

The Effects of Small Group Vocabulary Instruction on Second Grade Students'
Expressive Vocabularies

Laura Lester Fariss

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Curriculum and Instruction

Heidi Anne E. Mesmer, Chair
Mary A. Barksdale
Jerome A. Niles
Jennifer J. Powell

June 11, 2013
Blacksburg, VA

Keywords: vocabulary, small group, read alouds, oral language, and expressive
vocabulary

Copyright 2013

The Effects of Small Group Vocabulary Instruction on Second Grade Students' Expressive Vocabularies

Laura Lester Fariss

Abstract

The purpose of this study was to investigate the effectiveness of small group vocabulary instruction above and beyond whole group, read aloud vocabulary instruction, on second grade students' expressive vocabularies. This experimental study reflected a between-subjects design as three treatment groups were compared using a pretest, posttest within subjects variable methodology. A small group instructional intervention was administered to a treatment group in addition to the whole group, read-aloud based vocabulary instruction that the alternative treatment group received. Data was collected over an eight week intervention period. Results indicated that small group vocabulary instruction led to greater gains in second grade students' expressive use of target words than did read aloud-based instruction or no instruction (control). Additionally, students who received small group instruction retained more target word knowledge over time than students who did not receive small group instruction. Implications for practice and future research are included.

Dedication

This dissertation is dedicated to and was completed in honor and remembrance of my mother, Lisa Beth Lester.

Acknowledgements

I would like to thank my entire doctoral committee for all of the hours you have spent on my behalf, reading, analyzing, and reflecting on my work. I appreciate all of the support you have each provided me and for your constructive feedback that helped shape my writing and research. It has been a true pleasure to work with each one of you.

Dr. Barksdale, thank you for your ever-upbeat attitude. Your laugh is contagious and has helped me overcome many difficult moments in my work as a PhD student. Dr. Niles, I appreciate your true commitment to my work and consistently positive attitude towards my progress and ability to succeed as a young scholar. Dr. Jones, I appreciate your belief in me as a teacher and scholar. Your feedback regarding the practical applicability of my work in particular has helped steer me in the right direction.

Finally, I would like to acknowledge my adviser, Dr. Heidi Anne Mesmer. For all of the countless hours spent reviewing my work and conferencing with me, I am eternally grateful. You have been more than just an adviser on my work as a student, but have helped me gain perspective on life in academia. I appreciate all of your time and the commitment that you have shown to my growth and development as a student, teacher, scholar and researcher. Your dedication has been truly evident throughout my four years at Virginia Tech and I simply would not have reached this point in my young career without your guidance. You have believed in me every step of the way, even when my own belief wavered. Your attitude has helped me propel forward, to continue learning and seeking the next opportunity. Thank you all from the bottom of my heart.

Table of Contents

Chapter 1 Introduction	1
Purpose	1
Rationale and Background	2
Conceptual Framework	6
Chapter 2 Literature Review	15
Vocabulary Acquisition in the Early Years.....	15
Vocabulary Acquisition in the Elementary School Years.....	17
Influences of Early Vocabulary Knowledge	19
Considerations for Vocabulary Instruction	21
Scope and Sequence	24
Word Selection	25
Effective Features of Vocabulary Instruction	30
Conclusions	38
Chapter 3 Methodology	39
Purpose, Research Questions, and Hypotheses	39
Sampling Procedures	41
Experimental Conditions.....	43
Fidelity of Implementation.....	55
Measures.....	60
Data Collection.....	62
Word Selection	63
Limitations	67
Primary Data Analysis	68
Chapter 4 Results	71
Assumptions	71
Descriptive Statistics	73
3-Way ANCOVA	75
Ancillary Analyses	79
Participant Flow	87
Intervention Fidelity	88

Baseline Data.....	88
Statistics and Data Analysis	89
Adverse Events.....	89
Chapter 5 Discussion	90
The Use of Best Practices During Read Alouds Can Yield Better Than Average Word Gains.....	91
Clear and Cohesive Vocabulary Instruction Requires Consistency, Clustering, and Repetition	95
Expressive and Interactive Activities Lead to Greater Gains in Word Knowledge	102
A “Plain-old” Read Aloud is Only a Place to Start.....	105
Implications For Teachers.....	108
Implications For Future Research	112
Conclusion.....	113
References.....	115
Appendix A: Book List.....	130
Appendix B: Word List by Cluster	132
Appendix C: Fidelity Checklist.....	133
Appendix D: Lesson Plan Template for Small Group Instruction.....	134
Appendix E: Pilot Study of Assessment Measures	135
Appendix F: Expressive Target Word Assessment (TWA)	156
Appendix G: Word List with Biemiller’s <i>Words Worth Teaching</i> Classification.....	178
Appendix H: Standard Scores and Composites of Participants Standardized Assessments	180
Appendix I: Anecdotal Information from Classroom Teachers.....	182

List of Tables and Figures

Figure 1 Tiers of vocabulary words	28
Table 1 Descriptive statistics for pretest, posttest, EVT standard scores, and adjusted posttest means	74
Table 2 Tests of between-subjects effects	76
Table 3 Pairwise Comparisons	77
Table 4 Adjusted Means	78
Table 5 Descriptive Statistics from repeated measures ANCOVA	81
Figure 2 Raw Scores Across Time	81
Table 6 Tests of Within Subjects Effects	83
Table 7 Pairwise Comparisons	85
Figure 3 Adjusted Means of TWA Scores	86
Table A1 Correlation coefficients for summed scores of pretest and posttest	137
Table A2 Frequencies of receptive assessment items	139
Table A3 Student Test Scores on Test 1 and Test 2.....	140
Table A4 Point-biserial values for pretest items	142
Table A5 Point-biserial values for posttest items.....	143
Table A6 Correlation Coefficients for test items	146
Table A7 Frequencies for expressive assessment items.....	148
Table A8 Pre and posttest summed scores by student	149
Table A9 Point-biserial values for pretest items	151
Table A10 Point-biserial values for posttest items.....	152
Table A11 Composite Scores from PPVT and EVT.....	180

Chapter 1: Introduction

Vocabulary instruction is slowly making its way back into the spotlight of literacy research. On the recently released *What's Hot and What's Not Reading Today's Literacy Survey*, vocabulary/word meaning was considered “hot” by more than 50% of the respondents and “should be hot” by 100% of the respondents, indicating a renewed interest in vocabulary (Cassidy & Loveless, 2011). Indeed, vocabulary instruction needs greater attention. Hart and Risley (1995) proposed that although evidence shows that children enter kindergarten with vastly different vocabularies, indicating a need for instruction, eighteen years later, there seems to be very little systematic vocabulary instruction actually taking place in K-2 classrooms.

Purpose

The purpose of the present research was to investigate the effectiveness of small group vocabulary instruction on second grade student's vocabulary knowledge. Research has documented the effectiveness of whole group vocabulary instruction (e.g. Biemiller & Boote, 2006; Coyne et al., 2009); however, few studies have revealed data on the effects and benefits of small group instruction. In the era of ever-increasing state and national academic standards and expectations for student learning, it is pertinent for the development and progression of quality literacy instruction that researchers and teachers alike strive to develop innovative and effective instructional tools and protocols for teaching vocabulary. Based on the proficiency achievement levels set forth by the 2009 National Assessment of Educational Progress Reading Assessment (National Assessment Governing Board, 2007), Foorman and Connor (2011) proposed that the literacy expectations for primary grade students will likely continue to rise, with greater emphasis

on vocabulary instruction. Hence, the purpose of this study was to develop a vocabulary-based instructional protocol that could be implemented in primary grade classrooms and used to enhance the vocabulary acquisition of students. This study sought to answer the research question, “What are the effects on second grader’s expressive vocabulary knowledge, from receiving small group vocabulary instruction above and beyond that of whole group vocabulary instruction embedded within a read aloud?”

Rationale and Background

Although vocabulary instruction has recently received somewhat greater attention than in previous years, this component of literacy instruction still lags behind other components such as shared reading, word recognition, decoding, and comprehension in the amount of research-based focus received. However, recent efforts have been made to overcome this challenge (e.g. Beck and McKeown, 2007; Biemiller & Boote, 2006; Coyne, Simmons, Kame’enui, & Stoolmiller, 2004; Coyne, McCoach, & Kapp, 2007). The researchers cited above have each attempted to evaluate the effectiveness of particular vocabulary teaching methods. For example, Coyne, and colleagues explored the use of explicit (Coyne et al., 2004) and rich (Coyne et al., 2007) vocabulary instruction. Beck and McKeown (2007) expanded upon basic vocabulary instruction by teaching sophisticated target words. Biemiller and Boote (2006) infused extended word explanations into repeated read alouds so that students would have multiple exposures to novel words. The results of each of these studies showed positive results for students’ word learning, contributing to the growing body of research on vocabulary instruction and bolstering the rationale for why vocabulary instruction should be incorporated within daily literacy routines.

