, Harris, & Randle, 2002). However, while Horne and Lowe (1996) theoretically discussed the emergence of Naming, they stated that an extensive program of research was needed to explain how Naming occurred.

**Verbal Behavior Developmental Theory.** Greer and Keohane (2006) extended the original Naming theory identified by Horne and Lowe (1996) by seeking a means for inducing the emergent behavior within students. According to Greer and Keohane (2006), Naming was considered to be one of the three speaker-as-own-listener capabilities. Based on a program of research, Greer and Ross (2008) specified Naming as the joining of the listener and the speaker within the skin. Naming has two components: the listener and the speaker. Once individuals can
respond to a stimulus as a listener, those individuals can respond to the name of the same stimulus without direct instruction where as before they could not. Likewise, if a student possesses the speaker component of Naming, the student can respond to the stimulus as a speaker by hearing the name of a stimulus without direct instruction. Those students who possess Naming demonstrate learning language incidentally through their environment (Greer & Ross, 2008).

A series of research studies were conducted that demonstrated that multiple exemplar instruction was an effective procedure to induce Naming (Fiorile & Greer, 2007; Gilic, 2005; Greer, Stolfi, Chavez-Brown, & Rivera-Valdez, 2005; Speckman-Collins, Lee Park, & Greer, 2007). Greer, Stolfi, and Pistoljevic (2007) also related this emergent behavior as a higher-order class of verbal behavior. Multiple exemplar instruction functioned to be the source through which a history of reinforced relations emerged. Greer, Stolfi, and Pistoljevic (2007) even commented on the possibility of Naming as a relational frame, but neither mutual entailment or combinatorial entailment was accurately assessed before the experiment.

**Naming Joins Textual Responding.** Greer and Speckman (2009) defined verbal behavior as the language that functions both as speaker and listener through which the individual functions with others or within his or her own skin. While it is inherently more difficult to measure, behavior beneath the skin can be measured through overt speaker-as-own listener behaviors such as say-do, self-talk, and Naming (Greer & Ross, 2008). As stated before, the Naming capability demonstrates the presence of joint stimulus control from listener to speaker functions (Greer, Stolfi, Chavez-Brown, & Rivera-Valdez, 2005). Once established, this joint attentionality can be applied to other verbal operants, including textual responses. Greer (2008) suggested that students with the Naming capability and fluent phonemic decoding will
immediately acquire comprehension of the textual response as the textual response joins the
interlocking verbal operants for both speaker and listener. When applied to our current
discussion, the Naming Joins Textual Responding cusp begins to create a window for
understanding reading comprehension. Greer and Speckman (2009) provided an example of
Naming Joins Textual Responding with the word “elephant.” They explained that when a student
encounters a novel textual response and applies previously acquired phonemic decoding, the
listener within the skin hears “elephant” and as such comprehends the word in context with his or
her previous instructional history (Greer & Speckman, 2009).

A series of research studies were conducted to demonstrate the effects of acquiring fluent
decoding on Naming Joins Textual Responding (Helou-Caré, 2008; Lee Park, 2005; Reilly-
Lawson, 2008). In particular, Reilly-Lawson (2008), tested the effects of multiple exemplar
instruction on joining Naming to reading and writing. Experiment 1 tested three developmentally
delayed students who had previously acquired full Naming, but the Naming capability was not
joined to textual responses. Participants were presented with three types of probes in French: a
listener component where French words were matched with pictures; a speaker component where
participants tacted the image in French; and a written component where participants responded to
a picture with a written answer in French. Experiment 2 differed as the experimental materials
were altered to contrived stimuli; however multiple exemplar phonemic instruction (MEI)
remained as the intervention. Results demonstrated that the MEI intervention taught participants
to derive relations between seeing the printed letters, saying the sound of the letter, and saying
the name of the letter that corresponds with the sound, thus indicating that fluent phonemic
decoding was the source of the derived relations between Naming and reading comprehension
(Reilly-Lawson, 2008). Greer and Speckman (2009) later acknowledged that the Reilly-Lawson
(2008) study confirmed previous considerations between the role of reading fluency and Naming; the study demonstrated that mutual and combinatorial entailment are a part of the derivational relations between listening, speaking, and writing.

**Joint Control.** Lowenkron (1998) identified joint control as when a topography of a verbal operant, which has been evoked by one stimulus, can simultaneously evoke another stimulus; this event of joint stimulus control sets the occasion for a “special relation” between the stimuli (p. 327). These equivalent relations can be seen in our previous discussions of Stimulus Equivalence, RFT, Naming, and Naming Joins Textual Responding. Reading comprehension, similar to other types of verbal behavior, requires the development of joint stimulus control that includes instructional histories and multiple stimulus control (Greer, Yuan, & Gautreaux, 2003). Lowenkron (2006) utilizes the term joint control differently so as to include “the familiar notions” of operant stimulus control; however, the extended control is shown in instances where a single topography comes under the control of two verbal operants.

This concept becomes increasingly more important as our focus narrows itself to reading comprehension specifically. In his article, “Some Logical Functions of Joint Control,” Lowenkron emphasized that joint attention is required to comprehend written text as:

- Saying a word in response to its printed version (a textual response) provides a topography sufficient to allow the printed word to be selected in response to both its heard pronunciation and, if the word also functions as a tact, to the object it names.

  (Lowenkron, 1998)

This assertion parallels Reilly-Lawson (2008); with a prior history of reinforced tact productions, vocalizations of a textual response can set that occasion for a tact response from the reader. If a description is to be comprehended by an individual, its stimuli must be capable of functioning as
tacts for that subject (Lowenkron, 1998). Thus, in the case of an individual who “comprehends,” a textual stimulus may evoke a self-echoic that can also serve as tact. This stimulus equivalency can be included in a larger set of relations mediated through joint control.

When applied to our greater understanding of reading as verbal behavior, Lowenkron (1998) provides an example of when a string of textual stimuli work as autoclitics to the word “chair”. In this illustration, the phrase “small, red chair” would evoke three distinct tacts of “small,” “red,” and “chair.” The tacts “small” and “red” would then, in turn, be tacts as autoclitics, since the written stimuli describe the chair (Lowenkron, 1998). The relational responding between the stimuli would allow for the responses to generalize from previous instructional history to form a novel description of the chair. In this way, verbal behavior, whether written or spoken, builds upon “extensions of an autoclctic formula.” Skinner (1957) stated that this extension of the autoclctic formula orders fragments and responses within larger samples of verbal behavior (p. 331).

**Conditioned seeing.** Skinner (1951) maintained that all behavior is a function of our environment and any event in the universe is capable of affecting an organism; however, “part of the universe is enclosed within the organism’s own skin” (257). Therefore, concepts such as conditioned seeing can be one of the “most difficult problems in the analysis of behavior” (Skinner, 1953, p. 265). While conditioned seeing is inherently difficult to measure, Skinner (1953) defined conditioned seeing as a conditioned reflex that is emitted in the absence of the actual stimulus to a stimulus that had been previously paired the with presence of the stimulus. When applied to the Naming Joins Textual Responding cusp, it is possible that after fluently decoding the textual stimulus CAT, the individual may “hear” himself say “cat” and as such could then emit a “seeing” response with the image of a cat. Figure 1 is a proposed expansion of
the conditioned seeing diagram in Shanman (2013), where Naming Joins Textually Responding is joined with the initial presentations of stimuli.

**Figure 1.** Naming Joins Textual Responding diagram. Proposed expansion of Shanman (2013) as applied to the Naming Joins Textual Responding cusp.

With the joining of speaker-as-own listener and conditioned seeing, a reader could possibly comprehend sentences in context with his or her previous instructional history. Skinner (1957) asserted that large segments of verbal behavior resulting from autoclitic activity are called sentences (p. 345). He continued that a verbal stimulus may then “lead the listener to ‘see’” the sentences as a kind of conditioned seeing response (p. 363). Conditioned seeing could also explain why a reader may “see” the stimuli within context of one’s own previous history (p.
Figure 2 is a possible model of how larger segments of verbal behavior, in the form of sentences, could elicit an extended conditioned seeing response.

**Figure 2.** Proposed model of Skinnerian “sentences”. A possible model of a conditioned seeing in response to a sentence.

**Reading.** Skinner (1957) extended his concept of increasingly larger segments of verbal behavior into what he called “composition.” He explained:

Some sentences are more than the key responses on strong skeletal frames, or framed responses completed under the pressure to produce whole units. A set of variables may be so unusual or so complex that the past verbal behavior of the speaker yields no appropriate standard pattern. (p. 346)

The novelty of variables can be used to conditioned the behavior of the reader; the verbal stimuli can evoke a response from the reader that had affected the writer (p. 357). Because of this complexity, a reader could then experience a novel environment, which could be built on themes
that may have been withheld only because the occasion for the behavior had never occurred (p. 397). The reader then could “behave” within the novel scene as if he were able to behave in his environment. Skinner maintained that the novel’s description can evoke a “reader’s emotional reaction” and that this reaction can be considered to be “more than merely verbal” (p. 365).

In order to maintain the reader’s internal behavior with the text, Skinner (1951) asserted that the “reinforcing consequences” must be continued after the behavior has been acquired (p. 30). For any behavior, reinforcement is required to retain the responses’ strength (p. 31). When applied to reading specifically, a reader must continually have a positive interaction with texts so as to remain engaged with the material. A reader might then seek out other works of a given writer or other literature of a given type because of the reinforcement he has received (p. 272).

However, the opposite is also true: if there was a difficulty for the reader with the text due to issues such as clarity, familiarity of the terms, the ability to emit echoics, or the density of its autoclitics, a reader would not access positive interactions with the text (p. 367). This could eventually compound into overarching negativity about reading as the act itself is punishing.

With this sum of conceptualized understanding, we can consider how the ability to derive relations mediated through joint control and a possible “seeing” response can directly affect both reading comprehension as well as the reinforcement value of reading a text.

**Measurements of Reading**

Dennis Norris, a leading researcher from Cambridge University’s Medical Research Council Cognition and Brain Sciences Unit, stated that reading is an “impressive human achievement” that requires coordination of mastery in a constellation of perceptual and cognitive processes. He maintained this constellation gathers together an assortment of repertoires ranging from visual perception to eye-movement control, recognition of word forms, phonological
processing, and all the other higher-level linguistic processes required to recover an understanding of written words (Norris, 2013). In order for us to accurately measure reading behaviors and propose an effective reading intervention, we must examine how reading has been previously been measured and reported.

**Eye tracking.** Reading has been defined in various ways and at times the overall differences make identifying the process ambiguous for parents, educators, and researchers (Afflerbach, Pearson, & Paris, 2008; Tyson, 2014). In order to begin our discussion of the measurement of reading, we shall start with overt behaviors that can be observed. According to Rayner and Pollatsek (1989), eye fixation durations are commonly used as indicators of processing times for fixated words or segments during reading. Rayner, Pollatsek and Schotter (2012) state that while there are various ways of studying reading through eye tracking, word identification response-time measures are the most common. There are two opposing theories on how a reader identifies words: processing each letter and then the word as a whole or identifying words directly through a visual template of a word (Rayner, Pollatsek & Schotter, 2012).

However, the researchers assert that there is additional evidence that suggests a parallel-letter encoding model may be a better explanation of the reading than the simpler visual template model.

When applied to whole sentences, McConkie (1979) claimed that a reader’s eyes remain in place until a critical cognitive event occurs. Possible cognitive events include a completion of lexical access or a lack of information needed to make one. The movement of a saccade response, a simultaneous movement of both eyes in the same direction, defines the duration for eye fixation (Cassin & Solomon 1990). Ashby, Rayner, and Clifton (2005) found that highly skilled readers read high-frequency targets 17 ms faster than a group of average readers.
Interestingly, readers had similar gaze duration times for low-frequency words. The researchers indicated that for average readers, lexical access processes did not always control when the eyes moved, meaning that the average readers tended to move their eyes to the next word even though lexical access had not been completed (Ashby, Rayner, & Clifton, 2005). In general, processing theorists speculate that the decision of “when” to move the eyes is largely a function of cognitive processing (Rayner, Pollatsek & Schotter, 2012). For the current study, eye movement duration plays a key factor in observing students reading, however it is important to note that, Skinner (1957) cautioned that a text becomes “unintelligible” if read too rapidly, even in instances where the text was correctly echoed (p. 367).

**Comprehension.** While silent reading can be observed with eye tracking, reading silently to oneself presents an inherit difficulty – we cannot truly measure a covert response. To combat this obstacle, we need to explore supplemental research from other epistemologies. One theorist, Snow (2002), defined reading comprehension as the process of simultaneously extracting and constructing meaning through interaction and involvement with written language. Although not entirely scientific, the definition is effective at explaining the two key components: decoding words and the deriving meaning from them. Countless research studies has been conducted on reading comprehension and schema theory, but for the purposes of this study we shall explore the discourse of *situational models* and *language strategies*, which are cognitive psychology models that focus on the mental representations of verbally described situations (van Dijk & Kintsch, 1983).

According to van Dijk and Kintsch (1983), a language user, in both production and comprehension of verbal utterances used in communication, is consistently confronted with the performance and the understanding of an action. Zwaan and Radvansky (1998) reiterated the
statement by stating that readers construct situational representations in conjunction with text-based representations. This type of treatment of language resonates with the post-Skinnerian account of verbal behavior which was discussed earlier; arbitrarily applicable relational responding parallels the situational model in the sense that readers are constructing understanding of the text through the textual relations. While van Dijk and Kintsch were not directly referring to relational responding, they did state *language strategies* are “complement” to the account of verbal behavior.

Several researchers have maintained that the construction of a coherent situation model is crucial to the successful comprehension of a text (Zwaan, Magliano, & Graesser, 1995). Bransford et al. (1972) stated that written stimuli in the form of sentences are information that can used by a reader to create “semantic descriptions” of situation. Similarly, Zwaan and Radvansky (1998) claimed that “rather than treating language as information to analyze syntactically and semantically and then store in memory, language is now seen as a set of processing instructions on how to construct a mental representation of the described situation”. These sentiments were tested empirically when Yarkoni, Speer, and Zacks (2008) used functional magnetic resonance imaging (fMRI) to study the neural mechanisms supporting situation model processing. Twenty-nine participants were provided with blocks of sentences that were either unrelated to one another or formed coherent narratives. By differentiating the blocks, the experimenters sought to see if there were differences between narrative-level comprehension and sentence-level comprehension. Results illustrated that most brain regions showed activation during narrative-level comprehension, with a lesser extent during the sentence-level comprehensions. Furthermore, a subsequent memory analysis revealed that there
was an increase in comprehension and memory performance for the coherent narratives as opposed to the sentence-level or word-level encoding.

In a final study, researchers Schneider and Korkel (1989) compared the recall of soccer “experts” with novices in order to test situation model processing. One hundred and eighty-five children across 3-7th grade were given a multiple choice pre-assessment for previous knowledge of soccer. Students were grouped into experts and novices according to grade level. When provided with a narrative text dealing with a soccer game, experimenters tested participant recall of a soccer story that was comprised of 36 elements. Overall, third grade “experts” recalled more elements than both third and fifth grade novices and fifth grade “experts” outperformed 7th grade novices. The results indicated that the high-knowledge students had fewer problems constructing a situation model because they could assemble the situational model by retrieving relevant knowledge structures from their long-term memory. The researchers also stated that although domain-specific knowledge has been previously shown to have strongly influenced text processing, the inclusion of a domain of interest was particularly effective. This brings our discussion of reading back to Skinner’s argument: a reader seeks out other texts because of the reinforcement he has received from reading them. With this we can infer that readers who can comprehend a given text will be more likely to continue to read than those who struggle with decoding and mediating relations between words.

**Reporting.** During the last four decades, there has been a significant emphasis on reporting reading outcomes for students across the country. Both federal and state legislative reforms have been passed in order to help promote student growth especially in subjects such as reading and mathematics. In particular, standardized testing for reading comprehension has been implemented as a way to create a set level of requirements for academic performances. Thurlow
and Johnson (2000) stated that “high stakes” standardized tests are often used to determine (1) if a school is progressing in achieving annual yearly performance goals in accordance with the No Child Left Behind (NCLB) Act of 2001, (2) if a student will be promoted or retained based on individual performance on minimum competency exams, and (3) if a student is eligible for high school graduation diplomas based on mandatory exit exams.

All of these components combined foster a considerable amount of pressure on both the students and the school districts. High-stakes testing has increased the emphasis on accountability and is likely to challenge both educational institutions and the students affected by them (Katsiyannis, Zhang, Ryan, & Jones, 2011). While there have been dozens of articles written on the negative outcomes of high stakes testing for children with disabilities and children from low-socioeconomic backgrounds, there is limited research conducted on the general population (Havner, 2005; Lewis, 2004; O’Neill, 2003; Rowe, 2004).

Haertel (1999) stated that if reading standards were clearly identified and students were taught the material allowing them to meet the standards, standardized testing for reading is the logical approach to identify students who did not meet expectations, as well as the teachers of those students. This would then create a system for holding students, teachers, and schools accountable for ensuring that all students met expected reading standards (Haertel, 1999). Kent McGuire, Assistant Secretary from Office of Educational Research and Improvement (U.S. Department of Education), commented that “judgments need to be made about curriculum and instruction, expanded learning opportunities before and after school, developing the talents of teachers, and even improving the quality of the assessments used to measure performance” (2000). Federally approved state standards or adoption of the Common Core State Standards with the national high stakes tests, such as the PARCC and SBAC, can become strong indicators
of struggling readers and be able to provide data on which areas need improvement (Fitchett and Heafner, 2010).

There has been a struggle to find the balance of accountability for students, districts, parents and legislators. The current practice is to publish test scores; Smith (1991) advised against this practice since publishing test scores with schools’ rankings in local newspapers, pressured teachers to produce high test scores, thus causing teachers “anxiety, shame, loss of esteem, and alienation”. McGill-Franzen (2000) agreed with Smith (1991), commenting that the “top-down policy” is not helpful to teachers. Other authors have reiterated the opinion maintaining high-stakes test scores have detrimental effects on teaching (Johnston, 1998). Research suggests that the higher the stakes on a given test, the greater the chance of teachers teaching to the test, thus becoming a detriment to other aspects of teaching/learning (Herman & Golan, 1991; Johnston, 1998; Smith, 1991). This being said, Thomas J. Kane, director of the Center for Education Policy Research at Harvard University, pointed out that such a strong reaction against standardized testing would “be equivalent to saying ‘O.K., because there are some players that cheated in Major League Baseball, we should stop keeping score, because that only encourages people to take steroids” (Augustine, 2003). Despite the opposing sides, U.S. school teachers must accept the unavoidable trend: standardized tests are here to stay. Therefore, it is essential to include standardized testing as an additional measurement in relation to our analysis of reading repertoires.

**Research Models for Intervention Design**

As we start to conclude our discussion on how reading behaviors are established and measured, it is important to look towards prior research that can lend validity to the proposed intervention. In Experiment 3, the independent variable was a peer-yoked contingency procedure
that sought to increase the reinforcement value of reading by providing opportunities for participants to derive meaning from novel words and complete comprehension drawing tasks. After exiting the intervention, the experimenter assessed if increases in the reinforcement value of reading functioned to increase the acquisition of reading repertoires. The following review of previous research will create a base on which the final study’s foundation was built upon.

Greer, Becker, Saxe and Mirabella (1985) utilized a pair/test procedure in order objectively measure the process through which a stimulus is conditioned. The procedure has been replicated and applied to numerous items, such as books, blocks, toys, and print stimuli (Greer, Pistoljevic, Cahill, & Du, 2011; Longano, & Greer, 2006; Pereira-Delgado, Greer, Speckman, & Goswami, 2009). In their article Greer and Du (2013) also emphasized the importance of establishing conditioned social reinforcers as well; they identified pre-verbal foundations of verbal behavior to be the result of the onset of conditioned reinforcers for observing responses. The induction of social stimuli as reinforcers is thought to be the key to ontogenic development of language (Greer & Du, 2013). Once conditioned, reinforcing social stimuli can be applied to both academic interventions (Park, Pereira-Delgado, Choi & Greer, 2008) and the expansion of repertoires (Pereira-Delgado, Greer, Speckman, & Goswami, 2008).

A large body of research has sought to utilize the reinforcing qualities of social attention in order to shape additional conditioned reinforcers. Numerous studies have developed conditioning procedures that incorporate the observation of social attention, in particular social attention provided to peers. These studies were successful in inducing conditioned reinforcement for: candy (McCorkle, 1988), cereal (Greer & McCorkle, 1994), tokens (Greer & Sales, 1998), discs and strings (Greer & Singer-Dudek, 2008), and books (Singer-Dudek, Oblak, & Greer, 2011). Not only have observation procedures been shown to be effective at conditioning objects,
but they have also been show to establish conditioned reinforcement for teacher vocal praise (Greer, Singer-Dudek, Longano, & Zrinzo, 2008) and academic subjects (Lee, 2016; O’Rourke, 2006). Additionally, studies such as Davis-Lackey (2005), Gautreaux (2005), and Stolfi (2005) have utilized social attention provided by peers in a yoked contingency, which has shown to also establish capabilities such as observational learning.

The effects of conditioning procedures are not limited to increasing the reinforcement value of the intended stimuli. Studies such as Tsai and Greer (2006) and Buttigieg (2015) have demonstrated that conditioning procedures can also increase learning responses. Tsai and Greer (2006) examined the effects of conditioning books as reinforcers for observing responses on the learning of textual responses with preschool children. The study used a stimulus-stimulus paring procedure to condition books as a preferred activity and as a result the number of learn unit-to-criterion of textual responses decreased after the conditioning intervention. Buttigieg (2015) mirrored these results, however, the embedded independent variable allowed for three different methods of conditioning book stimuli: textual operant discrimination training, Pavlovian second order conditioning, and conditioning books through peer observation. In context of our current discussion, we start to evaluate how increases in the reinforcement value of reading could affect the acquisition of reading repertoires.

**Rationale**

In his book, *The Assayer*, Galileo asserted that a “book [of philosophy] cannot be understood unless one first learns to comprehend the language and read the letters in which it is composed” (1623). So then, how can we teach a reader to “comprehend” the language? This study seeks to expand the body of research pertaining to the development of a reader and how increases in reinforcement value can affect measurable outcomes in reading tests scores.
In Experiment 1, I sought to design a reliable, observable measurement for conditioned reinforcement for reading. It was theorized that if a student consistently chose to continue to read, the student was reinforced by the action of itself; therefore, by observing the duration of reading, experimenters could behaviorally measure the reinforcement value of reading. In addition, it was hypothesized that there would be a correlation between conditioned reinforcement for reading and standardized test scores. Experiment 2 searched for prerequisite reading repertoires that could affect the reinforcement value of reading. The final experiment worked towards finding an intervention that could increase the acquisition of reading repertoires as a function of increasing the reinforcement value of reading.

I propose that once an individual acquires the Naming Joins Textually Responding cusp, in conjunction with speaker-as-own listener and conditioned seeing behaviors, students will be able comprehend larger segments of verbal behavior. This comprehension can expand from Skinner’s initial definition of sentences, which are generated by adding autoclitics to available verbal operants (Skinner, 1957, p. 345). I agree with Skinner (1957) that verbal stimuli can then “lead the listener to ‘see’” larger segments of verbal behavior in a kind of kind of conditioned seeing response (p. 363). I believe that reading becomes increasingly more reinforcing as an individual can create stimulus equivalency for relations mediated by joint control and as such have a better ability to “see” the presented text.
Research Questions

The research questions addressed in this study are as follows:

1) Does acquiring conditioned reinforcement for reading increase reading test scores?

2) Are there prerequisite repertoires that can affect an individual’s reinforcement value of reading?

3) Can an intervention that rotates opportunities to derive meaning and draw pictures for comprehension with peers increase the reinforcement value of reading?

4) Do increases in the reinforcement value of reading also affect the acquisition of reading repertoires?
Chapter II

EXPERIMENT 1

Method

Participants

Eighteen students participated in Experiment 1, with equal numbers of males and females. Of these participants, 50% were white, 22% were black, 22% were Hispanic, and 5.6% were Asian. Participants had a mean age of 10.55 years (SD=0.28, Range 10.10-11.10). 44% of the students participated in New Jersey’s Free or Reduced Lunch Program, 22% had individualized education plans (IEPs), 16% were classified as having a learning disability (504 Plan), and one student was determined to be a Basic Skills student, which is a local school district categorization for students performing below grade level.

All participants attended a publically funded Title 1 School for grades 3-5 in a suburb outside of a metropolitan area in the Northeast. With respect to school-wide demographics, approximately 7% of the students spoke another language at home, 35.8% participated in the Free/Reduced Lunch Program, and 22% were classified with a disability (State of New Jersey Department of Education, 2014). The school had five 5th grade classrooms; one of which implemented the Comprehensive Application of Behavior Analysis to Schooling (CABAS®) Accelerated Independent Learner (AIL) educational model (www.cabasschools.org). The CABAS® model classroom applied empirically validated methods of pedagogy, curricula and sequence, classroom management, and training for staff and parents (www.cabasschools.org). Instruction within the classroom was based on short and long-term objectives derived from state and school standards. In addition to district and statewide assessments, the participants’ verbal
behavior development was assessed using the Verbal Behavior Development Assessment-Revised (VBDA-R) (Greer, 2010).

**Procedure and Measures**

The experimenter compiled participant data across a one-year longitudinal study. Data were collected through multiple pre/post measurements of outcome variables over the course of a school year. The measured variables included: 1) statewide standardized testing, 2) districtwide standardized testing, 3) 20-min observational probes, and 4) probes for participant preferences. The repeated measurements were administered with the initial pre-assessments in May and post-assessments completed in June. For example, every participant had three initial observational probes that occurred in September, 2013 and a second set of three observational probes in June, 2014. The repeated reading assessments included: NJASK Reading Scores, DRA scores, and observational probe trials conducted to measure the reinforcement value of reading.

Over the year, there were approximately 18 days of reading assessments. While no more than one reading assessment was administered a day, the durations of assessments varied. The NJASK Reading Assessment was conducted across the grade level over two consecutive days in the month of May, each session for 90 minutes. Short breaks were permitted and the administrator provided students with intermittent vocal praise. DRA assessments differed as the assessment was untimed; participants were provided with vocal reinforcement for completion of sections, but not accuracy. Figure 3 outlines the experimental timeline.
Figure 3. Experimental Assessment Timeline. This figure provides a visual representation of the sequence of the initial reading assessments conducted in 2013 and the subsequent reading assessments conducted in 2014.
**Statewide standardized testing scores: NJASK.** According to the State of New Jersey Department of Education (NJDOE), *New Jersey Assessment of Skills and Knowledge* (NJASK) was the statewide standardized assessment given in the month of May for grades 3 through 8. Educational Testing Service (ETS) developed the NJASK Assessment in order to evaluate student educational success rates and to provide informative data for school districts (ETS, 2013). 100,323 students partook in the assessment across the public school systems of the state during the 2014 window (NJDOE, 2015). At the time, the NJASK was accepted as having content validity as it aligned to the New Jersey Core Curriculum Content Standards (Rothman & Henderson, 2011). Tienken and Wilson (2014) reported that the ELA portion of the assessment had a validity range between .80 and .82. The assessment contained two separate scores; one score for English Language Arts (ELA) and one score for Mathematics. For the focus of this study, only the ELA scores were analyzed.

Over the course of two sessions, students completed multiple questions with a total of 62 opportunities to respond to both open-ended and multiple-choice options (NJDOE, 2014). Questions were generated to assess a range of skills, including sub-components of Reading and Writing such as literature, informational texts, and persuasive arguments. Scores from both Reading and Writing sections were reported and combined to create an ELA raw score. ETS then weighted the raw scores to establish a normal distribution across the state. Final NJASK ELA scores ranged from 150 to 300, where students were considered to be “Partially Proficient” with a score between 150-199, “Proficient” with a score between 200-249, and “Advanced Proficient” with a score between 250-300. Final percentiles reported for 2014 included 37.35% of participants scored as “Partially Proficient,” 54.35% as “Proficient,” and 8.3% as Advanced Proficient” (NJDOE, 2015).
**District-wide reading assessment scores: DRA.** The Developmental Reading Assessment (DRA) is a district-wide reading assessment that tests both reading fluency and comprehension level. Teachers and specialists individually administer the assessment twice a year in order to assess each student’s grade level equivalency. The DRA also provides an opportunity to test specific reading repertoires such as predictions, literal comprehension, summary, interpretation, reflection, and metacognitive awareness. Each of the repertoires are scored with a range of 1-4, 4 being considered as “Advanced”. Once completed, scores are summed and a final fluency and comprehension score are calculated. The resulting score corresponded with a reading grade level and a Lexile® Framework equivalence; for example, a DRA score of 40 would be a 4th grade reading level and a Lexile® score of 771.

Participants for the study were provided with four novella options, two of which were fiction and two of which were non-fiction. The four novellas ranged in length and difficulty based on the reading level. After selecting the preferred text, experimenters directed the participant to read a short section aloud. Data were collected on participant accuracy and rate. If the participant met criterion for both repertoires, the participant was directed to write predictions about the text. Participants recorded their responses and were instructed to finish reading the text independently and complete the given packet. Experimenters scored the final product; if the participant did not score within the “Independent” or “Advanced” range, a new assessment was administered on a lower reading level. For example, a participant score as “Intervention” on a level 40 DRA was required to retest at a level 38.