Five well-documented notions about vocabulary research support the need for the present study and related vocabulary research. First, many researchers have provided evidence that early vocabulary acquisition and knowledge contribute to reading comprehension by the third grade (e.g. Anderson & Freebody, 1981; Chall, Jacobs, and Baldwin, 1990; Hart & Risley, 1995; National Reading Panel, 2000; Silverman & Crandall, 2010; Stahl & Fairbanks, 1986; Cunningham & Stanovich, 1997). Cunningham and Stanovich (1997) proposed that vocabulary knowledge in the third grade was predictive of future vocabulary knowledge and influenced comprehension through the 11th grade. According to Chall and Jacobs (2003), children at this time are making a shift in their reading development from learning to read to reading to learn. In order to learn from independent reading, words encountered must be understood in a variety of contexts. If children have not acquired a sufficient vocabulary base at this point in their development, their ability to comprehend texts and individual words will be impeded. These children will also likely be unable to make this shift in their development, resulting in what is commonly referred to as the fourth grade slump (Becker, 1977; Chall & Jacobs, 2003).

Biemiller and Slonim (2001) suggested that most individual differences in vocabulary knowledge emerge in the primary grades and that once a child initially experiences difficulty, which typically occurs by the third grade, his or her ability to “catch-up” becomes increasingly difficult. Given this information, it seems essential that educators understand the most effective ways to teach vocabulary in the primary grades in order to prepare students for the literacy expectations of middle and upper elementary grades and beyond.

Second, prominent researchers, such as Anderson and Nagy (1992), Anglin (1993), and Nagy and Anderson (1984) proposed that the estimated number of words children encounter through books and other reading materials throughout their schooling exceeds 100,000 (Baumann, Kame'enui, & Ash, 2003). There is great potential for many of these words to be rare, decreasing the likelihood of being encountered incidentally in oral language. Although incidental word learning accounts for a large portion of language acquisition, children need direct instruction on many word meanings, along with instruction on how to figure out word meanings.

Third, there is a large gap in children's vocabulary knowledge when entering school (Biemiller & Slonim, 2001; Biemiller, 2005). Biemiller (2001; 2005) declared that the differences in word knowledge between students with advanced and below average vocabularies is approximately 4,000 words and these differences become apparent by second grade. For example, in a 2005 study, Biemiller found that children in the lowest quartile of word knowledge acquired 1.5 root word meanings per day, resulting in a total gain of 4,000 words per year, while children in the highest quartile acquired 3 root word meanings per day, resulting in a gain of 8,000 words per year. These differences reflect the result of oral and written language experiences that occur prior to the onset of school and are often correlated with socio-economic status (Becker, 1977; Duncan, Brooks-Gunn, & Klebanov, 1994; Graves, Brunetti, & Slater, 1982; Hart & Risley, 1995; Marzano, Pickering, & Pollock, 2001; McLloyd, 1998; Nagy & Herman, 1987). These early differences in children's language experiences can impact not only word knowledge, but cognitive development, reading competence (Anderson &

Freebody, 1981; Beck, McKeown, & Omanson, 1987; Nagy & Anderson, 1984; Nagy & Herman, 1987), and overall school success (Stanovich, 1986).

Biemiller (2010) suggested that children in primary grades can learn approximately four to five hundred words per year from direct instruction. However, given that students do not learn every single word taught, over a thousand words must be taught each year in order for children to learn four to five hundred (Biemiller, 2010). Hence, many researchers feel that in order to close the gap between children's varying levels of word knowledge, it is imperative that a research-based vocabulary program be developed (Biemiller 2004; Biemiller 2010; Byrnes & Wasik, 2009; Marulis & Neuman, 2010). Given these critical research reports, the present research is designed with the intent of developing a context in which primary-grade students can receive rich instruction of targeted words, in a systematic and explicit manner.

Fourth, this study was designed based on the notion that very little direct vocabulary instruction occurs in elementary classrooms prior to the third grade (Baumann, Kame'enui, & Ash, 2003; Biemiller, 2001; Lester, 2011). Most vocabulary instruction is situated within whole group read alouds. Research has shown that while children with high or above average vocabulary knowledge are better equipped to learn words incidentally through listening to books read aloud, children with low or below average vocabulary knowledge may not be as able to incidentally acquire vocabulary knowledge. This phenomenon recurrently contributes to the wide gap between children of varying vocabularies, placing many children at significant risk of language delay (Coyne et al., 2004; Robbins & Ehri, 1994; Stanovich, 1986, 2000).

Results from a recent survey of teachers' vocabulary practices corroborate the notion that little systematic vocabulary instruction is currently taking place in the primary grades (Lester, 2011). One hundred pre-school, kindergarten, first- and second-grade teachers in rural school districts in the midwest reported teaching vocabulary primarily in the context of a read aloud, in a whole group setting. Teachers reported teaching most words spontaneously as they were encountered in a read aloud text. Reports also alluded that more than 50% of class discussions around words occurred during reading, rather than before or after reading, further emphasizing the spontaneity of vocabulary instruction. Finally, teachers reported selecting few words from non-fiction domains such as science and social studies curriculums, while selecting most words from basal readers or fictional narratives. These data suggest that teachers are simply not aware of systematic approaches to vocabulary instruction or perhaps do not feel adequately prepared to implement research-based vocabulary instruction (Lester, 2011).

Conceptual Framework

The information presented in the Conceptual Framework aims to set the foundation for understanding the basic principles of vocabulary acquisition. For readability purposes, please note that at times in this review, the term vocabulary acquisition may be used interchangeably with the term *word learning*. First, the nature of word learning and basic principles are defined, followed by a discussion of research on incidental versus direct word learning.

The nature and principles of word learning.

Knowing a word is truly a complex skill, requiring an individual to understand not only the word's definition, but what the word represents and in what contexts the word

can be applied (Neuman & Dwyer, 2009; Stahl & Nagy, 2006). Additionally, truly knowing a word involves having the ability to make connections between bodies of old knowledge and gaining new concepts about what the word represents, and to apply this knowledge when encountering words in a variety of contexts. Nagy and Scott (2000) suggest five aspects of the complexity of knowing a word, including: incrementality, multidimensionality, polysemy, interrelatedness, and heterogeneity. Each of the aforementioned aspects will be individually elaborated on, to convey just how intricate and dense word learning can be.

Incrementality.

Incrementality reflects the notion that individuals can “know” words in varying degrees, implying that word learning is continuous and can solidify over time (Nagy & Scott, 2000). This aspect of word learning coincides with the parameters of word learning stages set forth by Dale (1965), who declared a progression of word learning beginning with having never seen a word, having heard the word but not knowing the meaning, recognizing the word in context, and finally, knowing a word well. Beck, McKeown, and Omanson (1987) also developed a continuum of word knowledge ranging from having no knowledge of a word, a general understanding, a narrow and context-bound understanding, generalized receptive knowledge, and rich decontextualized knowledge of a word. The work of these researchers is centered on the idea of incrementality, which defies the proposition that word learning occurs on an all or nothing basis.

Multidimensionality.

The next aspect of word complexity, multidimensionality, specifies that word knowledge involves understanding many qualitatively different forms of words (Nagy & Scott, 2000). For example, some words that are used primarily in oral language may be used very differently in written language, while other words have multiple meanings and can relate to more than one concept. Encountering words in multiple and varied contexts can help children expand their knowledge of basic word meanings to encompass variations such as the spoken, written, grammatical, syntactic and conceptual forms.

Polysemy.

Polysemy, the third aspect of word complexity, reflects the notion that words can have multiple meanings (Nagy & Scott, 2000). Although this theory may seem straightforward, the ability of words to portray different meanings in different contexts creates somewhat of a barrier in using a dictionary to extract word meanings. For example, when typing the word *charter* into Dictionary.com, the search yielded 10 possible definitions. Students would have to have acquired a multidimensional understanding of the word *charter* in order to understand it within the context of a book or conversation. The word *charter* could refer to a formal document or to lease or hire something, such as a boat or bus. Hence, dictionary definitions may prove to be inconsistent when applying word meanings to various contexts, as the more frequently a word is used in a particular language the more likely it is to have multiple meanings or be used figuratively. The result is that individuals are left with the delicate task of learning to infer word meanings from context, rather than relying on the basic or most common meanings. Discussing word meanings in the context of books can give children opportunities to hear and understand words used contextually.

Further, according to 2009 NAEP data, in order to achieve at a proficient or advanced literacy level, children must understand words at an “in-depth” degree, indicated by the ability to understand and apply the meaning of words beyond a basic definition (Foorman & Connor, 2011). For example, regarding the word *swept*, both of the following sentences reflect contextually correct uses of the word. The first reflects the literal meaning and the second, a figurative meaning: “I swept the floor because it was dirty” and “The angry librarian swept the noisy students out of the library quickly.” In the first sentence, the word *swept* was used literally, as someone used a broom to sweep the dirty floor. However, in the second sentence, the word swept meant that students were sent out of the library quickly because they were being noisy. Students who do not understand the polysemy of words may have a difficult time understanding the word *swept* in the second sentence. Allowing children to discuss personal examples and experiences related to word meanings, as well as, discussing word meanings in various contexts when reading aloud can provide exposure to the multiple meanings of words.

Interrelatedness.