**20-min observations of reading behaviors.** In order to measure the reinforcement value of reading, a duration probe was conducted in blocks of 20 min. The experimenter recorded observational data in 10 s whole intervals; the experimenter recorded a plus (+) if the participant
was emitting observable reading behaviors throughout the entire 10 s interval. Observable reading behaviors were defined as 1) the participant’s eyes tracking across the book page from left to right; 2) the participant’s eyes returning to the leftmost word and repeating to eye tracking behavior; and 3) the participant’s eye tracking behavior continues onto the next page after reaching the end of the text. The experimenter recorded a minus (-) if the student’s eyes were not tracking across the page or if the student was attending to an unrelated stimulus. As a way to accurately account for disturbances during the observant-ion, the experimenter continued to record a plus if the participant was distracted by an external stimulus for less than 3 s.

The experimenter summed the frequency of pluses (+) to calculate the observed reading. Scores could range from 0-120 intervals. Participants who were observed to read for 96 of the 120 intervals were considered to have conditioned reinforcement for reading (R+ Reading). The criterion percentage of 80% was based on previous research on conditioned reinforcement (Greer, Pistoljevic, Cahill, & Du, 2011; Pereira-Delgado, Greer, Speckman, & Goswami, 2009; Longano, & Greer, 2006). Observations were conducted for each participant during September and June. The data established a pre-post measure of R+ Reading for each student over course of one school year. See Appendix A for the 20-min observation data sheet.

**Participant reading preference.** As a fourth measure, the experimenter observed participants for five consecutive sessions to identify if reading was a preferred activity. Sessions were interspersed across the month and at different times of the day. Participants were asked to choose between completing a set of worksheets from previously mastered objectives (ELA/Math) or to “free read”. Experimenters defined “free reading” as the participant independently selecting a novel to read and quietly attending to the book. Data were collected in blocks of 5 opportunities so as to be sensitive to both internal and external influences. The
number of sessions in which the participant chose to “free read” was calculated, with the scores ranging from 0-5 sessions.

**Inter-observer and Inter-scorer Agreement**

Due to the variance in measure, measurement fidelity differed for each variable. Prior to the study, all scorers completed calibrations with the experimenter and the following agreements are between the independent scorer and the experimenter. For the behavioral measurements, the 20 min observation and the preference probe, the experimenter utilized inter-observer agreement. The inter-observer agreement (IOA) scores were calculated by dividing the total numbers of point-by-point observer agreements by the total numbers of agreements plus disagreements, and then multiplying that number by 100%. IOA was obtained for a total of 95% of preference probe sessions with 100% agreement. IOA was also conducted for a total of 10.53% of whole-interval probe sessions with 99.58% (99.17%-100%) agreement. The DRA Assessment was scored through an inter-scorer agreement (ISA) scores, which were calculated by dividing the total numbers of paired observer agreements by the total numbers of agreements plus disagreements, and then multiplying that number by 100%. Initial calibration was completed across scorers and final ISA scores was conducted for 35% of the DRA® assessments, with a mean score of 92%. NJASK scores were reported to the experimenter by the state, and as such had no ISA.

**Data Analysis**

All measurements were compiled and analyzed in IMB’s Statistical Package for the Social Sciences (SPSS), a software package used to evaluate data. The experimenter coded all reading assessments into two variable types, numerical and categorical variables, in order to perform multiple analyses. For example, a given participant would have two types of variables related to conditioned reinforcement for reading in 2014: a numerical value of the 10s whole-
intervals and a categorical value of attainment of 80% criterion. Additionally, participant
descriptions were coded for such variables as educational modifications and socioeconomic
status. Since preliminary findings showed that categorical factors such as IEP classifications, 504
classifications, and qualifications for Free/Reduced Lunch Program were not significant
predictors of test scores, the experimenter began examining relationships between standardized
testing scores and observational measurements. See Table 1 for more information about coding.
<table>
<thead>
<tr>
<th>Assessment</th>
<th>Variable Name</th>
<th>Type</th>
<th>Definition of Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th Grade NJASK</td>
<td>NJ4Raw</td>
<td>Scale</td>
<td>NJ4Raw score was defined as the number of correct responses to the 62 questions in 2013.</td>
</tr>
<tr>
<td>NJASK Assessment</td>
<td>NJ4ELA</td>
<td>Scale</td>
<td>NJ4ELA was defined as the weighted score of reading and writing responses reported in 2013.</td>
</tr>
<tr>
<td></td>
<td>NJ4Reading</td>
<td>Scale</td>
<td>NJ4Reading was defined as the weighted score for reading responses only reported in 2013.</td>
</tr>
<tr>
<td></td>
<td>NJ4Ord</td>
<td>Ordinal</td>
<td>NJ4Ord was coded as 0-4 for the categorization of Partially Proficient, Proficient, and Advanced Proficient in 2013.</td>
</tr>
<tr>
<td>5th Grade NJASK</td>
<td>NJ5Raw</td>
<td>Scale</td>
<td>NJ5Raw score was defined as the number of correct responses to the 62 questions in 2014.</td>
</tr>
<tr>
<td>NJASK Assessment</td>
<td>NJ5ELA</td>
<td>Scale</td>
<td>NJ5ELA was defined as the weighted score of reading and writing responses reported in 2014.</td>
</tr>
<tr>
<td></td>
<td>NJ5Reading</td>
<td>Scale</td>
<td>NJ5Reading was defined as the weighted score for reading responses only reported in 2014.</td>
</tr>
<tr>
<td></td>
<td>NJ5Ord</td>
<td>Ordinal</td>
<td>NJ5Ord was coded as 0-4 for the categorization of Partially Proficient, Proficient, and Advanced Proficient in 2014.</td>
</tr>
<tr>
<td>September 2013 DRA</td>
<td>DRASept</td>
<td>Nominal</td>
<td>DRASept was defined as the DRA Reading Assessment grade level in 2013.</td>
</tr>
<tr>
<td></td>
<td>DRASeptLexiEq</td>
<td>Scale</td>
<td>DRASeptLexiEq was coded as the Lexile® equivalent score of the DRA for the use of scale analysis in 2013.</td>
</tr>
<tr>
<td>June 2014 DRA</td>
<td>DRAJune</td>
<td>Nominal</td>
<td>DRAJune was defined as the DRA Reading Assessment grade level in 2014.</td>
</tr>
<tr>
<td></td>
<td>DRAJuneLexiEq</td>
<td>Scale</td>
<td>DRAJuneLexiEq was coded as the Lexile® equivalent score of the DRA for the use of scale analysis in 2014.</td>
</tr>
<tr>
<td>20 min Observation Probe 2013</td>
<td>W120Sept</td>
<td>Scale</td>
<td>W120Sept was defined as the observed number of 10 s whole-intervals a participant read in 2013.</td>
</tr>
<tr>
<td></td>
<td>AcquisSept</td>
<td>Nominal</td>
<td>AcquisSept was coded as 0 or 1 for participants who were observed to read for 80% of the probe trials in 2013.</td>
</tr>
<tr>
<td>20 min Observation Probe 2013</td>
<td>W120June</td>
<td>Scale</td>
<td>W120June was defined as the observed number of 10 s whole-intervals a participant read in 2014.</td>
</tr>
<tr>
<td></td>
<td>AcquisJune</td>
<td>Nominal</td>
<td>AcquisJune was coded as 0 or 1 for participants who were observed to read for 80% of the probe trials in 2014.</td>
</tr>
<tr>
<td>Acquisition of R+ Reading 2013-2014</td>
<td>AcquiredR+</td>
<td>Nominal</td>
<td>AcquiredR+ was coded 1-3 for the categorization of participants who Never Acquired R+ Reading, Acquired During 2013-2014, and Previously Acquired R+Reading.</td>
</tr>
</tbody>
</table>
Results

Spearman Analysis for Correlated Test Scores

Initial analysis was conducted to ascertain if there was a relationship between the two reading assessments during each of the probe windows. Based on these results, those participants who had higher reading testing scores on the NJASK during the 2013 year also tended to have higher scores on the DRA that was conducted in September, 2013. This was also the case for the post measurements analyzed from June, 2014. This relationship indicated that while the testing format and response types varied, participant data across the reading assessments retained a similar level of performance within the timeframe of the measurement occurred.

With the establishment of these relations between assessments, a series of analyses were completed between the scale scores of three variables: number of whole intervals of reading, NJASK ELA score, and the Lexile Level® equivalency for DRA scores. These variables were compared solely with the year in which they were collected, so as to investigate possible corresponding relationships between standardized testing scores and observational data. For example, NJASK 2014 scores were analyzed only with other variables measured from 2014. Both Pearson and Spearman analyses were conducted; trends were comparable, but only the Spearman’s rho correlations were reported as the sample size was limited. The results of the Spearman’s rho correlations found that the standardized assessment scores and observational scores based on the experimental timeline were significantly correlated with one another. Therefore, it can be stated that a participant in 2014, who was observed to have a higher number of reading intervals, was also more likely to have a higher score on his/her standardized tests score in 2014. See Tables 2 and 3 for exact correlation coefficients.
One-way ANOVA Analysis for Categorical Value of R+ Reading

A secondary analysis was conducted in order to find potential reading score differences between participants with distinctive acquisition levels for conditioned reinforcement for reading. Three categorical variables were created from the scale observational data: 1) student who never acquired R+ Reading, 2) students who acquired R+ during the 2013-2014 academic year, and 3) students who acquired R+ reading previously. These categorical variables were then
compared using an ANOVA. The results of the one-way ANOVAs showed an overall significance between the Acquisition of R+ and the following variables: NJASK ELA 5 ($F(2, 15) = 8.10, p = .004, \eta_p^2 = .519$); NJASK Reading 5 ($F(2, 15) = 10.38, p = .001, \eta_p^2 = .581$); June Lexile Score ($F(2, 15) = 28.65, p = .000, \eta_p^2 = .793$), and June Reading Preference ($F(2, 15) = 8.0, p = .004, \eta_p^2 = .518$). In all four analyses, those participants who did not acquire conditioned reinforcement during 2013-2014 had the lowest mean scores on reading assessments, while those who had acquired conditioned reinforcement previously had higher mean scores on the reading assessments. The third category, Acquired During the 2013-2014, all had mean scores between the two opposing categories; however, the mean scores varied in the closeness of range. See Figures 4-7 for information on the categorical mean scores.
**Figure 4.** One-way ANOVA Analysis with Categories for R+ for Reading and Mean NJASK5 ELA Score. This figure shows the mean scores from the NJASK ELA Assessment in 2014 when participants are placed into three categories based on their reading observational totals. Error bars denote standard error.

**Figure 5.** One-way ANOVA Analysis with Categories for R+ for Reading and Mean NJASK5 Reading Score. This figure shows the mean scores from NJASK Reading Assessment in 2014 when participants are placed into three categories based on their reading observational totals. Error bars denote standard error.
Figure 6. One-way ANOVA Analysis with Categories for R+ for Reading and Mean June Lexile Scores. This figure shows the mean scores from the June Lexile® Equivalency when participants are placed into three categories based on their reading observational totals. Error bars denote standard error.

Figure 7. One-way ANOVA Analysis with Categories for R+ for Reading and Selecting Reading. This figure shows the mean scores from the June Lexile® Equivalency when participants are placed into three categories based on their reading observational totals. Error bars denote standard error.
Discussion

The results of this study indicated that conditioned reinforcement for reading was highly correlated with reading test outcomes. These findings contribute to the growing body of literature for conditioned reinforcement, especially in context with Tsai and Greer (2006). The results of the Spearman’s rho correlations found that the standardized assessment scores and observational scores based on the experimental timeline were statistically significant. These findings demonstrated that those participants with lower scores on district and state testing also were observed to have shorter intervals of reading, while participants with higher scores on district and state testing were observed to read for longer intervals.

When participants were placed into three categories based their observational scores, the mean scores of each group reflected the previous results on all standardized assessments for both pre and post measures. The results confirmed that the acquisition of conditioned reinforcement for reading was correlated to all assessments required by the state and district during a given school year. The data analysis demonstrated that those students who never acquired conditioned reinforcement for reading also had the lowest mean reading scores, while students who had acquired conditioned reinforcement previously had the highest mean reading scores. Additionally, those who acquired conditioned reinforcement for reading during the school year, were shown to have higher mean scores than those who did not acquire the repertoire; however, those scores were still below the mean scores of those who had previously acquired conditioned reinforcement for reading.

The analysis was limited by the sample size of participants. With 18 students, each subgroup population ranged from 4–6. Inclusion of more participants would lend more weight to the analysis. After completion, it was also clear that in the next experiment required additional
testing measures to give more information about specific reading repertoires that were associated with conditioned reinforcement for reading. Both the DRA and NJASK assessments require repertories such as comprehension and inferencing; Experiment 2 sought to test for the correlations between variables that could affect reading scores.
Chapter III

EXPERIMENT 2

Method

Participants

Thirty-three students (48.5% male, 51.5% female) participated from inclusion fifth grade elementary classrooms. Of these participants, 72.7% were white, 12.1% were black, 12.1% were Hispanic, and 3% were Asian. Eight of the 33 students participated in New Jersey’s Free or Reduced Lunch Program, 8 students had individualized education plans (IEPs), 4 students were classified as having a learning disability (504 Plan), and 3 students were identified through the Intervention and Referral Services (I&RS), which is a district-wide program for students performing below grade level. Overall, 45.5% of the participants required academic modifications to district and state reading assessments, while the remaining 54.5% did not receive any additional accommodations. Experiment 2 was conducted in the same Title 1 school for grades 3-5 as Experiment 1. As stated before, the classroom implemented the Comprehensive Application of Behavior Analysis to Schooling (CABAS®) Accelerated Independent Learner (AIL) educational model and the participants’ cusps and capabilities were assessed using the Verbal Behavior Development Assessment-Revised (VBDA-R) (Greer, 2010).

Procedure

The experimenter utilized a two-year cross-sectional design in order to create a larger sample size for the study. Data were collected through measurements of outcome variables during two school years 2014-2015 and 2015-2016. All probe trials were conducted between the months of May and June. For example, complete data sets were collected from 11 participants between the months of May and June of 2015; this included: a 20-min observation probe, a
combined conditioned seeing and derived response probe trial, a DRA assessment, and an ELA PARCC score. Data from the other 22 participants were collected similarly in May and June of 2016. The assessment durations ranged from 3 non-consecutive days to 20 minutes depending on the measurement. The experimenter did not give multiple assessments in a single day.

Measures

The experimenter compiled data from the following assessments: 1) statewide standardized assessment (PARCC), 2) a district-wide reading assessment (DRA), 3) a 20-minute reading observation, and 4) 15-target response probe for conditioned seeing and derivational responding.