The theory of interrelatedness suggests that an individual’s understanding of a word is intertwined with his or her understandings of other words (Nagy & Scott, 2000). Potentially, understanding the interrelatedness between word meanings can contribute to an enhanced ability to make connections between words and be more useful than the representations of word meanings in dictionaries (Nagy & Scott, 2000). For example, if a child has previously encountered the word *tides* in either oral language or text and has developed an understanding of what a *tide* is, he or she would likely be able to

understand a book about how shells and starfish are washed upon the sand when the tide is high. This child would also be able to relate the word *tide* to other words associated with the beach, ocean, sand, etc. However, this concept may be foreign and difficult to understand for a child who has never been to the beach or read about tidal waves. Hence, the theory of interrelatedness implies that exposure to words in a particular context has potential to trigger the knowledge of related words encountered in a different context, contributing to greater overall reading comprehension. One instructional strategy that highlights the interrelatedness of words is clustering words by semantic relatedness. This strategy for teaching words is later described in this literature review.

Heterogeneity.

Finally, Nagy and Scott (2000) identified the theory of heterogeneity in terms of different kinds of words, such as function or content words, which have different dimensions (Nagy & Scott, 2000). Knowing a word depends on the type of word that it is. Children may follow a progression of how well they know a word, similar to the continuums suggested by Dale (1965) or Beck, McKeown, and Omanson (1987), for some types of words, but not for all.

In sum, word learning is not simple; it is complex and can be hindered or influenced by an individual's own experiences and knowledge, as well as the instruction he or she receives (Pearson & Hiebert, 2007). In order to acquire word knowledge that is sufficient enough to enhance comprehension of oral and written language, it seems that the meanings of words and applicability of those words to various contexts must be understood receptively and expressively.

Incidental versus direct word learning.

Researchers have supported the notion that across all ages and stages of literacy development, the majority of word learning occurs incidentally, or naturally, through language experiences (Beck & McKeown, 1991; Nagy & Anderson, 1984; Sternburg 1987). Beck & McKeown (1991) suggested that of the approximate 3,000 words learned on average per year by children in the preschool and primary grades, only a mere 300 words are learned through direct and explicit instruction. Biemiller's (2010) claims of approximately word learning per year, which was discussed earlier in this review, corroborated this estimate. During the preliterate years, from approximately birth through age two, the majority of words learned are acquired through incidental exposure. This form of language acquisition continues until at least and even beyond the onset of the formal schooling. With the onset of formal schooling, word meanings may be additionally acquired through instructional activities and discussions around novel words. A caveat to this broad understanding of incidental word learning is that the way in which words are learned may be at least in part dependent on what children bring to the table, meaning their prior word knowledge, language exposure and literacy experiences. As previously mentioned, children enter school with varying levels and degrees of word knowledge. Children with large vocabularies may therefore be able to quickly acquire new word meanings incidentally, whereas, direct instruction of specific words may be necessary to bolster the word learning for children with smaller vocabularies (Robbins & Ehri, 1994).

Incidental versus direct teaching.

Although a debate about the effectiveness of incidental and direct word learning has existed in vocabulary research for decades (e.g. Biemiller, 2010; Nagy & Scott,

2000), both seem to have a place in children's development and schooling, perhaps in conjunction with one another. Incidental word learning is supported by prominent researchers, who have suggested that the number of words an individual can acquire incidentally far exceeds the number acquired through direct instruction (Nagy & Herman, 1987; Nagy & Scott, 2000). Percentages of 25-50% of annual vocabulary growth are attributable to incidental exposure to novel words through reading experiences (Nagy, Anderson, & Herman, 1987). Undeniable is the notion that incidental word learning does occur and is a primary contributor to vocabulary acquisition. However, there are several drawbacks to relying solely on incidental learning as the main avenue for word learning.

First, Nagy and Scott (2000) reported that even 40 encounters with a word often do not bring children to a ceiling and four instructional encounters are not enough to significantly make a difference in comprehension, supporting the notion that word learning is complex. Second, incidental word learning fosters receptive language acquisition, as children have opportunities to understand words spoken aloud, but not necessarily reproduce the words orally in various contexts. Expressive vocabulary knowledge refers to the ability to reproduce words orally or in writing and within various contexts, which incidental language exposure does not provide. Kamil and Hiebert (2005) suggested that words understood receptively are typically not known as well or as in-depth as words used expressively, through speaking or writing. Expressive language use requires a higher and more in depth level of vocabulary knowledge, which is not supported in incidental word learning. Third, Nagy and Scott (2000), who are proponents for incidental word learning, suggested that children at the greatest risk for language delay frequently have not been provided with rich exposure to a vast number of words.

These researchers proposed that in order to close this gap in children's varying amounts of word knowledge, immersion in written and oral language is necessary. Not plausible is the notion that children at-risk of language delay will receive sufficient exposure to words without providing them some amount of direct vocabulary instruction.

In contrast to the research supporting incidental word learning, Beck and McKeown, with various colleagues (Beck, Perfetti, & McKeown, 1982; McKeown, 1985; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985; McKeown, Beck, & Sandora, 2012) have primarily explored the effects of direct instruction on children's vocabulary acquisition. In a review of 50 studies, the National Reading Panel (NICHD, 2000) found that children acquired novel words best when vocabulary instruction was embedded within meaningful instructional contexts, rather than in isolation.

Related to Nagy and Scott's (2000) theory of incrementality, students may require engagement in both types of vocabulary instruction, direct and incidental, in order to fully understand word meanings in multiple contexts; however, direct vocabulary instruction simply cannot be dismissed. With the ever-increasing gap in language opportunities and exposure between students from various socioeconomic groups, it is critical to provide young children with exposure to rare words through direct vocabulary instruction beginning in the primary grades (Biemiller & Boote, 2006; Biemiller & Slonim, 2001; McKeown, Beck & Sandora, 2012; Weizman & Snow, 2001). In sum, delving into the uncharted waters of incorporating systematic vocabulary instruction into primary grade classrooms is essential in enhancing the vocabularies of all children and keeping up the expectations of today's literacy standards.

The purpose of the present study was to understand the added benefits of small group vocabulary instruction on second grade students' expressive vocabulary knowledge, above the benefits of read aloud-based vocabulary instruction. This study was designed to address one central research question:

- What are the effects on second grader's expressive vocabulary knowledge, from receiving small group vocabulary instruction above and beyond that of whole group vocabulary instruction embedded within a read aloud?

Chapter 2: Literature Review

The following review of literature is organized into three sections. The first section addresses vocabulary acquisition and development in a child's toddler years. The second section presents a theoretical basis for word learning. This section encompasses the notion of what it means to know a word and how word learning occurs. The third section is centered on recent research of vocabulary instructional practices. Within this section, research on whole group vocabulary instruction, embedded within a read aloud, as well as small group vocabulary instruction is discussed. Effective features of vocabulary instruction, such as using repeated reading and semantic clustering of words to enhance student learning of novel words, are also explained. The purpose of this review of literature is to provide background on how children acquire words from a young age and what research-based instructional practices are most effective in bolstering vocabulary, particularly expressive vocabulary. Gaps in the research are identified as they relate to the overall goals of the present study, which are centered on identifying and solidifying best practices found to contribute to enhanced vocabulary learning in the primary grades.

Vocabulary Acquisition in the Early Years

Beginning in the early stages of infancy and toddlerhood, children typically acquire vocabulary incidentally through oral language experiences, such as: listening to or participating in verbal conversations where adults and older children name objects or informally define words in various contexts, listening to books read aloud, and watching or listening to media sources. Toddlers, typically considered preliterate or just emerging in their literacy, seem to first acquire words with high affective value, such as names of

familiar individuals and enjoyable activities (Ninio & Snow, 1996). During the preschool years, children begin to acquire words embedded in narrative texts read aloud (Snow & Goldfield, 1983). Children who are emerging in their literacy up until approximately third grade are unlikely to acquire vocabulary through books read independently, given that much primary reading material consists of high-frequency words, decodable text, predictable storylines, and patterns with little sophisticated vocabulary. Beck and McKeown (2001) suggested that although young children's word recognition and decoding skills do not allow for independent reading of sophisticated vocabulary words, through oral language experiences these children have the potential to understand rich language and concepts.

The more word knowledge a child develops, the more efficiently he or she is able to acquire new words. Therefore, as children age and are exposed to language in a variety of contexts, word learning seemingly becomes easier. This ease in word acquisition may stem from the ways in which children understand the paradigmatic relationships among words and develop the ability to quickly fill in the gaps with new words when encountered (De Temple & Snow, 2003). This notion is related to Carey's (1978) hypothesis of word mapping. Carey's (1978) hypothesis suggested that children "map" new words in their mind when initially encountering them and over time, as more encounters are experienced, develop deeper and more nuanced understandings. For example, when a child has mapped basic color words, such as: *red*, *blue*, *yellow* and understands the meaning of the word: *color*, it is easier to acquire knowledge of rare color words such as: *pink* or *fuchsia* (De Temple & Snow, 2003). These are common occurrences for young children, which often happen without knowing that word learning

is taking place. This stage of development is characterized quite differently in terms of how vocabulary is acquired than the way in which vocabulary is acquired in the fully literate years.

Vocabulary Acquisition in the Elementary School Years

Children in the primary grades continue to learn language primarily through exposure to rich oral language experiences and text read aloud. By the middle of elementary school, around grade three, however, children begin to interact more independently with texts, at which time word learning becomes more metalinguistically and metacognitively demanding than when word learning occurred primarily through incidental exposures to words (Nagy & Scott, 2000). Nagy and Scott explained that metalinguistically, children must realize that certain words have more than one meaning, perhaps a literal and an implied meaning and be able to reflect on and manipulate these various meanings. Children must learn to make associations and differentiate between words and word meanings in order to understand sophisticated language and advanced text. Metacognitively, children must become aware of the processes involved with word learning. For example, when encountering an unknown word meaning in text, children must consider ways in which they can figure out the meaning. Using context clues, root words, antonyms and synonyms are possible strategies for figuring out unknown words; however, children have to possess the metacognitive ability to realize that they do not know the word meaning and make decisions concerning the use of strategies that will allow them to understand. At this stage in literacy development, children have to bridge together their knowledge of syntactic and grammatical rules to understand individual word meanings in context and the general message of a written passage as a whole.

Unlike the typical words encountered during independent reading in the primary grades, by the third grade children are likely to encounter novel words that are outside of their current vocabulary levels. At this point in their development, children should have a basis for many word meanings and be able to acquire new word knowledge from texts. The ability to acquire new word meanings from text is a skill necessary for furthering and enhancing word knowledge, as Cunningham and Stanovich (1998) suggested that students can only acquire a limited number of novel words through oral language alone. This notion stems from an exploration of the differences found in word rarity in various contexts. Many rare words, or words used infrequently in everyday oral language, are unknown to children because they have not yet encountered them through oral language experiences. If children are only exposed to language orally, the likelihood that they will acquire rare words is low. For example, per 1,000 words, 30.9 rare words occur in children's books and only 17.3 in conversations by adults with college degrees (Cunningham & Stanovich, 1998).

Children beyond the primary years are more likely to encounter a greater numbers of rare words when reading than when engaging in verbal conversations. Children, therefore, need to be exposed to rare words orally and in text in order to acquire new word meanings and enhance their vocabularies. Supporting this hypothesis, prominent researchers estimated that if students engage in a mere 25 minutes of reading each day and read 200 words per minute for 200 days, a million words could be encountered in a year (Herman et al. 1987; Nagy, Anderson, & Herman, 1987; Nagy & Herman, 1984; Nagy, Herman, & Anderson, 1985a, 1985b). These researchers also pointed out that children can learn approximately one in twenty new words encountered in text; therefore,

if a million words are potentially being encountered each year, there are considerable opportunities for acquiring new words through independent reading. A high volume of reading is simply the key to novel word exposure, as the more students read the more likely they are to encounter novel words repeatedly and solidify their understandings of word meanings. However, this form of word acquisition has to be accompanied by some direct instruction in order for children to learn how to acquire novel words from text.

Influences of Early Vocabulary Knowledge

From infancy through second grade, children's ability to rapidly absorb language is similar to that of a sponge soaking up water. Literacy and language learning in the toddler and preschool years occur in the contexts of a child's environment and can be conceptually understood through Bronfenbrenner's (1979) ecological theory. The message suggested in this theory is that learning develops within a variety of interrelated contexts, such as one's home, child care setting, neighborhood and community (Weigel, Martin & Bennett, 2005). There is reason to believe that children exposed to rich language in these early years of life will absorb, or soak up, thousands of novel words.

One of the most commonly cited research reports of influences on vocabulary development is that of Hart and Risley (1995). Hart and Risley found that children from high socioeconomic status received three times more experience to language than children from lower socioeconomic status. Additionally, Hart and Risley estimated that by age four children from families on welfare had received exposure to thirteen million fewer words than children from working class families. Possible explanations for this discrepancy are related to the consequences of low socioeconomic status, such as: less education received by mothers, greater prevalence of single parent families, fewer

resources, and greater survival demands for parents (Qi, Kaiser, Milan, & Hancock, 2006).

McKeown, Beck, and Sandora (2012) reported findings from several recent research reports on the importance of vocabulary exposure at home. For example, more words and longer utterances produced by mothers led to faster recognition of orally spoken words by 24 month old children (Marchman and Fernald, 2008). Other researchers found the complexity of sentences spoken by parents to be related to children's usage of complex sentences (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). Hence, the amount of language children absorb is dependent, at least in part, on the type of language exposure received in their familiar surroundings (Biemiller & Slonim 2001). Furthermore, these experiences at home in the early years of development contribute to later vocabulary knowledge.

Unfortunately, not every child has equal exposure to rich language and literacy experiences before attending school. Research findings have shown that when parents supply their children with high-quality literacy materials (e.g. books), engage in shared book reading, personally value literacy, provide rich oral language experiences, and support their children when reading and speaking, emergent literacy development and motivation for reading is enhanced (Weigel, Martin & Bennett, 2005). Children need more than general exposure to words. Upon the entrance into school, children need explicit and systematic vocabulary instruction in order to obtain sufficient word knowledge.

Considerations for Vocabulary Instruction

Given the background provided on the influences of children's word development prior to formal schooling, selecting a method for vocabulary instruction is a delicate and intricate task for teachers. Researchers have portrayed a long standing debate on the benefits of two instructional approaches to teaching vocabulary: breadth versus depth, finding that although different, both are beneficial and can result in overall word learning (Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009; Biemiller & Boote, 2006). The general difference between these two types of vocabulary instruction is that the depth approach fosters acquisition of fewer word meanings, but more nuanced understandings of individual words, while the breadth approach fosters acquisition of a greater number of words, but more basic or general understandings of individual words.

Teaching for breadth.

Biemiller and Slonim (2001) estimated that by second grade, children considered to have above average vocabularies know roughly 4,000 more words than children considered to have below average vocabularies. Anderson and Nagy (1992) suggested that it is impossible to directly teach all the words needed to close this wide gap in children's vocabulary knowledge, resulting in instruction focused on breadth. Teaching for breadth often reflects an embedded approach to vocabulary instruction in which novel words are discussed throughout the day in a variety of contexts, but not within a formal setting devoted specifically to vocabulary instruction. Although teaching for breadth is at times necessary, as instructional time does not allow for explicit teaching of every word in the English language, children have been found to learn more words when explicit teaching is providing for individual words.

Biemiller and Boote (2006) revealed findings from two studies concentrated on the breadth of vocabulary instruction. Specifically, the researchers explored the effects of repeated readings of storybooks with instructed and non-instructed target words on children's word learning. In the first study, children exhibited a 22% gain in word knowledge for instructed words and a 12% gain for non-instructed or incidentally learned words. In this study, one to four repeated readings were incorporated and students were responsible for providing word explanations. When necessary, the teacher clarified the explanations. In the second study, children exhibited a 41% gain in word knowledge. Again, repeated readings were conducted, but teachers were the sole providers of the word explanations, instead of students, and incorporated two review sessions of the target words. Study two yielded greater word learning than study one, due to teacher-led word explanations and more in-depth review. This study incorporated elements that support research-based vocabulary strategies, such as: rereading books, providing word explanations during reading, and reviewing word meanings. This study reflected the breadth approach of vocabulary instruction as 4-10 word meanings were explained per day. Given the earlier report that children can only acquire roughly three new word meanings per day, Biemiller and Boote's (2006) study, particularly study one, was not centered on the teaching words in-depth, but on exposing children to larger quantities of words each day. Although these data provided evidence that children can learn words incidentally by listening to repeated readings, when greater review and discussion around particular words occurred in study two, children learned more words.

Embedded instruction, such as that used in Biemiller and Boote (2006), allows for discussion around words to be incorporated into meaningful contexts, such as that of a

storybook or content related instruction (National Reading Panel, 2000). The downside to embedded vocabulary instruction is the limited time of focused attention that each word receives (Coyne et al., 2009). Given that children need multiple, varied, and repeated exposures to novel words in order to fully gain nuanced understandings of word meanings, relying solely on an embedded approach to vocabulary instruction may simply not be sufficient.

Teaching for depth.

In contrast to teaching for breadth, teaching for depth typically occurs in the form of extended instruction. One particularly important implication of teaching for depth is the extent to which individual word meanings are known and can be applied when comprehending novel texts (Coyne et al., 2009). Beck and McKeown (2007) explored instruction focused on developing depth of vocabulary knowledge, by providing rich instruction of individual word meanings following the reading of storybooks. In this study, rich instruction consisted of providing explanations of words, reviewing the meanings of words in original and new contexts, and discussing examples of the word meanings. Results of an initial study found that students made significantly greater gains in target word knowledge when they received rich and extended instruction.

An extended approach to vocabulary instruction is beneficial as it provides children with more encounters to novel words, allows for interaction with word meanings in a variety of rich contexts in addition to storybooks (Coyne et al., 2009), and aids in the development of greater word consciousness as children have opportunities to discuss the processes by which they figured out word meanings (Nagy, 2007). An obvious downside to using an extended instructional approach is the time it takes to address individual

words, creating a situation where potentially fewer words are discussed. In summation, researchers such as Coyne et al., (2009) suggested that it may be more beneficial for children to know fewer words well, than to know many words ostensibly. Although storybook reading is an excellent domain for embedded vocabulary instruction, reflective of the breadth approach, rare words or content-specific words may warrant greater instruction and attention, which would be found in a depth-approach to vocabulary instruction.