Statewide standardized testing scores: PARCC. Partnership for Assessment of Readiness for College and Careers (PARCC) is a group of states collaborating to develop assessments that provide better information to educators so that standardized testing can enhance instruction to meet the individual needs of the student body. The PARCC assessment is comprised of two subjects: English language arts and mathematics. The PARCC ELA assessment covered both fiction and nonfiction reading, as well as multi-media videos that provided additional information on the tested topic. Nichols-Barrer, Place, Dillon, and Gill (2015) reported that those students who scored “college-ready” on PARCC’s English/language arts section had an 89 percent probability of earning at least a C average across all their freshman-year courses. The testing was completed on a computer-based assessment system supplied by Pearson, where students logged into the TestNav system and were presented with both multiple choice and essay responses. Similar to the NJASK, ELA PARCC reported raw scores and weighted scores to establish a normal distribution across the state.
**District-wide reading assessment scores: DRA.** Developmental Reading Assessment (DRA) was a district-wide reading assessment that tested both reading fluency and comprehension level. This was the same district measurement used in Experiment 1. To review, teachers and specialists administered the assessment in order to assess each student’s grade level equivalency. The resulting scores were matched to both a reading grade level and the district Lexile® Equivalent Guided Reading Leveling Chart.

**20-min observations of reading behaviors.** The experimenter observed each participant for a duration of 20 min in order to record student reading. As described earlier, the experimenter recorded observational data in 10 s whole intervals; the experimenter coded a plus (+) if the participant was emitting observable reading behaviors throughout the entire 10-second interval. Conditioned reinforcement for reading was defined as the participant reading for 96 of the 120 intervals.

**15-target response probe for conditioned seeing and derivational responding.** In addition to the previous measurements found in Experiment 1, two novel repertories were tested: conditioned seeing and derivational responding. Data for each repertoire was recorded separately, but the overall probe procedure delivered the antecedents in conjunction to one another. The probe design sought to test how a novel textual stimulus could be comprehended through both a derivational relation of coordination between textual stimulus and how the novel stimulus could be “seen” in the context of a sentence.

In order to create the probe trial stimuli, the experimenter selected researched novel textual stimuli and generated a novel sentence that included the targeted textual stimuli. For example, the experimenter wrote the sentence, “The pirate struck the **linstock** and lit the cannon’s gunpowder with its flame.” The novel sentence was then submitted to a committee of
three educators; the committee were asked to 1) draw a picture of the entire sentence, and 2) write a possible derivational relation of the target stimuli in bold. If the correct response was emitted by all three committee members, the novel textual stimuli was considered to be valid stimuli for participant probe trials. See Table 3 for stimuli.
### Table 4

**Textual Stimuli and Novel Sentences Used in Experiments 2 and Experiment 3**

<table>
<thead>
<tr>
<th>Textual Stimuli</th>
<th>Novel Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>lazer</td>
<td>The <strong>lazer</strong> stayed in the hospital for weeks before the doctors could diagnosis his illness.</td>
</tr>
<tr>
<td>maw</td>
<td>After eating a six-course meal, my <strong>maw</strong> was so full that I thought I would explode.</td>
</tr>
<tr>
<td>bulwark</td>
<td>We had to create a <strong>bulwark</strong> in order to protect ourselves from the storm.</td>
</tr>
<tr>
<td>guidon</td>
<td>A <strong>guidon</strong> was hung high on the pole, waving in the wind.</td>
</tr>
<tr>
<td>linstock</td>
<td>The pirate struck the <strong>linstock</strong> and lit the cannon’s gunpowder with its flame.</td>
</tr>
<tr>
<td>abolutions</td>
<td>The woman quickly finished her <strong>abolutions</strong> in the shower, as she scrubbed and rinsed her hair.</td>
</tr>
<tr>
<td>dale</td>
<td>The girl skipped down through the <strong>dale</strong> and admired the flowers in its peaks and hills.</td>
</tr>
<tr>
<td>redolent</td>
<td>As the man walked into the flower shop, he was overwhelmed with the <strong>redolent</strong> smell of roses.</td>
</tr>
<tr>
<td>liniment</td>
<td>The nurse handed the patient a small bottle of <strong>liniment</strong> to rub into the deep cut.</td>
</tr>
<tr>
<td>skein</td>
<td>The hair dresser carefully pinned the <strong>skein</strong> of client’s soft hair into the other curly waves.</td>
</tr>
<tr>
<td>bodkin</td>
<td>The knight swung his <strong>bodkin</strong> as the villain attacked the princess.</td>
</tr>
<tr>
<td>pallium</td>
<td>Little Red Riding Hood pulled her <strong>pallium</strong> over her shoulders to block the wind.</td>
</tr>
<tr>
<td>fust</td>
<td>When the girl opened the container that had been in the refrigerator for weeks, she saw that <strong>fust</strong> had grown all over the bread.</td>
</tr>
<tr>
<td>bilboes</td>
<td>The prisoner’s <strong>bilboes</strong> around his feet kept him from running away.</td>
</tr>
<tr>
<td>gib</td>
<td>The <strong>gib</strong> pounced on the mouse as it scurried across the living room.</td>
</tr>
</tbody>
</table>
Once validated with the committee, the experimenter compiled 15 target textual stimuli and their corresponding sentences. The target stimuli were randomly ordered into three probe groups of five stimuli. The experimenter created brief Microsoft PowerPoint presentations, where each target stimulus was presented to the participants in the following order: 1) a 30 s opportunity for the participants to read the sentence with the target stimulus in bold; 2) a slide with the antecedent to “Draw a picture of the entire sentence” in order to measure comprehension of both the novel textual stimulus and its relation to other textual stimuli in the sentence; and 3) a final slide with the antecedent to “What does the word BLANK mean?” to assess the accuracy of the derived relation. While all slides were only presented to the participants once, the duration of participant response for the second two response slides were untimed.

Each presentation was blocked into 5 target stimuli opportunities. Three sessions were conducted over the course of four weeks for a total of 15 conditioned seeing responses and 15 derived responses. The experimenter presented the slides on a projector in the front of the classroom. All desks were oriented in the same direction as the presentation. Participants were directed to silently read the sentence with the target stimulus in bold. The experimenter changed the slide after providing a 30 s opportunity to read the sentence. Participants were then provided with a blank 10.16 cm by 13.97 cm piece of paper and were given both a vocal and a written antecedent to “Draw a picture of the entire sentence”. After completion, the experimenter collected the conditioned seeing response paper with the participant’s drawing response. A second blank, lined piece of paper was then supplied to the participants and the experimenter instructed the participants, both vocally and visually, to silently write a possible derivational relation of the novel stimulus. Following the collection of the derived response, the second target
stimulus was displayed with the novel sentence. This process continued until all five of the target stimuli were presented for that session. See Figure 6 for a diagram of the conditioned seeing/derived responding probe procedure. Appendix B displays versions of correct and incorrect responding.

Figure 8. Procedure for Collecting Probe Data for Conditioned Seeing and Derivational Responding. The figure above displays both the PowerPoint slides shown to the participants and the procedure for collecting the probe data.
**Inter-observer and Inter-scorer Agreement**

Like Experiment 1, measurement fidelity differed for each variable; all scorers completed calibrations with the experimenter before the study and the following agreements are between the independent scorer and the experimenter. The experimenter utilized inter-observer agreement for the 20 min reading observations. The inter-observer agreement (IOA) scores were calculated by dividing the total numbers of paired observer agreements by the total numbers of agreements plus disagreements, and then multiplying that number by 100%. IOA was obtained for a total of 27.2% of whole-interval probe sessions with 90% agreement. Inter-scorer agreement was used for the DRA assessment. The inter-scorer agreement (ISA) scores were calculated by dividing the total numbers of paired observer agreements by the total numbers of agreements plus disagreements, and then multiplying that number by 100%. ISA was conducted for 21% of the DRA® assessments with a mean score of 96%.

The conditioned seeing and derived response probe trials also utilized ISA, however the process was conducted through a committee. The committee consisted of three educators; two of the educators were CABAS trained while the third was a blind observer. For each session, two of the educators were randomly selected to complete initial scoring; both scorers recorded a plus or a minus for each response. If the scorers agreed, the data were recorded and considered the valid score. However, if there was a disagreement, the third observer would score the response. The committee scored all conditioned seeing and derived response probe trials (100%) with a mean agreement score of 86%.
Data Analysis

Similar to Experiment I, all measurements were compiled and analyzed in IMB’s Statistical Package for the Social Sciences (SPSS). Each participant had multiple categorical and scale variables. For example, a participant had categorical variables such as educational plan and scale variables which were collected from district and statewide test scores. Additionally, categorical values were created to provide opportunities to compare the differences in participants acquiring specific repertoires, including the acquisition of conditioned reinforcement for reading (R+Reading), the acquisition of conditioned seeing (R+ CS), and the acquisition of derived responding (R+ DR). The experimenter analyzed the data set through Pearson correlations, independent T-tests, a one-way ANOVA and multiple linear regressions.

Results

Preliminary Findings

Initial analysis sought to determine if any participant categorical data were correlated with testing scores. Each participant was coded with gender, ethnicity, socioeconomic status, and educational modification data. All categorical values were then evaluated in relation to acquisition of repertoires (R+R, R+ CS, and R+ DR), test scores (DRA and PARCC), and number of correct responses during the reading observations. Gender, ethnicity, socioeconomic status, 504 eligibility, and IRT eligibility were not correlated with any testing data. However, it was found that participants who were eligible for an IEP were more likely to have lower correct responding to: a) number of observed reading intervals (r(33) = -.590, p = .000), b) number of correct responses to CS probes (r(33) = -.456, p = .008), and c) number of correct responses to DR probes (r(33) = -.570, p = .001). Participants with IEPs were also correlated with lower
scores for both ELA PARCC Score ($r(27) = -.734, p = .000$) and DRA Score ($r(33) = -.679, p = .000$). Due to this inconsistency, subcategories were originally analyzed separately; however, the Shapiro-Wilk test for normality found that the ELA PARCC (S-W = .971, $df = 27, p = .641$) as well as the DRA (S-W = .944, $df = 33, p = .091$) had normal distributions. It was decided that because the results demonstrated normality across standardized tests and was being conducted in an inclusion setting, all participant data would be included for analysis.

**Independent t-Test for Correlated Test Scores**

Following that information, independent t-tests were conducted to test for correlations between the acquisition of the three tested repertoires and the variable scale scores. Variable scale scores included: 1) DRA score, 2) ELA PARCC score, 3) number of intervals observed to be reading, 4) number of correct responses to conditioned seeing probe trials (CS), and 5) number of correct responses to derivational response probe trials (DR). Overall, those participants who did not acquire any of the repertoires were more likely to have lower district and standardized testing scores than their peers who acquired all three repertoires. This was also the case those who had acquired conditioned reinforcement for reading and the number of intervals for CS and DR, acquisition of conditioned seeing and the number of intervals for R+ and DR, and the acquisition of derived responses and the number of responses to CS probes. The only variables that did not correlate were the acquisition of derived responding and the number of intervals of observed reading; this was interpreted to mean that the acquisition of derived responding was not directly related to longer durations of reading. See Tables 5-7 for exact correlations.
### Table 5

*Independent Group t-test between Acquisition of R+ Reading and Variable Data*

<table>
<thead>
<tr>
<th></th>
<th>Did Not Acquire R+ Reading</th>
<th>Acquired R+ Reading</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>DRA</td>
<td>841.777</td>
<td>103.419</td>
<td>945.000</td>
</tr>
<tr>
<td>ELA PARCC</td>
<td>748.642</td>
<td>25.400</td>
<td>777.538</td>
</tr>
<tr>
<td>#CS Responses</td>
<td>7.611</td>
<td>4.146</td>
<td>11.666</td>
</tr>
<tr>
<td>#DR Responses</td>
<td>9.722</td>
<td>4.026</td>
<td>12.400</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01 N=33

*Note. M=Mean. SD=Standard Deviation. CS=Condition Seeing. DR=Derived Response*

### Table 6

*Independent Group t-test between Acquisition of R+ Seeing and Variable Data*

<table>
<thead>
<tr>
<th></th>
<th>Did Not Acquire R+ Reading</th>
<th>Acquired R+ Reading</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>DRA</td>
<td>853.250</td>
<td>99.065</td>
<td>943.230</td>
</tr>
<tr>
<td>ELA PARCC</td>
<td>752.562</td>
<td>25.043</td>
<td>777.090</td>
</tr>
<tr>
<td>Reading Intervals</td>
<td>59.550</td>
<td>47.305</td>
<td>98.923</td>
</tr>
<tr>
<td>#DR Responses</td>
<td>9.950</td>
<td>12.461</td>
<td>12.461</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01 N=33

*Note. M=Mean. SD=Standard Deviation. DR=Derived Responding*
Since all three acquired repertoires were significantly correlated with district and standardized testing scores, the next set of analyses sought to find potential score differences between participants with distinctive acquisition levels. One categorical variable, Combinations of Acquisition, was coded for each of the participants; each number corresponded to a combination of possible repertoires. For example, if a participant did not demonstrate acquisition for any repertoire, the score coded was 0, while a participant who acquired both R+ Reading and R+ DR were coded as a 5. Numerical values were not a continuous scale, and therefore did not have definitive value. All possible combinations were systematically added in order to assess if there were any repertoires were more likely to be associated with high mean scores.

Two ANOVA analyses were conducted, one for the DRA scores and one for ELA PARCC scores. Both one-way ANOVAs demonstrated that there were significant differences
between the combinations of acquisitions and mean testing score for the DRA ($F = 3.012; df = 6; p = .023, \eta_p^2 = .41$) and the ELA PARCC ($F = 4.01; df = 6; p = .008, \eta_p^2 = .546$). There were minimal differences found between the categorical variable and the order of mean scores in the two ANOVA results. Those participants who had only acquired repertoires such as conditioned seeing and derived responding were more likely to have lower mean scores on assessments. In contrast, those who had acquired conditioned reinforcement for reading were more likely to have higher test scores than their peers who did not. See Figures 7 and Figure 8 for additional information.

**Figure 9.** One-way ANOVA Analysis between Acquisition of Reading Repertoires and Mean Lexile Scores. This figure displays the mean Lexile scores for each reading repertoire and their possible combinations.
**Figure 10.** One-way ANOVA Analysis between Acquisition of Reading Repertoires and Mean ELA PARCC Scores. This figure displays the mean ELA PARRC scores for each reading repertoire and their possible combinations.

**Regression Models**

A final analysis was conducted to investigate if incremental gains in each repertoire would have a corresponding effect on testing scores. A stepwise-regression analysis was used to estimate a model that would correctly predict the effect of gains of each repertoire scale score on both the DRA and ELA PARCC scores. For example, would the number of correct responses to the derived responding probes help predict DRA scores? As stated earlier, prior to conducting the analysis, descriptive statistics were generated to examine the test assumptions such as normality for distributions. Two regressions were completed, one for each of the testing scores.
Additional step-wise variables were included into the models based on the ANOVA mean score found in prior results.

Regression 1 sought to estimate a model that would correctly predict the increases of DRA scores based on scale scores of three variables: number of correct responses to derived probe trials, number of correct responses to conditioned seeing probe trials, and the number of intervals of observed reading. Each model step included the previous predictor and calculated the increase in the change. Therefore, Model 1 had one variable: number of correct responses to derived probe trials. Model 2 expanded into two variables: number of correct responses to derived probe trials and the number of correct responses to conditioned seeing probe trials. The final Model 3 was comprised of all three variables. The results demonstrated that Model 3 better explained the variance of DRA scores, which mirrored results found in ANOVA data. Additionally, the change in each Model had a similar increase to the previous Model. See Table 8 for Regression 1 data.
Regression 2 examined estimate models that would predict the effects of ELA PARCC scores based on the same three scale scores variables. 3 identical step-wise model inputs were used, which were based on previous ANOVA data. The results demonstrated that Model 3, with all three variables, better explained the variance of ELA PARCC scores. While these results are similar to Regression 1, the change in each step was not as proportional. The inclusion of number of correct responses to condition seeing probe trials did not change the means of model as greatly as the other variables. See Table 9 for details.

Table 8

Summary of Hierarchical Regression Analysis for Predicting Lexile Scores (N = 33)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
<td>β</td>
</tr>
<tr>
<td>#DR</td>
<td>14.79</td>
<td>4.27</td>
<td>.528</td>
</tr>
<tr>
<td>#CS</td>
<td>11.59</td>
<td>3.51</td>
<td>.501</td>
</tr>
<tr>
<td>#Reading Intervals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.279</td>
<td>.471</td>
<td></td>
</tr>
<tr>
<td>$F$ for $\Delta R^2$</td>
<td>11.997**</td>
<td>10.862**</td>
<td>8.556**</td>
</tr>
</tbody>
</table>

Note. #DR = Number of correct responses to Derivational Responses. #CS = Number of correct responses to Conditioned Seeing Responses. #Reading Intervals = Number of intervals read during observation.

**p < .01.
The results of Experiment 2 demonstrated that conditioned reinforcement for reading continued be to highly correlated with reading text outcomes. Additionally, results indicated that repertoires such as conditioned reinforcement for seeing and derivational responding may also effect reading outcomes as well. Overall, those participants who did not acquire any of the tested repertoires were more likely to have lower district and standardized testing scores than their peers who acquired one or more of the repertoires. Furthermore, those participants that had only acquired repertoires such as conditioned seeing and derived responding were more likely to have lower mean scores on assessments in comparison to their peers that had acquired conditioned

---

**Discussion**

The results of Experiment 2 demonstrated that conditioned reinforcement for reading continued to be highly correlated with reading text outcomes. Additionally, results indicated that repertoires such as conditioned reinforcement for seeing and derivational responding may also affect reading outcomes as well. Overall, those participants who did not acquire any of the tested repertoires were more likely to have lower district and standardized testing scores than their peers who acquired one or more of the repertoires. Furthermore, those participants that had only acquired repertoires such as conditioned seeing and derived responding were more likely to have lower mean scores on assessments in comparison to their peers that had acquired conditioned reinforcement.
reinforcement for reading. These findings mirror to the previous study and help provide evidence for possible reading interventions.

While the data were primarily analyzed to find replications to support Experiment 1, the comparison of related reading repertoires and their specific effects on reading test scores can also be used to better inform educators about subcomponents required for success. In particular, t-values from the first analysis can be used to interpret which reading repertoires are most influential for different reading assessments. For both the PARCC and the DRA, the acquisition of conditioned reinforcement for reading was the highest indicator of test scores with t-values of 3.456 and 3.556 respectively. This was also true of the acquisition of conditioned seeing with PARCC t-vales of 2.696 and DRA t-values of 2.888. Most interestingly, there is a considerable difference in the effects of derivational responding on testing. Whereas the DRA has a t-value of 2.324, the PARCC has a t-value of 3.164. These data suggest that the most effective way to change reading scores for the DRA would be to teach conditioned seeing repertoires over derivational responding. Conversely, the more effective way to increase reading scores on the PARCC would to be concentrate on deriving relations from novel words as their t-value correlation for derivational responding is only three tenths less than conditioned reinforcement for reading. This implication lends to our previous discussion of comprehension. It would seem that assessments, which primarily measure fluency and comprehension through written responses, are more centered around participants emitting speaker-as-own listener responses; whereas assessments that involve multiple choice responses may be better suited to test inferring the meaning of key vocabulary would require a strong instructional history with derivational responding.
Results from the ANOVA tables also demonstrate key differences in the measurement of reading repertoires. In both analyses it is clear that the more reading repertoires acquired, the higher the testing scores. Those participants who acquired conditioned reinforcement for reading were more likely to have higher test scores than those who did not. The ANOVA data helps confirm earlier statements about the two reading assessments; DRA mean scores significantly increased with the inclusion of the acquisition of conditioned seeing, just as the PARCC mean scores increased when a participant was more effective at deriving a relation of a word. With this confirmation, an unexpected factor was also discovered: the conditioned seeing repertoire was never independent of any other reading repertoire. This could support Shanman (2013) as it presents conditioned seeing as a relation between a Naming response with visual stimuli. It is possible that speaker-as-own-listener and conditioned seeing behaviors, which have been discussed as the building blocks of comprehension, must be emitted simultaneously in the presence of the textual stimulus for effective reading. Conditioned seeing may not have been emitted independently as the response may be what Skinner called a “conditioned reflex” (Skinner, 1953).

The final stepwise-regression analysis was used to estimate a model that would correctly predict the effect of gains of each repertoire scale score on both the DRA and ELA PARCC scores. Like what was indicated with the ANOVA analysis, increases in reading repertoires were also related to better explanations of variances between testing scores. When analyzing the DRA regression, each reading repertoire better explained the model with similar increases in the change. For example, the change of variance between the inclusion of derived responding and conditioned seeing was calculated at 11.997 and 10.862 respectively. The PARCC variance did not have comparable changes, but could be interpreted to mean that those participants who had
not acquired any of the reading repertoires were significantly below those who had acquired one or more repertories. Additionally, the inclusion of number of correct responses to conditioned seeing probe trials did not change the means of the model as greatly; it is possible that the responses tested in the PARCC may have a larger concentration of derivational responding over other reading repertoires. Experiment 3 sought to tested the effects of a peer-yoked contingency procedure on reading comprehension as a function of the establishment of conditioned reinforcement for reading.
Chapter IV

EXPERIMENT 3

Method

Participants

Four students participated in this study. All attended the same publically funded Title 1 School for grades 3-5 in a suburb outside of a metropolitan area in the Northeast during academic school year 2016-2017. As stated in earlier studies, the inclusion classroom implemented the Comprehensive Application of Behavior Analysis to Schooling (CABAS®) Accelerated Independent Learner (AIL) educational model. In addition to district and statewide assessments, the participants’ level of verbal behavior and cusps and capabilities were also assessed using the Verbal Behavior Development Assessment-Revised (VBDA-R) (Greer, 2010). See Table 10 for additional participant information.

The experimenter conducted pre-intervention probes, during which the experimenter tested for previously acquired prerequisite repertoires such as conditioned seeing for novel sentences and derived relations for novel textual stimuli. The experimenter also recorded individual participant reading durations through the same 20 min reading procedure as the earlier studies. With these data, students who did not demonstrate conditioned reinforcement for reading, but had acquired conditioned seeing for novel sentences and derived relations for novel textual stimuli, were identified as possible participants. The experimenter assigned participants into dyads based on the results of the conditioned reinforcement for reading probes and district reading assessments.
Setting and Materials

All experimental probes were administered in either the participants’ classroom, or in the hallway directly outside of the classroom. Specifically, experimenters recorded duration data in a large group setting. Before the duration began, the experimenter directed all students to select a preferred novel that was on their preassigned reading level. All novels in the classroom were labeled with a color coordinated system for reading grade levels. Students were informed of their assigned color, but also had the option to choose a personal book from their library or household. If the preferred novel did not come from the classroom library, the experimenter confirmed the grade-level equivalency through the publisher. The experimenter did not conduct duration probes immediately after the selection so as to allow participants the opportunity to change their preferred text.

Table 10

Participant Descriptions for Experiment 3

<table>
<thead>
<tr>
<th>Participant</th>
<th>PARCC ELA</th>
<th>DRA</th>
<th>i-Ready Reading/Grade Equivalency</th>
<th>Naming Joins Txt</th>
<th>OL</th>
<th>TSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant A</td>
<td>766</td>
<td>40</td>
<td>565/4th</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Participant B</td>
<td>732</td>
<td>40</td>
<td>560/4th</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Participant C</td>
<td>730</td>
<td>40</td>
<td>577/4th</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Participant D</td>
<td>735</td>
<td>40</td>
<td>556/4th</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Data collected for the relational analogy probe were conducted outside in the hallway to limit exposure to the experimental stimuli and procedure. Relational analogy stimuli were created through a combination of resources. The experimenter examined and altered analogies that were designed as test preparation materials for the Secondary School Admissions Test Middle Level Assessment (SSAT). The SSAT Middle Level Assessment is a standardized test utilized by admission officers to assess the abilities of students who intend to enroll in an independent school who are currently in grades 5-7; the SSAT measures the verbal, math, and reading skills students need to be successful in independent schools (SSAT, 2016).

Of the 16 analogies created for relational analogy probe, three examples were generated from Cracking the SSAT & ISEE, 2017 Edition. The experimenter used the initial relation, the correct response, and one of the three possible exemplars for the experimental questions 2, 3, and 10. For example, the antecedent for Question #2 was “Rest is to exhaustion” and the correct response was defined as “as water is to thirst”. This relation was an exemplar that was previously published by The Princeton Review (2016). The Kaplan SSAT & ISEE 2016: For Private and Independent School Admissions test preparation book was also utilized for Question #11 for the initial relation, but not for the correct response or the other exemplars. All other relational analogies were produced by the experimenter.

Once all relational analogies were compiled, the experimenter tested the possible probe materials on three naïve adults. The committee members answered all 16 opportunities independently without the experimenter present, but were then provided an opportunity to discuss the stimuli for clarity of errors. Three analogies required alterations. An additional 10 naïve committee members were chosen and completed the altered relational analogies. Each responded correctly with at least 15 of the 16 correct responses. Those that errored found that #1
had two possible correct responses. The antecedent “Sky is to earth” was edited to “Cloud is to ground” which refined the spatial relational analogy to exclude a possible incorrect response of as “chair is to table”. A final pre-experimental probe was conducted with a non-participating student from another classroom to ensure that the analogies retained the primary focus of being accessible to students between 5-7th grade. The student performed above grade level in math and reading and responded correctly for 13 of the possible opportunities.

All intervention sessions were also conducted outside in the hallway to limit exposure to the participant pair responses. Intervention materials were designed by the experimenter and were contained in a single binder that could be shared as the participant dyads responded to one another through written feedback. Graphic organizers were used to establish two distinct components of the intervention: an opportunity for participants to derive meaning from novel words and the completion of the comprehension drawing task. See Figure 9 for examples of the intervention materials.
Figure 11. Intervention Materials. This figure displays the two components to the yoked contingency procedure: an opportunity for participants to derive meaning from novel words and the completion of the comprehension drawing task.

**Dependent Variables**

There were four dependent variables in this study: 1) 20-min observation of reading behaviors, 2) the Woodcock Johnson III Diagnostic Reading Battery, 3) the Gray Silent Reading Tests, and 4) a 16-opportunity relational analogies probe.

**20-min observations of reading behaviors.** To briefly reiterate, the experimenter measured the reinforcement value of reading by observing participants reading for 20 min durations. The experimenter recorded observational data in 10 s whole intervals; the experimenter recorded a plus (+) if the participant was emitting observable reading behaviors throughout the entire 10 s interval. Observable reading behaviors were defined as 1) the...
participant’s eyes tracking across the book page from left to right; 2) the participant’s eyes returning to the leftmost word and repeating to eye tracking behavior; and 3) the participant’s eye tracking behavior continues onto the next page after reaching the end of the text. The experimenter recorded a minus (-) if the participant’s eyes were not tracking across the page or if the participant was attending to an unrelated stimulus. The experimenter continued to record a plus if the participant was distracted by an external stimulus for less than 3 s, but returned back to the text. Participants who emitted a reading behavior for 96 of the 120 intervals (80%) were considered to have conditioned reinforcement for reading (R+ Reading). Two 20 min durations, for a total of 40 minutes, were recorded as pre-probes and another two 20 min durations were recorded as post probes.

**Standardized reading assessment: Woodcock Johnson III Diagnostic Reading Battery (WJRB).** In order to ascertain reading grade level equivalencies for both overt and covert responding, the experimenter chose two standardized reading assessments. The first was the Woodcock Johnson III Diagnostic Reading Battery (WJRB) which required all responses to be emitted vocally; the reading battery consisted 10 separate reading subtests that ranged from skills such as letter-word identification to oral reading comprehension. For the purposes of this study, Subtests 2 and 4 were used to complete the “Reading Comprehension” cluster. All subtests were recorded as raw scores, where the basal was established for each subtest through provided algorithms based on age and approximate grade level. Antecedents became increasingly more difficult as the subset continued. The experimenter concluded the assessment after four incorrect responses. Raw scores were then imputed into the Woodcock Johnson III Diagnostic Reading Battery software. The software program used the raw scores to create standardized scores for each subtest and overarching clusters. Standardized testing scores were correlated with
both grade level and age level ranges, as well as normalized percentiles. The cluster has been reported to have a median reliability of .91 in the age 5-19 range (Schrank, Mather, & Woodcock, 2004).

Test 2, “Passage Comprehension”, required that participants read a short passage and vocally respond with a missing key word that matched the context of the passage. The experimenter directed the participants to silently read the sentence/passage and emit a single vocal response. If the participant did not respond after 30sec, the experimenter provided a vocal prompt to attempt the response. If the participant did not respond after the second prompt, the experimenter instructed the participant to “Try the next one.” This process continued until the participant incorrectly responded to four consecutive items. In the case of varied correct responding, the incorrect count was reset each time, but the participant did not receive credit for any of the incorrect responses. For example, the approximate basal level for Participant A began at number 17. The participant silently read the antecedent, “A is for APPLE and ______ is for BAT.” The assessment’s answer key defined that “B” was the only correct response. Participant A continued to emit correct responses until question number 36, which stated:

With the stock market continuing its meteoric rise, investment publications and newspapers are rife with articles about operations set up to dupe the finically naïve. In response, several advisory services have been established to warn any prospective _______ of the scams

Participant A emitted the incorrect response “people” and the first instance of consecutive incorrect responses was recorded. After three additional incorrect responses were recorded, the
experimenter ended the session. Participant A’s final raw score for Test 2 was recorded as 28 with the basal score at 16 plus the 12 correct responses.