Scope and Sequence

The Living Word Vocabulary (LWV) is one of the earliest comprehensive assessments of words known by children (Dale & O'Rourke, 1981). Dale and O'Rourke's LWV is comprised of 44,000 words organized by grade level and implies a possible scope and sequence for word learning. Biemiller and Slonim (2001) explored the acquisition of root words by elementary students by adapting Anglin's (1993) lists of roots and Dale and O'Rourke's (1981) LWV to create the Root Word Inventory. Researchers used the Root Word Inventory to assess kindergarten through fifth grade children's expressive and receptive vocabulary knowledge. Results indicated that children learned the words in roughly the same sequence and rapidly developed root word knowledge through the second grade, at which time root word learning slowed significantly, until another peak occurred after sixth grade. Biemiller and Boote (2006) corroborated the results of this study as they also found that students learned words in a similar sequence. The implication of this research is that after grade two, children's acquisition slowed and they began to learn words at more similar rates; therefore, by providing systematic vocabulary instruction in the primary grades, it is possible that by

second grade vocabulary knowledge will be enhanced and increased (Biemiller & Slonim, 2001).

Word Selection

A critical element of vocabulary instruction, as presented by Biemiller and Boote (2006) and Coyne et al. (2004) for instance, is the manner in which words are selected for instruction. Coyne et al. (2004) suggested that although much research highlights findings that showcase effective ways to teach vocabulary, when it comes down to selecting the words for instruction, there is an obvious gap in the literature. One difficulty underlining word selection is that there are thousands of unfamiliar words present in children's reading materials and trade books, often leaving educators bewildered at how and which words to choose, given that there are entirely too many rare words to teach them all (Nagy & Anderson, 1984; Nagy & Scott, 2000). Word selection is further complicated by the varying vocabulary levels of students in any given classroom (Douglas-Hall, Chau & Koball, 2006).

Nagy and Hiebert (2011) recently proposed some beginning steps toward a theory of word selection, indicating several factors that can impact the selection of words for vocabulary instruction. These researchers discussed the dispersion index of words as a possible factor in the selection of words for instructional purposes, arguing that only the most frequently used words in text (such as the first 100 high-frequency words) have the likelihood of being dispersed across a variety of content area texts. However, words outside of that 100 word corpus are essential for developing understandings of specific content area information. For example, Nagy and Hiebert described the word *combustible* as having a low dispersion index and the unlikelihood of its appearance

across a variety of texts. However, there may be a time in children's schooling when teaching the word *combustion* or *combustible* would be critical to understanding particular content. Without direct teaching of this word, it is unlikely that children would ever be exposed to the word in an in-depth manner enough to result in learning of the word. This example reflects a caveat of word selection, in that although many words are rare they are still important and warrant teaching.

In a more practical sense, the question remains: From what corpus of words should teachers select from when choosing words for instruction? Beck, McKeown, and Kucan (2002) have proposed three tiers of words that shed light on the differences in words and help teachers and researchers carefully consider what it is that students can gain from learning words from the various tiers. Tier One words are the most basic, such as: *child*, *sad*, and *book* and typically do not necessitate direct teaching as they will likely be encountered in daily language or beginning reader texts. These are words that are used commonly in everyday language, increasing the likelihood that students will hear and use these words without being directly taught. Sight words, or high frequency words, would be included in Tier One. Tier Two words are still viewed as somewhat high-frequency, but are intended for the comprehension levels of mature readers and therefore, typically require more instruction than Tier One words. Specifically, Beck and McKeown (2007) claimed, "...words for vocabulary instruction should be selected from the portion of word stock that comprises sophisticated words of high utility for mature language users and that are characteristics of written language." (p. 253) For example, the words: *coincidence*, *absurd*, and *industrious* are words that students may hear used in everyday language by adults or older peers and siblings, but for which some direct

teaching is needed in order to clarify the meaning of the word and how exactly to use them independently. Tier Three words are very specific to a particular domain or content area, such as the science-based terms: *proton*, *neutron*, *electron*, and *atom*. Tier Three words are usually taught within the context of a content area, such as science or social studies, as these words are not commonly found in beginning reader texts or in everyday oral conversations. Students are likely to have never heard or used Tier Three words prior to a science lesson centered on the words. Hence, for the purposes of instruction, Tier Two and Tier Three corpuses consist of words that should be taught in order to expand students' vocabularies. Children are likely to encounter Tier Two words during reading and more likely to encounter Tier Three words during specific content instruction. Without instruction of these words, students may never have opportunities to hear or use the words in conversations. This notion has huge implications for when students encounter Tier Two or Tier Three words independently when reading, as there will be increasing difficulty discerning the meaning of the words and understanding the subject matter as a whole. The progression of words from sight words, through each tier of words are presented below in Figure 1.

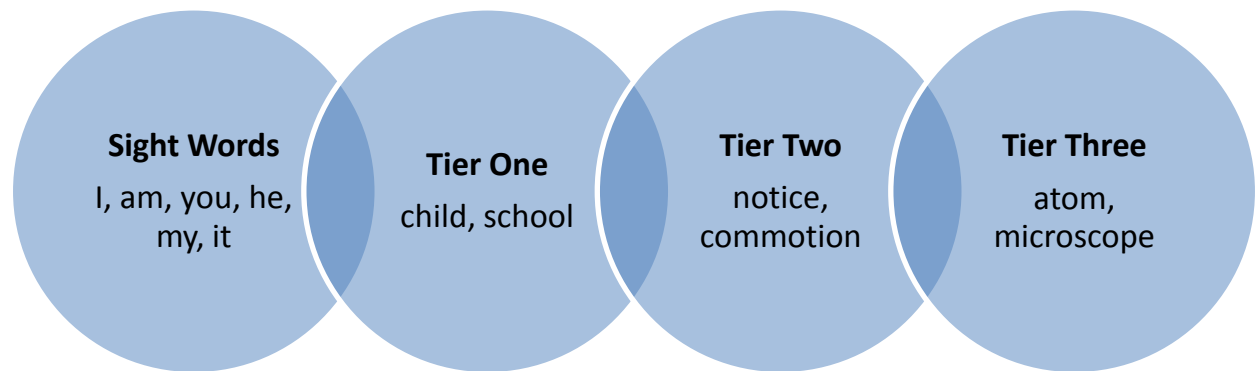


Figure 1. Tiers of vocabulary words. This figure illustrates three tiers of words, including sight words.

Although these tiers of words can be helpful in providing a basis for word selection, in order for vocabulary acquisition to be meaningful for children and contribute to the overall comprehension and understanding of instructional concepts, word selection must not take place without consideration of curriculum.

Nagy and Heibert (2011) explored another dimension of word selection, referred to as a word's role in language. Specifically, these researchers emphasized the considerations of morphological and semantic relatedness between words.

Morphologically, children are typically exposed to words with only one or two morphemes in the primary grades and can typically generalize their understanding of morphemes to inflected forms of words. For example, if children know *happy*, the likelihood of understanding *happiness* or *happily* is great. However, by approximately the third grade children begin to encounter morphologically complex words (Anglin, 1993; Nagy & Anderson, 1984; Nagy & Scott, 2000). Nagy and Heibert (2011) provide

examples of morphologically complex words such as *standardized* and *standardization*. Although these words are derived from the same root word, *standard*, teaching the morphological relationship among the words can be challenging as one word is a noun and the other a verb. Nagy and Hiebert described two additional challenges related to teaching morphological relationships between words. First, some words that are related morphologically may diverge from the root word in spelling and pronunciation, such as *divine* and *divinity* (Nagy & Heibert, 2011). These words have different stress patterns and pronunciations although they are both derived from the same morpheme. Second, the morphological relationships between words are not always semantically related, such as the words *vision* and *visage* (Nagy & Heibert, 2011). Although these words are visually similar and are derived from the same morpheme, knowing the meaning of vision to be eyesight may not assist a child in understanding visage as meaning tolerant. Hence, these challenges can impede vocabulary instruction if words are selected solely based on morphological relationships between words.

Semantic relatedness is an effective means of word selection for students in the primary grades, for several reasons. Nagy and Hiebert (2011) explained that goals of teaching typically involve strengthening students' knowledge of concepts, which inherently involves teaching relationships among concepts. Teachers must be careful when selecting words for instruction as the acquisition of some new words simply involves learning a new label or slightly more nuanced understanding of an already known concept, while other words may involve learning both a new concept and label (Marzano & Marzano, 1988). For example, if a child understands the concept of *cry*, the word *sob* may be easily learned, given that it has similar semantic properties, but a new

label. Learning the word *evaporation*, on the other hand may involve learning an entirely new concept and label. Learning new labels for already known concepts may occur more commonly through incidental exposures in oral language or text; however, learning new concepts and words often involves more concentrated instruction and experiences with the words than simple, incidental exposure. Marzano and Marzano (1988) suggested that when teaching novel, but semantically related words, staggering the introduction of each word within the cluster can help children slowly build an understanding of the relatedness between the words. Also, highlighting one word within a cluster as somewhat of an anchor word can provide children with a basis for the semantics involved in the cluster.

In sum, word selection is very difficult and complex. Children must, however, have opportunities to make connections between word meanings in the early stages of their development. As these understandings of relationships between word meanings begin to solidify, reading comprehension skills have potential to be enhanced.

Effective Features of Vocabulary Instruction

Repeated read alouds.

Reading aloud to children is a common practice that involves a teacher reading a book aloud to her students, as he or she builds background knowledge, addresses novel vocabulary words, and models comprehension, for instance (Beck & McKeown, 2001; van Kleeck, Stahl, & Bauer, 2003). Read-aloud experiences can contribute to enhanced vocabulary development, as there is potential for robust vocabulary learning to arise through the explicit teaching of novel words in context and the exposure of words incidentally (Coyne et al., 2009).

However, a caveat of vocabulary instruction during read-alouds is that not all students have been exposed to rich oral language experiences, such as book reading, prior to attending school and therefore, may not be able to incidentally acquire sufficient words from listening to stories read aloud. Unfortunately, many read aloud approaches lack the intensity needed to bridge this gap (Fien et al., 2011), resulting in a need to extend read-alouds with an instructional component (Biemiller, 2004; Coyne et al., 2007).

One research-based approach to expanding vocabulary exposure during a read aloud is to conduct multiple readings. During repeated readings, the primary form of word learning taking place is incidental, unless the reading coincides with other instructional strategies. For example, Feitelson, Kita, and Goldstein (1986) engaged in six-month interventions where storybooks were read aloud to first grade students and words were directly explained throughout the readings. Repeated readings were incorporated and students exposed to the additional reading made greater gains in language development and reading comprehension than students who were not exposed to the additional readings. More recently, Biemiller and Boote (2006) studied the use of repeated readings in conjunction with vocabulary explanations during read-alouds. Their measures assessed the effectiveness of instruction on vocabulary acquisition, finding that word learning for kindergarten students was positively influenced by the number of repeated readings that took place. Students that listened to four readings, with instruction, made a 22% gain in word learning. The results of this study corroborated the results of 13 other studies reviewed by Biemiller and Boote (e.g. Brabham & Lynch-Brown, 2002; Brett, Rothlein, & Hurley, 1996; Elley, 1989; Hargrave & Senechal, 2000; Robbins & Ehri, 1994; Senechal, 1997; Senechal, Thomas, & Monker, 1995). For

example, several studies found that the average increase of students' knowledge of word meanings from listening to a single reading with word explanations was 15% (Nicholson & Whyte, 1992; Senechal, 1997; Senechal & Cornell, 1993; Stahl, Richek, & Vandevier, 1991), while the average increase of students' knowledge of word meanings from listening to repeated readings without word explanations was 9% and 26% with word explanations (e.g. Brabham & Lynch-Brown, 2002; Brett, Rothlein, & Hurley, 1996; Elley, 1989; Hargrave & Senechal, 2000; Robbins & Ehri, 1994; Senechal, 1997; Senechal, Thomas, & Monker, 1995). Similarly, Penno, Wilkinson, and Moore (2002) compared two conditions of book reading for six year old children. Both conditions involved three repeated readings of a book, but condition two also consisted of teacher-provided explanations of target vocabulary words. Although repeated reading alone has been proved an effective method for teaching words, additional features of the reading experience, such as providing explanations, is most critical in enhancing word learning.

Talk during read alouds.

There is great variability in the types of book reading that teachers engage in with their students, as well as the types of instructional practices and talk that occur before, during, and after reading. Dickinson and Smith (1994) propose that book reading can be categorized into two general styles: co-constructive or interactive versus performance. Characteristic of a co-constructive/interactive style of book reading is the occurrence of rich discussion between teachers and students, before, during and after a story is read. In contrast, rich discussion does not occur during a performance style of book reading, as the intent is for students to enjoy listening to uninterrupted reading by the teacher (Dickinson & Smith, 1994). Although both types of read alouds have a place in primary

grade classrooms, an interactive style of book reading that is analytic in nature, has greater potential to enhance children's vocabulary acquisition as group discussion and reflection on the vocabulary used within the book is an essential element (McKeown & Beck, 2003).

Text talk, a specific program designed to engage students in meaningful decontextualized conversations about words in text, originated from observations of kindergarten and first grade teachers' read-aloud practices (Beck & McKeown, 2001). Researchers found that most discussion taking place during read-alouds consisted of teacher-directed questions that elicited answers of one or two words. The content of the question and answer discussions was reflective of the most basic information in the story, such as "Where did the boy go?" To overcome the question-answer format, the following components of Text Talk were created to enhance the discussion of novel words in text: introducing the story, asking open-ended questions throughout the book, asking follow-up questions, summarizing the story, and engaging students in a variety of activities centered on vocabulary words. Much research has shown that when children become active participants in their own learning and have multiple opportunities to engage in conversations about word meanings, more potential for gains in word knowledge exists (Senechal, Monker, & Thomas, 1995; Justice, Meier, Walpole, 2005; Ard & Beverly, 2004; Blewitt & Walsh, 2006). Finally, Neuman and Dwyer (2009) recommended that word learning involves more than simply labeling and identifying novel words. The talk that occurs during read alouds can and should help students make connections between words and concepts addressed in the text. Children in the primary grades need multiple experiences making meaning of words in context and understanding

the concepts that words represent. Therefore, engaging students in an interactive style of book reading can provide the oral language experiences needed to understand words and make connections between word meanings that will eventually be encountered in written text (Neuman & Dwyer, 2009). Additionally, interactive book reading gives children opportunities to expressively use language.

Small group vocabulary instruction.

Small group vocabulary instruction differs from large group vocabulary instruction in that it has potential to target a more in-depth level of discussion about words, allow students more opportunities to expressively use words, and allow teachers to provide more specific feedback about word use. Although whole-group vocabulary instruction infused into the context of a read aloud has great potential to serve as an effective method of vocabulary instruction, an important critique of the field of literacy research is that few studies have explored the effectiveness of incorporating small group vocabulary instruction into daily literacy blocks.

Several researchers have tested this effectiveness and found positive results in favor of small group instruction above and beyond that of whole group vocabulary instruction. Baker, Fien, and Baker (2010) suggested that small group vocabulary instruction may be an effective method in closing the achievement gap for children at risk of language delay and difficulties in reading comprehension. This may be due to the limitations that whole group setting instruction places on students' abilities to use words expressively. On one hand, whole language instructional settings are ideal for teachers in that they are able to introduce and explain concepts to many children at the same time. However, this setting is less than ideal when aiming to give each child opportunities to

use language independently and with one another. Children at risk for literacy or language delay are often among the population of students that perpetually fall behind throughout their entire schooling. These children often are reluctant to speak out in a whole group setting and when speaking out, often do not receive the level of specific feedback needed to support their learning. Therefore, it seems essential to provide children with opportunities to learn within whole and small group settings.

Fien et al., (2011) investigated the effectiveness of small group vocabulary instruction for students at risk of language delay. The population of students targeted in the study was already receiving whole-group direct and explicit vocabulary instruction through a high-quality read aloud program, prior to the onset of the study. The small group instructional component provided added benefits for children above and beyond the benefits of whole group instruction. Measures of retellings were included to give students opportunities to orally explain words in the context of books read aloud. Results showed that student's progress in expository retellings was enhanced through participation in small group intervention. Hence, although children not at risk for language delay may be able to acquire sufficient word knowledge incidentally and from whole group instruction, the key to unlocking learning for many children, such as those at-risk for language difficulties, strongly points to small group vocabulary instruction.

As mentioned, one reason small group vocabulary instruction has such great potential to enhance word learning for children is that teachers can offer more specific, focused and differentiated activities centered on novel words when working with a small group of children. Perfetti and Hart (2002) have conducted many literacy-based works that have contributed to the development of the lexical quality hypothesis. This

hypothesis maintains that when individuals store high-quality mental representations of word meanings, they are able to rapidly retrieve the meanings when encountering words in context. In order to help children build quality word representations, it is necessary for vocabulary instruction to present words in a variety of contexts, allow for multiple and varied exposures to words, and to foster active processing of word meanings by thinking and discussing ideas around words (Mezynski, 1983; Stahl and Fairbanks, 1986; Perfetti & Hart, 2002; Perfetti, 2007). Although there are benefits to whole group instruction, it is simply not an effective stand-alone method for high-quality vocabulary instruction. Small group settings can offer a safe and secure environment for children to discuss words orally and become active participants in discussions.

Clustering words.

A final effective feature of vocabulary instruction that is critical to discuss in regards to the present study is the clustering of words based on semantic properties. The rationale for this strategy stemmed from the efficiency for which children can store words in memory when learned in relation to others in a common semantic class (Nagy & Hiebert, 2011). Learning words grouped in clusters by semantic similarities should aid the brain in establishing connections and contrasts between words meanings (Beck, McKeown, & McCaslin, 1983; Finkbeiner & Nicol, 2003; Nagy & Hiebert, 2011).

Clustering words can be particularly useful when engaging children in three vocabulary-based instructional activities, including: labeling known concepts with new words, experiencing new concepts and words, and relating new words to known words (Marzano & Marzano, 1988). Although teachers cannot explicitly teach each and every word meaning that occurs in the curriculum, grouping words into clusters for instruction

may be an effective method in which greater numbers of words can be learned, interrelationships between words can be identified, and overall depth of learning can be enhanced. As previously mentioned, Nagy and Hiebert (2011) suggested that a goal of vocabulary instruction, as well as overall classroom instruction, is to help students understand relationships between concepts. Teaching novel words in clusters can therefore, help students capitalize on instruction by not only acquiring novel word meanings but also making associations between words. For example, if children have even some knowledge of at least one word within a cluster, there is potential to build upon that knowledge and acquire new, related word meanings.