Test 4, “Reading Vocabulary”, included three categories: synonyms (4A), antonyms (4B), and analogies (4C). There were equal numbers of opportunities to respond to the synonyms and antonyms antecedents (24), however the analogy responses were limited to 21. For Tests 4A and 4B, the experimenter asked the participants to silently read the word and provided a vocal response. Only one vocal response was accepted for each question. For example, in 4A the word “kitten” was presented on a blank sheet; a “1” was recorded if the participant responded with the word “cat”. This continued until a ceiling of four incorrect responses were emitted. 4C differed as the participant was asked to read the analogy aloud and then make a vocal response. Analogies consisted of frames of coordination, opposition, comparison, and hierarchical; however, the other four frames were not tested. For example, the written opposition analogy “dessert…dry lake…” required a correct response of “wet”. As stated earlier, all raw scores were imputed into the Woodcock Johnson III Diagnostic Reading Battery program. Both raw scores and standardized scores are reported pre and post intervention.

**Standardized reading assessment: Grays Silent Reading Tests.** The second standardized reading assessment was chosen as a couplet to WJRB to measure participants’ covert comprehension reading grade level so as to compare overall comprehension grade level equivalencies. The Grays Silent Reading Tests (GSRT) are standardized reading assessments that use short passages with five multiple choice responses that are directly related to the text. One of the key advantages of the GSRT included having two forms (A and B), which limited participant exposure to the reading material. Similar to the WJRB, the assessment provides an initial basal starting point based on the participant’s age. Participants then completed progressively more
difficult stories until a ceiling of three incorrect responses were emitted for one story. All stories were exactly one paragraph; however, the difficulty of the vocabulary and the mean sentences length increased as the participant proceeded to the next story. For example, Story 3 had sentences such as “A girl ran out of the white house into the backyard,” while Story 6 had more complex sentences such as, “Much of our understanding about sharks come from the discoveries of Eugenie Clark.” After reading each Story, participants were directed to recorded their responses on a multiple-choice bubble answer sheet and each story was administered one at a time. The experimenter did not mark the multiple-choice answer sheet during the administration of the test, though the number of correct responses were monitored for the exiting criterion.

**Relational analogies probe.** The experimenter expanded the initial coordinational derivation probes to assess other possible derivational relationships between two stimuli. Relations included: opposition, distinction, comparison, spatial, temporal, hierarchical, and causal. In order to effectively measure the ability to identify these different types of relations, the experimenter created a relational analogies probe that consisted of 16 analogies; participants were provided with two opportunities to respond to each of the eight relations in a multiple choice format. The relational analogy opportunities were randomly arranged on one piece of paper (8.27 in × 11.7 in) and the duration for completion of the probes were untimed. Each assessed relation was presented in identical format with no prompting towards the type of relation being measured. See Appendix C for probe materials.

For example, when assessing the capacity for identification of stimuli in opposition of one another, the experimenter generated an initial analogy: “sick is to healthy”. Below the first analogy, four possible analogies were presented, each with a different relationship between them. Participants were asked to circle the analogy that had the same relation as initial exemplar. In
this opportunity, the correct response was defined as the participant selecting Option B, “jailed is to free”, as sick and healthy are in opposition to one another, just as jailed and free are in opposition. See Table 11 for the relational stimuli.

Table 11

*Relational Analogy Stimuli and Type of Derivational Relationship Used in Experiment 3*

<table>
<thead>
<tr>
<th>Type of Derivational Relationship</th>
<th>Relational Analogy Stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Causal</td>
<td>Laughter is to joke as cry is to pain</td>
</tr>
<tr>
<td></td>
<td>Rest is to exhaustion as water is to thirst</td>
</tr>
<tr>
<td>Comparison</td>
<td>Cube is to sphere as square is circle</td>
</tr>
<tr>
<td></td>
<td>Week is to year as pint is to quart</td>
</tr>
<tr>
<td>Coordination</td>
<td>Clog is to shoes as cap is to hat</td>
</tr>
<tr>
<td></td>
<td>Sun is to solar as moon is to lunar</td>
</tr>
<tr>
<td>Opposition</td>
<td>Sick is to healthy as jailed is to free</td>
</tr>
<tr>
<td></td>
<td>Drip is to pour as bend is to break</td>
</tr>
<tr>
<td>Spatial</td>
<td>Cloud is to ground as head is to foot</td>
</tr>
<tr>
<td></td>
<td>Basement is to roof as roots are to leaves</td>
</tr>
<tr>
<td>Temporal</td>
<td>September is to December as middle is to end</td>
</tr>
<tr>
<td></td>
<td>Breakfast is to dinner as sunrise is to sunset</td>
</tr>
<tr>
<td>Distinction</td>
<td>Raspberry is to strawberry as glove is to mitten</td>
</tr>
<tr>
<td></td>
<td>Pen is to pencil as right hand is to left hand</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>Hanger is to airplane as stable is to horse</td>
</tr>
<tr>
<td></td>
<td>Bird is to migration as bear is to hibernation</td>
</tr>
</tbody>
</table>
Independent Variable

The independent variable was a peer-yoked contingency procedure that sought to increase the reinforcement value of reading. There were two components: opportunities for participants to derive meaning from novel words and the completion of the comprehension drawing task. Both components were separated into two steps with a total of four intervention steps for each session. The intervention steps were as follow: 1) overt reciprocal reading, 2) selection of reinforcing textual stimuli and deriving meaning of novel words, 3) covert independent reading, and 4) a comprehension drawing task. Steps 1 and 2 established a pairing of shared reinforcement for textual stimuli; the rotation of listening to one another read required a joint attention for the text and the subsequent step created an opportunity for reinforcement for attending to the task. Steps 3 and 4 mirrored this process of joint attention; however, the steps were completed independently. In order to find the initial paragraph length for Steps 1 and 3, the experimenter calculated the mean duration of the pairs’ reading across all 20-min conditioned reinforcement probes. When both participants correctly identified comprehension drawing in one opportunity, the duration length of the reciprocal reading and independent reading were expanded. For example, after Participant A and B both emitted the correct paragraph and page number to the comprehension drawing task with the reading duration of seven minutes, the reading duration was increased to nine minutes. All duration intervals were increased by two minutes regardless of the dyads initial mean rate. Criterion for exiting the procedure was set at 100% accuracy across three ascending duration intervals. See Figure 12 for details.
| Step 1 | Reciprocal Reading | - Pairs jointly read the novel  
- Rotation between paragraphs  
- Intermittent vocal reinforcement for peers |
|--------|-------------------|---------------------------------------------------------------------|
| Step 2 | Selection of Reinforcing Textual Stimuli Task | - Pairs select:  
- 2 novel textual stimuli that was difficult to derive the meaning of  
- 1 textual stimulus that was reinforcing  
- Both participants record and share the stimuli  
- Reinforcement for completion of task |
| Step 3 | Independent Reading | - Pairs independently read the novel  
- Highlight preferred paragraphs for reference |
| Step 4 | Comprehension Drawing Task | - Pairs independently draw a preferred scene  
- Pairs switch drawings and respond with the paragraph and page number of the picture.  
- Reinforcement if both are correct. Recycle if either are incorrect. |

**Figure 12.** Peer-yoked Contingency Procedure. This figure provides additional information on each of the procedural steps and examples of materials. The first column labels the step completed, the second describes the components required in each step, and the final column displays a visual of example of the procedural steps.
Peer yoked-contingency procedure. Each participant dyad selected one book that both students were to read. Once agreed upon, the dyad sat together and completed all 4 steps of the procedure. For Step 1, reciprocal reading occurred as the dyad jointly read the novel, each taking turns reading out loud, rotating between paragraphs outside in the hallway. After the dyad finished reading the assigned duration, each participant selected a preferred textual stimulus that was read during the reciprocal reading. Both copied the stimulus in the provided graphic organizer that contained three columns: 1) textual stimulus, 2) page number the stimulus was found on, and 3) an opportunity for the peer to derive the meaning of the word. The participants recorded their textual stimulus and page number, but left the final section blank. The dyad rotated stimuli and were directed to write possible meanings of the text. Reinforcement was delivered on the completion of the task regardless of correct responding.

Following Step 1 and Step 2, the experimenter directed the dyad to independently read for the assigned duration back inside of the classroom. Participants were provided with highlighter and were directed to annotate preferred paragraphs. The experimenter provided both participants with a second graphic organizer that was comprised of a comprehension drawing task. The organizer contained a large blank box where the participants were given the written antecedent to, “Draw a picture of your favorite scene,” in order to establish a possible conditioned seeing response. The participants then switched the comprehension drawings and were required to find the page and paragraph number in the text that their peer had drawn. The participant and peer were presented with an opportunity to respond with written feedback. If both identified the correct page and paragraph number, the dyad received reinforcement. While the dyad had an opportunity to select any shared reinforcer, in this study participants selected “free time” as their primary reinforcer. If a participant did not correctly respond to the peer’s drawing,
participants were directed to complete a correction procedure. The correction procedure required that the participant who emitted the incorrect response was provided with the correct paragraph and page number; the participant was then asked to draw his or her own version of the scene that was identified by the peer. See Appendix D and E for exemplars of correct and incorrect responding.

**Design**

A delayed multiple probe design was used to test the effects of the peer-yoked contingency procedure on the reinforcement value of reading, and if the increase in independent reading duration functioned to increase a participant’s reading assessment scores. Four types of probes were completed for each participant; the first was two 20-min observational probes followed by the WJRB, the GSRT, and the relational analogies probe. After completion of all probe trials, participants were then placed into the peer-yoked contingency procedure, where both participants were required to accurately respond to one another with 100% accuracy across three increasing duration intervals. Once the Participant A and Participant B completed the intervention, the experimenter conducted post-conditioned reinforcement for reading probes; if the participants read for the criterion duration of 16 out of 20 mins, the experimenter provided the participants with the post reading assessments. If participants did not demonstrate increases in reading duration, they returned to the intervention with an increased duration interval. Following the completion of post probes, the second set of pre-intervention probes for Participant C and D were completed. The sequence repeated until all participants were through the intervention. See Figure 13 for the experimental sequence.
Figure 13. Experimental Sequence for Experiment 3.
**Inter-observer and Inter-scorer Agreement**

Like Experiments 1 and 2, fidelity of treatment differed for each variable. The inter-observer agreement (IOA) scores were calculated by dividing the total numbers of paired observer agreements by the total numbers of agreements plus disagreements, and then multiplying that number by 100%. Four components were monitored with IOA: the 20-min observations of reading, the WJRB Passage Comprehension subtest, the WJRB Vocabulary subtest, and the intervention sessions. IOA was obtained for a total of 60% of 20-min observations of reading with mean agreement of 98.5%. IOA was obtained for a total of 75% of WJRB Passage Comprehension subtest with mean agreement of 100%. IOA was also obtained for a total of 75% of the WJRB Vocabulary subtest with a mean agreement of 100%. Finally, 66% of intervention sessions were calculated for IOA for Participants A and B with a mean score of 100%; Participants C and D were calculated at 55% of the sessions with 100% IOA. Inter-scorer agreement (ISA) was used for the relational analogies probe and Grays Silent Reading Tests. ISA was calculated for 80% of the relational analogies probe with a mean agreement of 100%. ISA was calculated for 75% of the Grays Silent Reading Tests with a mean agreement of 100%.
Results

Dependent Variable: 20-min Observations of Reading Behaviors

Figure 14 showed the number of 10 whole intervals of observable reading behaviors for Participants A, B, C, and D. Experimenters observed each participant for a total of 1 hour and 40 mins across pre and post intervention probes. During pre-intervention probes, Participant A was observed to be reading for 41 of the 120 intervals (33%) and 34 of the 120 intervals (28%). After meeting criterion initially for the intervention, Participant A was observed to be reading for 75 of the 120 intervals (63%). Because Participant A had not met the 96 intervals for exiting the intervention, Participant A was placed back into the intervention; after a second round of sessions, Participant A was observed to be reading for 108 intervals (90%) twice. This process was comparable with all participants. Participant B initially was observed to be reading for 23 (19%) and 24 (20%) intervals. Following the first sets of intervention sessions, Participant B was observed to be reading for 70 (58%) of the intervals. Final observations were made after the second set of intervention sessions and Participant B was observed to be reading for 112 (93%) and 110 (92%) of the 120 intervals. Participant C was observed to be reading during pre-intervention probes for 2 (2%) and 18 (15%) of the intervals. Participant D was observed to be reading for 14 (12%) and 10 (8%) of the intervals. After completing the first set of intervention, Participants C and D were observed to be reading for 91 (76%) and 87 (73%) respectively. Post intervention observations found that Participants C and D had both demonstrated additional increases with observed reading intervals. See Figure 15 for results represented with pie charts of the total duration in minutes for each participant.
Figure 14. 20-min Observations of Reading Behaviors. The total possible intervals for all four participants is shown with white bars. Observed intervals are shown with the black bar.
Figure 15. Pie Graph Representation of Durations Pre to Post Intervention. The total possible 20 minutes is shown with the white circle. Gray coloring indicates the total number of minutes and seconds observed. The duration is labeled in each graphic.

**Dependent Variable: WJRB Passage Comprehension Subtest**

Figure 16 and 17 shows the results from the WJRB Passage Comprehension subtest with grade level equivalencies. Participant A emitted 28 correct responses to the pre-intervention probes and emitted 30 correct responses to the post-intervention probes, which was estimated for an increase between grade equivalencies by 0.8. Participant B emitted 25 correct responses to the pre-intervention probes and emitted 30 correct responses to the post-intervention probes with an estimated increase of 1.6 grade levels. In the second dyad, Participant C emitted 32 correct responses to the pre-intervention probes and emitted 35 correct responses to the post-intervention probes with an estimated increase of 1.2 grade levels. Participant D emitted 31 correct responses
to the pre-intervention probes and emitted 36 correct responses to the post-intervention probes with an estimated increase of 1.3 grade levels.
Figure 16. The Grade Level Equivalencies to the Woodcock Johnson III Passage Reading Subtest Pre and Post Intervention.
Figure 17. Increases in Grade Level Equivalencies to the Woodcock Johnson III Passage Reading Subtest Pre and Post Intervention.