However, research on the effectiveness of learning words in clusters versus randomly through rich instruction is mixed (Stahl, Burdge, Machuga, & Stecyk, 1992). This issue may be derived from a common confusion between the effects of teaching words grouped by semantic relatedness versus semantic similarity. Words grouped by semantic relatedness may help students associate words with one another, potentially improving word learning; whereas, words grouped by semantic similarities may put children at risk for confusion and negatively impact word learning (Nagy & Heibert, 2011; Tinkham, 1997). For example, words that are related semantically may include: airplane, pilot, tarmac, runway, and propeller. Words that are semantically similar may include: rat, mouse, chipmunk, and squirrel. These words, in comparison to the semantically related words, are similar in that they are all small, rodent-like animals. The difficulty lies in distinguishing between the slight, nuanced differences in the cluster. With semantically related words, each word has its own distinct meaning, but is related somehow to other words in the cluster. In general, clustering words that are semantically

related can be an effective strategy in teaching words, but may need to coincide with other strategies to significantly enhance overall word learning.

Conclusions

In sum, it is evident that infancy through second grade is a timeframe of rapid vocabulary growth and represents a critical period when children's vocabulary acquisition is most salient. Evident from this review of literature, is the existence of a paradox. Paradoxically, researchers have pinpointed that the greatest need for vocabulary instruction is in the primary grades; however, researchers also have proved that the least amount of systematic vocabulary instruction is actually taking place in the primary grades. This contradiction needs greater attention and consideration for change. All children, especially those children who enter school with below average vocabulary knowledge and those who are at-risk of future language delay and literacy difficulties have potential to benefit from systemic, explicit vocabulary instruction that encompasses the strategies deemed effective by researchers. Based on the research presented in this review, the conclusion can be drawn that systematic vocabulary instruction, which is reflective of research-based best practices, should encompass both read alouds and small group-based instruction where teachers foster discussions around the complexities of word meanings and provide opportunities for students to expressive use words in various contexts. According to Coyne, Capozzoli-Oldham, and Simmons (2012) "early intervention matters, and it matters more for children who enter with less" (p. 52).

Chapter 3: Methodology

The purpose of this chapter is to discuss the methodology used to conduct the present study. Components of this chapter include: purpose of the study, research questions, hypotheses, research design, sampling, experimental interventions, instrumentation, procedures, fidelity, measures, data collection, and word selection.

Purpose, Research Questions, and Hypotheses

As stated in chapter one, the research question that was addressed in this study was: What are the effects on second grader's expressive vocabulary knowledge, from receiving small group vocabulary instruction above and beyond that of whole group vocabulary instruction embedded within a read aloud?

The research hypothesis for the present study was that students in the treatment group, receiving small group and read aloud-based vocabulary instruction, would make greater gains in vocabulary knowledge of target words than students in the read aloud or control groups, who only received read aloud-based whole group instruction or no vocabulary instruction. The null hypothesis (H_0) was that students who received whole and small group vocabulary instruction (independent variable) of clustered target words would show no statistically significant difference in vocabulary knowledge of target words (dependent variable) from students who received only whole group vocabulary instruction or no instruction. Evident in the literature review, was that the field of literacy education has highlighted a large body of research suggesting that direct and systematic vocabulary instruction has potential to bolster primary grade students' vocabulary knowledge (e.g. Baumann & Kame'enui, 2004; Biemiller, 2003; Biemiler,

2004). This study was designed to test the notion that small group vocabulary instruction would be the key ingredient in helping solidifying students' knowledge of target words.

Research Design

In order to analyze the effects of small group vs. read aloud vs. control vocabulary instruction on second graders' expressive vocabularies, a pretest/posttest, experimental design was used, with ANCOVA as the intended data analysis method. Three between subject variables were created with one control group and two treatment groups, selected using stratified random assignment of participants. The between subject variables consisted of three groups, including: read aloud (RA), small group and read aloud (SG) and control (C). Pretesting students using both standardized (EVT-2 and PPVT-IV) and target word assessments (TWA), prior to randomly assigning groups, aided in equalizing the groups statistically, as the PPVT and EVT were used for grouping purposes and the TWA pretest and EVT were used as covariates in the analysis (McMillan, 2008). The research design is illustrated in Table 1 below.

Table 1

Research Design

	Within Subjects Variable		
	Pretest	Posttest	
Treatment Group -RA and SG (read aloud + small group)			Between Subjects Variable
Alternative Treatment -RA (read aloud)			
True Control -C (no treatment)			

Two within subject variables were also used, including pretest and posttest. A main advantage of using within-subject variables was the need for fewer participants, as statistically, power was enhanced by inclusion of the pretest and posttest (Shuttleworth,

2009). The stratified randomly sampling technique for the between subjects variables was another advantage of this study, as it allowed for careful monitoring of the effects on individuals and lessened possible teacher effects, as students from more than one classroom received the treatment. However, Shuttleworth described a disadvantage of a within-subjects design as being the possibility of carry-over effects, meaning the influence of pretest on posttest. Two potential adverse effects included fatigue and practice. These threats were carefully considered and overcome by the duration of the study and limited number of assessments. Specifically, given that only a pretest and posttest were administered, rather than multiple tests throughout the intervention period reduced the number of exposures students had to the test.

Sampling Procedures

Student Participants.

The research question: “What are the effects of small group vocabulary instruction on second graders’ expressive vocabularies, above and beyond that of whole group vocabulary instruction embedded within a read aloud” warranted a sample of second grade students in multiple classrooms. The sample was acquired from one elementary school consisting of three second grade classrooms, comprised of approximately 20 students each. All students had an opportunity to participate, as each student received a consent form, but only those students who returned a consent form were included in the study. Once consent forms were returned, a total sample of 46 students was obtained. Although 46 students returned consent forms, four students were unable to participate in the study because they received Title I services each morning

during the time of intervention, preventing them from participating in the read aloud or small group instruction. This resulted in a final total of 42 student participants.

Demographics of the school were obtained from the National Center for Education Statistics, from the Institute of Educational Sciences (IES). Ethnicity demographics for students at the participating elementary school were as follows: .2% American Indian/Alaskan, .2% Asian/Pacific Islander, 8% Black, .6% Hispanic, and 87% White. The county locale of the elementary school was considered rural fringe and the county population of students was 4,722 (Institute of Education Statistics [IES], n.d.). The specific school that participated in this study was a Title I school, in which 43.7% of the students received either free or reduced price lunch. Total student population was 437, 51% male and 49% female, and 28 teachers.

Teacher Participants.

All three second grade teachers at the targeted school were asked to participate in the study and all agreed. To overcome or lessen potential teacher effects, students that received either small group instruction (SG) or read aloud (RA) were split between two teachers' classrooms. This helped control for potential effects on student's vocabulary learning from the classroom teacher, such as a particular teacher's own ability to effectively teach vocabulary above and beyond that of another teacher's ability. Therefore, after students were pretested, two teachers implemented whole group, read aloud (RA) instruction as an alternative treatment and the researcher administered small group instruction to two groups in each classroom, as SG treatment. The third teacher's classroom served as the control group, receiving neither whole nor small group instruction, beyond the vocabulary instruction already embedded into the second grade

curriculum. This teacher did not play a role in the study other than allowing her students to be assessed for comparison purposes with the treatment group participants.

Experimental Conditions

Read aloud whole group vocabulary instruction.

The rationale for including a read aloud component in this study was based on research depicting whole group vocabulary instruction, embedded into a read aloud, as an effective method for promoting vocabulary development in primary grade classrooms (e.g. Coyne et al., 2009). Reading aloud to students, while embedding extended vocabulary instruction, was deemed a tried and true method of providing a high-quality context for vocabulary learning. Hence, a teacher-led read aloud was incorporated in this study as a component of both the treatment and alternative treatment groups. Students that received the small group instruction also received the read aloud instruction. Books were pre-selected for the read alouds by the researcher and were chosen based on their content and inclusion of target words. A list of book citations can be found in Appendix A.

Repeated Readings. In addition to the discussion accompanying each read-aloud book, which is described below, repeated readings were conducted to provide students with multiple exposures to the target words. Researchers support the usage of repeated readings as an effective instructional strategy for promoting vocabulary acquisition. For example, in numerous studies, primary grade students made greater gains in word knowledge and comprehension when storybooks were read aloud more than once (e.g. Biemiller & Boote, 2006; Feitelson et al., 1986; 1991; Penno, Wilkinson, and Moore, 2002). In order to provide students with researcher-based vocabulary instruction, two repeated readings were conducted during read alouds, in conjunction with extended word

explanations. The first reading was provided with minimal or only naturally occurring interruptions, such as a student asking a question. The second reading involved explicit vocabulary instruction on two target words. A protocol was developed by the researcher and included structured opportunities for teachers to read definitions and examples and students to discuss target words in a systematic fashion. The protocol is presented and described on the following pages.