Dependent Variable: WJRB Vocabulary Subtest

For the purposes of visual examination of subtests, Figure 18 shows the results from the WJRB Vocabulary Subtest pre and post intervention with grade level equivalencies. During pre-intervention probes, Participant A emitted 11 correct responses to Test 4A Synonyms, 13 correct responses to Test 4B Antonyms, and 5 correct responses to Test 4C Analogies with an overall grade equivalency of 5.3. Post intervention scores demonstrated increases in Test 4A (15), Test 4B (13), and Test 4C (11) with a grade equivalency of 7. Participant B initially emitted 11 correct responses to Test 4A, 12 correct responses to Test 4B, and 9 correct response to Test 4C.
with a grade level equivalency of 6.4. Post intervention scores showed the Participant B increased her grade level equivalency to 7.5. Participant C emitted 11 correct responses to Test 4A, 11 correct responses to Test 4B, and 11 correct responses to Test 4C during pre-intervention probes with a grade level equivalency of 6.7. Post intervention probes indicated that Participant C increased her grade level equivalency to 8.2 with increases in Test 4A and 4B. Participant D also emitted increased numbers of correct responding with gains of 5 in Test 4A, 2 in Test 4B, and 1 in Test 4C with a total increase of one grade level. See Figure 19 for increases in grade level equivalencies.
Figure 18. The Number of Correct Vocal Responses to the Woodcock Johnson Vocabulary Subtest with Grade Level Equivalency Labels.
Figure 19. Increases in Grade Level Equivalencies to the Woodcock Johnson III Reading Vocabulary Subtest Pre and Post Intervention.

**Dependent Variable: Grays Silent Reading Tests**

Figure 20 shows the results from the Grays Silent Reading Tests with grade level equivalencies. Pre-intervention measures were tested with the Grays Silent Reading Form A, while post measures were tested with the Grays Silent Reading Form B in order to reduce exposure to testing materials. Form A and B differed slightly as certain raw scores were correlated with slightly different grade level equivalencies; no raw score varied more than two points and all participants demonstrated gains. Participant A emitted 27 correct responses to Form A and 35 correct responses to Form B with an increase of 3 grade levels. Participant B emitted 24 correct responses to Form A and 27 correct responses to Form B with an increase of 0.7 grade levels. Participants C and D emitted increased numbers of correct responding by 9 and
7 points respectively. Participant C’s final grade equivalency increased by 2.3 grade levels and Participant D score was correlated with a grade equivalency of 6.2. See Figure 21 for increases in grade level equivalencies.
Figure 20. The Grade Level Equivalencies to the Grays Silent Reading Tests.
Figure 21. Increases in Grade Level Equivalencies to Grays Silent Reading Tests Pre and Post Intervention.

**Dependent Variable: Relational Analogies Probe**

Unlike the other dependent variables, there were minimal increases in correct responding. This difference will be discussed later in the analysis. Participant A emitted 11 correct responses to both the pre-intervention and post intervention probe. Participant B emitted one additional correct response in the post probe. Participants C and D each increased their responding by 2 additional correct responses from 7-9 and 10-12 respectively. Table 12 for additional information on correct responding across assessments.
Comparison of Overt and Covert Reading Repertoires

Figure 22 and 23 illustrate the differences in grade level equivalencies across reading assessments for each participant. Figure 22 displays comparison graphs pre and post intervention for the overt reading comprehension measured by the WJRB Passage Comprehension and the covert reading comprehension by the Grays Silent Reading Tests. During pre-intervention probes, Participants A and B had similar grade level equivalencies across overt and covert reading comprehension; however, while Participant B made incremental gains in both reading repertoires, Participant A had a significant gain in covert reading comprehension with an increase of 3.8 grade levels. Participants C and D also similar gains in both overt and covert reading comprehension. In comparison, Participant A, C, and D all increased their overt reading grade level equivalences between .8 and 1.5 grade levels. Each of these participants increased
their covert reading comprehension at higher proportion with grade level equivalences increases of 2.3-3.8.

Figure 23 shows a comparison between the WJRB Vocabulary Subtest with the relational analogies probe. Because of the differences in scoring, there are two axes: the left axis is the grade level equivalency from the WJRB Vocabulary and the right axis is the number of correct response to the relational analogies probe. The comparison displays that although there increases in grade equivalencies in overt responses such as synonyms, antonyms, and analogies, there were small changes between pre and post intervention scores for the relational analogies probe with increases between 0-2 correct responses.
Figure 22. A Comparison of Grade Level Equivalencies across Overt and Covert Reading Comprehension. The gray bar is the grade level equivalency for the overt reading assessment, the WCJIII Passage Comprehension Subtest, and the black bar is the covert reading assessment, the Grays Silent Reading Tests.
Figure 23. A Comparison of Overt and Covert Reading Repertoires for Derived Relations.
Intervention Results

Figures 24-27 show the intervention data for the participants. For Participant A and B, the initial intervention was completed in 9 sessions, while their second set of intervention sessions was completed in 3 sessions. The number of pages read between intervals increased steadily as the duration increased, with a final increase of 2 pages in the Session 12. Participants C and D completed the first set of intervention sessions in 5 sessions and completed the second set in 4 sessions. For the first set of sessions, Participant C and D consistently co-read the same number of pages as when they read independently, with an exception during Session 4. Notably, the dyad increased their number of pages independently read during the second set of sessions by a mean of 3.25 pages.
Figure 24. Number of Correct Yoked-Responses to Intervention Sessions for Participant A and B

Figure 25. Number of Pages Read During Intervention Sessions for Participant A and B
Figure 26. Number of Correct Yoked-Responses to Intervention Sessions for Participant C and D

Figure 27. Number of Pages Read During Intervention Sessions for Participant C and D.


**Discussion**

The results from Experiment 3 supported the earlier findings in Experiment 1 and Experiment 2. The peer-yoked contingency procedure was effective at increasing the reinforcement value of reading and the increases in participant independent reading durations were associated with increases in standardized reading assessments. After two sets of intervention sessions, all four participants increased their reading durations and acquired conditioned reinforcement for reading, as defined earlier in the study. The results also indicted that all participants demonstrated increases in both standardized reading tests. However, the relational analogies probe had limited growth. These results were unexpected, but possible explanations would include a key difference between testing stimuli. Overall, participants increased their overt and covert repertoires for comprehension and derivational relations with novel textual stimuli.

During the pre-intervention 20-min observations, each of the four participants emitted limited reading behaviors, where the maximum duration of consecutive intervals across participants was 1:20. This indicated that during the 20 min durations, reading behaviors stopped being emitted after 1:20 as the participants attended to other stimuli in the environment. While each participant redirected back to the text, this indicated that reinforcement value for reading was minimal. After the first set of intervention sessions, Participant A and Participant B emitted a similar increase in reading behaviors with emission levels at approximately 72 intervals. Interestingly, Participants C and D’s first post-intervention probes had a higher increase despite having a lower initial duration for the intervention. Participants C and D also required fewer sessions to meet the exiting criterion than Participants A and B. While instructional history plays a role in these results, it is possible that the number of pages read during the intervention could
also been used as an indicator of later performance; Participant C and D read the same number of
pages independently as they completed in reciprocal reading, whereas Participant A and B’s
number of pages increased during the independent reading portion of the intervention. It is
possible that as Participants A and B read independently, they were not emitting the same
response covertly as they did when the experiment forced an overt response. This may then
explain the variations between participant data in the first 20 min post probes, as Participant A
and B exited the intervention with lower durations of reading as opposed to their participant
peers C and D.

All four participants increased their comprehension grade level equivalency in both the
overt WJRB Passage Comprehension and the covert Grays Silent Reading Tests. WJRB Passage
Comprehension grade level equivalencies increased with a mean of 1.2 grade levels and a range
of 0.8-1.6. More impressively, the covert reading grade level equivalencies measured with the
Grays Silent Reading Tests were calculated to have a mean score of 2.3 grade levels with a range
of 0.7-3.8. These results indicate that the yoked-contingency procedure was more effective at
teaching covert reading comprehension than overt reading comprehension. It is possible that
since the comprehension drawing task required participants to initially read covertly and then
reread the sections covertly in order to correctly respond to their peer, this continual referring to
the text may have been more effective at increasing covert comprehension. The reinforced
comprehension drawing task may have established a type of conditioned seeing response that
may have aided participants in producing a more accurate understanding of the text.

Each participant also demonstrated increases in grade level equivalencies for deriving
relations between novel textual stimuli. The WJRB Vocabulary subtest 4A and 4B measured two
district relations: coordination relations, which was measured with synonyms, and opposition
relations that were measured with antonyms. The third subtest 4C measured vocal analogies that were primarily comprised of relations of coordination, opposition, hierarchical, and comparison. In contrast, the experimental relational analogies probes measured all the possible eight frames. By not directly addressing additional frames in the intervention, it is possible that the participants had limited exposure of additional relations which would account for minimal difference in pre/post responses. Across all participants, the largest increases were demonstrated in the synonyms section with a mean increase of 15.75% and range of 8%-20%. This may have been a function of Step 2 in the intervention, when participants recorded novel stimuli and were required to derive the meaning of the word.
Chapter V

GENERAL DISCUSSION

In three experiments, I tested the effects of conditioned reinforcement for reading on reading comprehension with 5th graders. In Experiment 1, I analyzed the relations between reading behaviors, as measured by 20-min durations, and two reading assessments. The NJASK and the DRA were chosen to accurately depict the current reading assessment practices across the state. By using a pre and post measurement design, the results showed that not only were there statistically significant relations between both reading assessments, but there were also corresponding relationships with participants’ durations of reading. This demonstrated that the durations of 20-minute observations of reading were related to reading assessment scores; in order to measure the effects of this relationship, I defined conditioned reinforcement for reading as emitting reading behaviors for 16 min of the 20-min probe. When analyzed as three categorical groups, those participants who never acquired conditioned reinforcement for reading had the lowest mean scores on both reading assessments, those participants who acquired conditioned reinforcement for reading previously had the highest mean scores, and those participants who acquired conditioned reinforcement for reading over the 2013-2014 school year had higher testing scores than those who did not, but did not surpass those who had acquired it previously.

Experiment 2 sought to ascertain if additional reading repertoires, such as conditioned seeing and derivational responding, also affected reading assessments scores. In order to measure a possible conditioned seeing response and to measure derivational relations, I created a 15-opportunity PowerPoint where participants were provided with a sentence with a novel word.
Participants were directed to read the sentences for 30 s; the slide changed to the antecedent “Draw a picture of the entire sentence” and conditioned seeing responses were recorded. Finally the participants were shown a slide that stated “What does the word BLANK mean?” and were asked to derive a possible synonym for the novel textual stimulus. Once completed, I compiled each of the 15 responses for the conditioned seeing portion and the 15 responses for the derivation relations. With these data, if a participant emitted 12 or more correct responses to the conditioned seeing probe, he or she was defined as having conditioned seeing in repertoire. Similarly, if a participant emitted 12 or more correct responses to the derivational probe, he or she was defined as having derivational responding in repertoire. Utilizing these three variables, I analyzed the effects of conditioned reinforcement for reading, conditioned seeing, and derivational responding on reading assessment scores. Akin to Experiment 1, conditioned reinforcement for reading was shown to have a statistically significant relationship with reading assessments; this was also true of the repertoires of conditioned seeing and derivational responding. ANOVA analysis also found that with each additional repertoire mean scores on reading assessments increased.

In Experiment 3 I tested the effects of a peer-yoked contingency on reading comprehension as a function of the establishment of conditioned reinforcement for reading. Four participants were divided into dyads based on previous reading assessments such as the DRA and PARCC scores. With the intention of measuring the previously described reading repertories, there were four dependent variables. The Woodcock Johnson “Reading Vocabulary” and a created Relational Analogies probe were designed to measure participants correct responding to deviational relations; The Woodcock Johnson “Passage Comprehension” and the Grays Silent Reading Tests were utilized to measure overt and covert reading comprehension. The peer-yoked
contingency procedure rotated both overt and covert reading, as well as comprehension drawing tasks and derivational opportunities. Results indicated that through a pairing of peer attention and yoked reinforcement, all participants increased correct responding to The Woodcock Johnson “Reading Vocabulary,” Woodcock Johnson “Passage Comprehension,” and the Grays Silent Reading Tests. These results suggest that the yoked contingency procedure acted to enhance repertoires such as reinforcement value of reading, reading fluency, conditioned seeing, and derived responding as well as pairing social collaboration.

**Major Findings**

Results from Experiment 1 and 2 provided a framework for the latter study. In both experiments, conditioned reinforcement for reading was found to be an indicator of reading success in the established assessments for the district and state. The reinforcement value of reading was directly related to increases regardless of when the reinforcer was established, as shown by those participants who acquired conditioned reinforcement for reading over one academic school year. This finding is essential as we discuss how children of different ages may be affected by reading interventions that are similar to Experiment 3.

Results of Experiment 3 provide evidence for three of the suppositions. The first being that when an individual acquires the Naming Joins Textually Responding cusp, in conjunction with speaker-as-own listener and conditioned seeing, the individual will be able comprehend larger segments of verbal behavior. We can see that when the intervention was applied, participants were required to emit overt comprehension drawing tasks. Through the repetition of drawing full paragraphs that were reinforcing, the intervention may have established a conditioned seeing response that was closely aligned to Skinner’s 1957 theory that verbal stimuli can “lead the listener to ‘see’” larger segments of verbal behavior in a type of conditioned seeing.
response (p. 363). The conditioned seeing was strengthened as peers experienced positive attention for their response and this pairing resulted in increases of reading durations even when removed from the peer interaction.

Experiment 3 also provided evidence to support that the covert conditioned seeing responses may be inhibited if the textual stimulus being read is novel. To combat this, readers need to acquire derived relations to create a stimulus equivalency with the novel stimuli; this would then aid the creation of coordination relations that help overall comprehension of a text. Once established, reading can become increasingly more reinforcing as an individual can create stimulus equivalency for relations mediated by joint control and, as such, have a better ability to “see” the presented text. These results may provide additional support to Barnes-Holmes, Barnes-Holmes and McHugh (2004) as they stated that verbal relational skills can be correlated with educational achievement, cognitive skills, and the development of flexibility.

This then bring us to the final assumption: when reading is reinforcing to an individual, he or she may be more likely to have higher reading assessment scores compared to their peers who have not acquired this repertoire. All three experiments provided evidence of this case. Additionally, in Experiment 3, each of the four participants were shown to have increases in both overt and covert comprehension as well as increases in relational frames such as coordination and opposition. It would seem that increases in the reinforcing value for reading can be a significant indicator of later academic success.

**Implications**

If reinforcement value for reading is a significant indicator of academic success, it is possible that conditioned reinforcement for reading may allow students to learn in ways that they could not before; the students’ exposure to novel textual stimuli and complex written texts would
increase as a function of the change in reinforcement value. While this application with comprehension differs from previous research, the overarching principle had been examined extensively. According to Greer and Speckman (2009), verbal developmental cusps and capabilities are behaviors that, when present allow children to learn things and learn in new ways that they could not before. Cusps are verbal milestones that allow children to learn things that they could not learn before. Greer and Ross (2008) defined verbal developmental cusps as cusps that once in a person’s repertoire, allow them to acquire new verbal operants (Greer & Speckman, 2009). Some examples include: preverbal cusps, such as conditioned reinforcement for voices (Greer, Pistoljevic, Cahill, & Du, 2011; Maffei, Singer-Dudek, & Keohone, 2014) and conditioned reinforcement for 2D stimuli (Delgado, Greer, Speckman, & Goswami, 2009) as well as listener/speaker capabilities such as Naming (Gilic; 2005, Greer & Longano, 2010; Stolfi, 2005). If conditioned reinforcement for reading is similar, as it allows students to learn new operants, it is possible that conditioned reinforcement for reading may be a reading cusp as it has been shown to increase reading comprehension through the pairing of reading with social collaboration.

Greer and Du (2015) stated that the attainment of social reinforcers is an essential component for true verbal behavior development. As children acquire language through listener/speaker social contracts, and this communication exchange must be continually reinforced through the verbal episodes (Greer & Du, 2015). It is theorized that different types of reinforcers can control different types of speaker and listener responses; these responses build to create complex verbal behavior and the corresponding conditioned motivating conditions (Greer & Du, 2015). The outcome of this collaboration between the listener/speaker is what emerges when both individuals are motivated to obtain a reinforcer. The joining of motivations yokes the
two individuals together, establishing a collaborate effect that creates verbal behavior exchanges. Behaviorists have defined the collaboration as a yoked-contingency, where the attainment of reinforcement requires interacting by listening and speaking (Greer & Du, 2015). Yoked-contingency procedures have been shown to be effective pedagogical (Greer & Ross, 2008; Broto; 2011) and verbal developmental practices (Stolfi, 2005; Davies-Lackey, 2005; Reilly-Lawson & Walsh, 2007; Rothstein & Geatreaux, 2007; Gold, 2013; Choi & Jung, 2014). With the correct application, yoked contingencies provide social opportunities for a peer to observe reinforcement, which, in turn, increases the opportunity to establish additional conditioned reinforcers.

Social reinforcers also been shown to have effects on the acquisition of conditioned reinforcement for academic repertoires. Lee (2016) tested the effects of a social condition on the direct reinforcement value of writing and indirect conditioned reinforcement for writing; results indicated that participants acquired new reinforcers from the social conditions. As mentioned earlier, Tsai and Greer (2006) and Buttigieg (2015) found that establishing conditioned reinforcement for books with pairings of social reinforcers affected the rate of acquisition of textual responses. Additionally, Weber (2016) and O’Rourke (2006) found that the effects of a social conditions can also affect conditioned reinforcement for math and can establish resistance to extinction for attempting untaught math problems. The results from the current study add to this body of research, as reading comprehension increased as a function of the establishment of conditioned reinforcement for reading through the use of peer collaboration.

Limitations

In all three experiments, there is a limitation of participants. Although this study has been conducted for over four years, the students who have participated are all from the same
classroom model in the same school district with the same experimenter. The study would greatly benefit from replications across different states with different subgroups of participants. Replications could also include other more popular reading assessments used across the country. This would lend to a more complete picture of verbal behavior and its applications to reading.

A second limitation is the use of conditioned seeing as a measurement. Experiment 2 had this difficulty as participants were asked to draw a picture of a sentence they had read. This sentence, however, was not present when the response was emitted. It was intended that without the sentence, the participants’ responses would be a true measurement of conditioned seeing. However, since conditioned seeing is a behavior that occurs beneath the skin, a true measurement was not possible. Participants may have had a covert conditioned seeing response that was not accurately measured in the probe. This limitation was also present in Experiment 3 when participants were asked to draw their favorite scene from the text. Despite the comprehension drawing tasks having inter-observer agreement, and a peer response, it is possible that comprehension drawings were correct for that participant, but were marked incorrectly due to nonconformity to other individuals’ own response.

**Future Research**

Among many avenues for future research, this study could be extended to possibly account for differences between age levels. All participants were between the ages of 10-11 years; it is possible that if the intervention is introduced earlier in their education the intervention may result in larger gains. Furthermore, the study could be extended to evaluate the necessity of other relational frames in context with reading achievement. Is it possible that frames such as spatial and temporal are not relevant to reading repertoires?
**Educational Significance**

Rowman and Littlefield (2014) asserted that current reading education emphasizes word-level skills in literacy such as word recognition, decoding, and spelling, and has little emphasis on text-level skills such as comprehension and writing. While the research behind the effectiveness of word-level skills is numerous (Ehri & Wilce, 1987; Henry, 1993; Perfetti, 2010), it is essential to consider how to teach students the content of a text. Robbins (2011) stated that reading comprehension is generically taught with reading questions that are presented following a prose reading. This approach asks students to “test their understanding” with these prompted questions; Robbins (2011) suggested, instead, that educators must provide varied “active” and “meaningful” responding in order to effectively measure activities that lends its self to learning.

By including a rotation of different response topographies, the results from Experiment 3 supports Robbin’s (2011) suggestion for “meaningful” instruction. While meaningful in this case refers to instruction that is significant and has a purpose, it also can be applied to the quality of the experience for the student. Students cannot draw a meaningful relationship with academic subjects, such as reading, if the instruction does not provide opportunities to actively engage with the material. Taking this thought a step farther, a student cannot find a meaningful relationship with a subject if they are unable to access the “meaningful” activities. As Skinner (1974) stated, The meaning of a response …is to be found in its antecedent history. In other words, meaning is not properly regarded as property of either response or a situation but rather the contingencies responsible for both the topography of the behavior and the control exerted by the stimuli. (p. 93-94).

When students have difficulties due to their instructional history, the task of learning “meaning” becomes overly demanding and can decrease motivation. By teaching students to enjoy reading
more by establishing repertoires such as conditioned seeing and derived responding, students may then find the larger task of reading increasingly more complex novels reinforcing.

**Conclusion**

Norris (2013) stated that reading is an “impressive human achievement” that requires coordination of mastery in a constellation of perceptual and cognitive processes. She maintained this constellation gathers together an assortment of repertoires ranging from visual perception to recognition of word forms, phonological processing, eye-movement control, and all of the other higher-level linguistic processes required to recover an understanding of written words (Norris, 2013). While this is a beautiful notion, reading repertoires must be considered more concretely. They are not a coordination; they are condensation; a condensation of previously mastered repertoires becoming increasingly more fluent. We need to continue to establish opportunities for individuals to learn new reading repertoires, apply these repertoires to increasingly more difficult novels, and maintain the sort of wonder Norris is referring to.

The current study provides evidence that increasing the reinforcement value of reading can affect comprehension scores that are frequently measured in reading assessments. The study also provides evidence that expanding verbal behavior by including additional components to sentences, whether previously acquired or derived, to available verbal operants, may increase comprehension and establish a type of conditioned seeing response. Although these results have been found to be significant in this study, it is essential that the discussion of effective reading repertories continues.
REFERENCES


Progress in International Reading Literacy Study (PIRLS) (2011) Retrieved from https://nces.ed.gov/surveys/pirls/


## Appendix A

### 20-min Reading Observation Data Sheet

#### Whole Interval R+ Reading

<table>
<thead>
<tr>
<th>Student:</th>
<th>Date:</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0:10</td>
<td>4:10</td>
<td>8:10</td>
<td>12:10</td>
<td>16:10</td>
<td></td>
</tr>
<tr>
<td>0:20</td>
<td>4:20</td>
<td>8:20</td>
<td>12:20</td>
<td>16:20</td>
<td></td>
</tr>
<tr>
<td>0:30</td>
<td>4:30</td>
<td>8:30</td>
<td>12:30</td>
<td>16:30</td>
<td></td>
</tr>
<tr>
<td>0:40</td>
<td>4:40</td>
<td>8:40</td>
<td>12:40</td>
<td>16:40</td>
<td></td>
</tr>
<tr>
<td>0:50</td>
<td>4:50</td>
<td>8:50</td>
<td>12:50</td>
<td>16:50</td>
<td></td>
</tr>
<tr>
<td>1:00</td>
<td>5:00</td>
<td>9:00</td>
<td>13:00</td>
<td>17:00</td>
<td></td>
</tr>
<tr>
<td>1:10</td>
<td>5:10</td>
<td>9:10</td>
<td>13:10</td>
<td>17:10</td>
<td></td>
</tr>
<tr>
<td>1:20</td>
<td>5:20</td>
<td>9:20</td>
<td>13:20</td>
<td>17:20</td>
<td></td>
</tr>
<tr>
<td>1:30</td>
<td>5:30</td>
<td>9:30</td>
<td>13:30</td>
<td>17:30</td>
<td></td>
</tr>
<tr>
<td>1:40</td>
<td>5:40</td>
<td>9:40</td>
<td>13:40</td>
<td>17:40</td>
<td></td>
</tr>
<tr>
<td>1:50</td>
<td>5:50</td>
<td>9:50</td>
<td>13:50</td>
<td>17:50</td>
<td></td>
</tr>
<tr>
<td>2:00</td>
<td>6:00</td>
<td>10:00</td>
<td>14:00</td>
<td>18:00</td>
<td></td>
</tr>
<tr>
<td>2:10</td>
<td>6:10</td>
<td>10:10</td>
<td>14:10</td>
<td>18:10</td>
<td></td>
</tr>
<tr>
<td>2:20</td>
<td>6:20</td>
<td>10:20</td>
<td>14:20</td>
<td>18:20</td>
<td></td>
</tr>
<tr>
<td>2:30</td>
<td>6:30</td>
<td>10:30</td>
<td>14:30</td>
<td>18:30</td>
<td></td>
</tr>
<tr>
<td>2:40</td>
<td>6:40</td>
<td>10:40</td>
<td>14:40</td>
<td>18:40</td>
<td></td>
</tr>
<tr>
<td>2:50</td>
<td>6:50</td>
<td>10:50</td>
<td>14:50</td>
<td>18:50</td>
<td></td>
</tr>
<tr>
<td>3:00</td>
<td>7:00</td>
<td>11:00</td>
<td>15:00</td>
<td>19:00</td>
<td></td>
</tr>
<tr>
<td>3:10</td>
<td>7:10</td>
<td>11:10</td>
<td>15:10</td>
<td>19:10</td>
<td></td>
</tr>
<tr>
<td>3:30</td>
<td>7:30</td>
<td>11:30</td>
<td>15:30</td>
<td>19:30</td>
<td></td>
</tr>
<tr>
<td>3:40</td>
<td>7:40</td>
<td>11:40</td>
<td>15:40</td>
<td>19:40</td>
<td></td>
</tr>
<tr>
<td>3:50</td>
<td>7:50</td>
<td>11:50</td>
<td>15:50</td>
<td>19:50</td>
<td></td>
</tr>
<tr>
<td>4:00</td>
<td>8:00</td>
<td>12:00</td>
<td>16:00</td>
<td>20:00</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Positive and Negative Exemplars of Conditioned Seeing Responses

Positive Exemplars for Conditioned Seeing:

*The knight swung his bodkin as the villain attacked the princess.*

Negative Exemplars for Conditioned Seeing:

*The knight swung his bodkin as the villain attacked the princess.*
Appendix C

Relational Analogies Probe for Experiment 3

1. Cloud is to ground as
   a. flower is to plant
   b. head is to foot
   c. chair is to table
   d. phone is to ear

2. Rest is to exhaustion as
   a. dog is to bark
   b. water is to thirst
   c. noise is to muffle
   d. tide is to ocean

3. Sick is to healthy as
   a. masked is to clown
   b. written is to author
   c. saddened is to tragedy
   d. jailed is to free

4. Garage is to car as
   a. hammer is to nail
   b. plant is to farm
   c. stable is to horse
   d. dog is to eat

5. Pen is to pencil as
   a. hair is to color
   b. right hand is to left hand
   c. ring is to finger
   d. nose is to toes

6. Week is to year as
   a. pint is to quart
   b. cup is to teaspoon
   c. arch is to horizon
   d. temperature is to thermometer

7. Clog is to shoes as
   a. cap is to hat
   b. sock is to foot
   c. shirt is to pants
   d. glasses are to face

8. Drip is to pour as
   a. sun is to shine
   b. storm is to clouds
   c. bend is to break
   d. planet is to turn

9. Laughter is to joke as
   a. read is to book
   b. question is to answer
   c. cry is to pain
   d. talk is to conversation

10. Cube is to sphere as
    a. line is to angle
    b. area is to perimeter
    c. cylinder is to can
    d. square is circle

11. Sun is to solar as
    a. pond is to marine
    b. moon is to lunar
    c. planet is to soil
    d. tower is to building

12. Raspberry is to strawberry
    a. glove is to mitten
    b. tree is to woods
    c. snow is to mountains
    d. wheel is to tire

13. Basement is to roof as
    a. windows are to doors
    b. books are to pages
    c. roots are to leaves
    d. cars are to roads

14. Breakfast is to dinner as
    a. sunrise is to sunset
    b. clock is to time
    c. wind is to hurricane
    d. song is to music

15. Bird is to migration as
    a. bear is to hibernation
    b. traveler is to location
    c. parrot is to cage
    d. giraffe is to plains
    e. 

16. September is to December as
    a. bone is to arm
    b. ship is to captain
    c. lake is to puddle
    d. middle is to end
Appendix D

Positive Exemplars of Intervention Responses

Positive Exemplar:

**Lexical Neophilia: Opportunities to Derive Relations**

Participant C recorded three textual stimuli. The first two were required to be novel textual stimuli and the third was a textual stimulus that she enjoyed reading (in purple).

Participant D then was asked to look back to the provided page number and derive a possible synonym (in blue).

Participant C read Participant D’s responses and marked a check if she agreed that the derived response could be a synonym for the textual stimulus (in purple).

Positive Exemplar:

**Comprehension Drawing Task**

Participant C chose to draw the following paragraph:

_I spent two glorious days this way, the best since before Mother had died. I saw deer and hares, and once, at twilight, I swear I saw a phoenix rise, trailing in smoke._

Participant D correctly identified this scene and Participant C marked a plus.
Appendix E

Negative Exemplar with a Recycle for Intervention Responses

Negative Exemplar:

Comprehension Drawing Task

Participant B incorrectly responded with the following paragraph:

"Have you ever seen anything as splendid is this?" He passed me the object he'd had in his lap...

Negative Exemplar:

Comprehension Drawing Task Recycle

Participant B was provided with the correct paragraph:

Mandy lifted the hems of her shirt and five petticoats to reveal feet that were no longer than mine...

Participant B then redrew the paragraph for a the correction.