Number of words taught. Six words were taught each week, with two words taught per day during three instructional sessions per week. There were two weeks when exceptions were made and only two instructional sessions occurred. This change was pre-made, given that the school district's calendar reflected teacher workday/professional development days scheduled on an intervention week. During the course of the intervention, only one school day was missed for snow and this intervention session was made up during the same week, when students returned to school.

The rationale for teaching two words per day stemmed from a line of research suggesting that children can acquire at most, two to three novel words per day. Elley's (1989) study with second grade students, involved rereading books three times and explaining eight or nine words during each reading. Of the eight or nine words taught, students acquired and retained approximately three word meanings per day. Biemiller (2003) also suggested that by the end of elementary school most students have acquired roughly 9,000 words. Hence, in order to acquire this amount of word knowledge, children must learn 2.2 words per day between ages one and eight (Biemiller & Slonim, 2001). Therefore, for the purposes of this study, two words were introduced during each day's read aloud, two or three times per week. Within a week's time, students received

explicit instruction, in a whole-group setting, on either four or six novel words (depending on whether the week included three sessions or two). Words were continually reviewed throughout the week during each read aloud session, as each set of weekly words was semantically clustered and relationships between words were discussed in the books read. A word list can be found in Appendix B.

Whole group instruction. The whole group read aloud session began with the aforementioned repeated readings of the preselected book of the day. First, teachers read aloud the book with minimal interruptions, as the purpose was to expose students incidentally to the words and provide an opportunity for mapping the word when encountered initially in the context of the story (Carey, 1978). After incidentally exposing students to the words, teachers briefly discussed the two words of the day prior to reading the book a second time. The words were written on large index cards and held up for students to see as the teacher introduced the words. During the second reading, teachers stopped when encountering the first target word for the first time and asked students to repeat the word orally. Then the teachers read the researcher-developed definition. Each definition was written on the back of the word cards so that teachers could hold up the word cards for students to see as the definitions were read aloud. Once the definition of the word that was encountered had been read and students had been given a second opportunity to say the word orally, teachers reread the sentence containing the target word and continued reading the book. This was repeated for the second target word of the day. Following the second reading, the two target words of the day were reviewed by the teacher, as the definition of each word was reread. Students were then given the opportunity to discuss the word meanings in a think-pair-share activity. During

a think-pair-share, each student is asked to turn to a partner and share one word and its definition, so that a total of two words are discussed. The think-pair-share experience provided students with opportunities to expressively use and discuss word meanings. The matrix below showcases the read aloud-based extended vocabulary instruction in greater detail. This protocol was used during each whole group, read-aloud instructional session.

Whole Group Instruction	
First Reading	<ul style="list-style-type: none"> • Read through the book without stopping to discuss target words • Respond to student questions and comments as you normally would
Introduction of target words	<ol style="list-style-type: none"> 1. Tell students that you are going to show them two new words. 2. For each card shown, state the word three times: “Today we are going to talk about two new words. The first word is _____. It is written on this card. Raise your hand if you have heard the word _____ before. Please say the word _____ with me.” 3. Repeat step 2 for the second word. 4. Tell students that you are going read the book again and to listen carefully for the words just discussed.
Second Reading	<ol style="list-style-type: none"> 1. Read the book a second time. 2. Stop at the end of the sentence that first contains one of the target words. 3. Ask students to tell a partner the target word. 4. State the target word again for students, while holding up the corresponding index card containing the written word. 5. Ask all students to chorally state the target word with you. 6. Provide students with a definition (developed by the researcher), consisting of four essential components: definition, super ordinate term, examples, non-examples, and description of the word in a picture representation. 7. Do not avoid natural conversation that may evolve when discussing the definitions. After reading a provided definition, you may add to the definition or respond to student questions and comments as you normally would. 8. Ask students to state the word chorally, one more time, for pronunciation purposes.

	9. Reread the sentence containing the target word and continue reading until encountering the next target word for the first time.
After Reading	<ol style="list-style-type: none"> 1. After reading, restate both of the target words while showing the word card and reread each of the two definitions. 2. Think-pair-share: Ask students to turn to a partner and one person share the first word meaning. Then ask the other person to share the second word meaning.

This protocol was developed based on research findings and adapted from Maynard, Pullen, and Coyne (2010). The study conducted by these researchers consisted of a whole group instructional design and yielded significant results, as improvement in student word learning was evident for those receiving “rich” instruction (Maynard, Pullen, and Coyne, 2010). Additionally, Coyne et al., (2010) implemented an intervention where three words were selected during a storybook reading. Each word was introduced by the teacher, pronounced by the children, read in the context of sentences from the book, reread by replacing the word with a synonym, discussed in terms of related pictures, and related to other examples and non-examples.

Efforts were made to establish and maintain consistency between the vocabulary instruction delivered by each classroom teacher. To establish this consistency, teachers completed a training session provided by the researcher. During the training session, the researcher and teachers engaged in a conversation about the vocabulary instruction that was already taking place in their classrooms. Then the researcher provided teachers with an in-depth overview of the study and the research supporting the whole group read aloud protocol. During the second training session, the researcher modeled the protocol to be used for extended vocabulary instruction during a read aloud. The training took place approximately two weeks before the onset of the study so that teachers had ample time to practice the routine and ask questions. According to teachers in the control and treatment

classrooms, a daily read aloud session was already in place in their routine and a general understanding of how to conduct a read aloud was held.

Scripts for the word explanations of each target word were provided in addition to the routine schedule used to implement the whole group instruction. The basis for including definitions, super-ordinate terms, examples, non-examples, and related pictures into the definitions of target words is based on research supporting extended vocabulary instruction during read alouds that engage children in discussion of words beyond simple dictionary definitions (Beck, McKeown, & Omanson, 1987; Kame'enui, Carnine, & Freshi, 1982). Hence, the following examples are researcher-developed word explanations that were used during the whole group instruction:

(DEFINITION) A *sepal* is a part of a flower that covers the bud before it opens. (SUPER ORDINATE) A *sepal* is a part of a flower plant. (EXAMPLE and DESCRIPTION USING PICTURE REPRESENTATION) Here is a picture of a *sepal*. Here is another picture of a *sepal*. See how it is wrapping around the bud to protect it? The *sepal* is usually green. (NON-EXAMPLE) Here is a leaf below the flower (point to the leaf in the picture). That is not a *sepal*. A leaf does not close around the bud or protect it. Remember that the *sepal* is the part that protects the bud.

(DEFINITION) A *summit* is the highest point on a mountain. (SUPER ORDINATE) A *summit* is a part of a mountain. (EXAMPLE and DESCRIPTION USING PICTURE REPRESENTATION) Here is a picture of a *summit*. Here is another picture of a *summit*. See how the *summit* is the highest or tallest part of a mountain? A *summit* is not a small hill or peak. Here is a small hill or peak (point to the hill or peak in the picture). That is not a *summit*. (NON-EXAMPLE) This hill or peak is not the highest point on a mountain. Remember that the *summit* is the highest point on the entire mountain.

A fidelity checklist was also developed based on the routine for whole-group instruction and can be found in Appendix C. The researcher conducted three fidelity checks during weeks two, four, and six by observing each teacher during their read aloud and completing the checklist of required read aloud elements. Results were shared with

teachers after each fidelity check and suggestions were provided as needed, to enhance the implementation of the intervention and maintain consistency between the instruction provided by each of the two classroom teachers.

Small group vocabulary instruction.

Although whole group vocabulary discussions around the various contexts of read alouds are proven to be effective in enhancing word learning, this setting often lacks the intensity and differentiation needed to close the ever-widening gap between students with advanced and at-risk vocabularies (Biemiller, 2004; Coyne, McCoach, & Kapp, 2007). While a smaller body of research is centered on the effects of small group vocabulary instruction compared with whole group vocabulary instruction, results of certain studies have shown that it is an effective means for promoting word learning (e.g. Fien et al., 2011). As a result of these findings, students randomly selected for the treatment condition also received a small group vocabulary intervention in addition to the whole group instruction described above.

The additional small group vocabulary instruction that the treatment group received was designed to foster expressive vocabulary use. Given the nature of small group instruction, the researcher was able to provide students with specific and individualized corrective feedback on their word usage. Maynard, Pullen and Coyne (2010) found similar instruction, which they called “rich instruction” to be highly effective in enhancing the word knowledge of target words for first grade students. Beck et al., (2002) proposed that effective vocabulary interventions should provide definitional and instructional information about words, multiple exposures to words in, and encourage deep processing of words. Given these research-based vocabulary strategies and

recommendations for instruction, small group instruction was implemented into the present study so that students could have multiple exposures to words and opportunities to engage in discussions and activities around the words.

The small group instruction took place after the whole group read aloud sessions three times per week during classroom literacy rotations. The students who were randomly assigned to the SG intervention did not miss out on guided reading or other classroom teacher-led literacy centers, but were pulled out for the small group instruction during independent seat-work times. To maintain the typical schedule of the classroom teachers and to coincide with the typical length of second grade guided reading sessions and literacy centers, each group session lasted 15 minutes.

Grouping for small group intervention. Grouping for the small group intervention was systematic and thoughtful, given the paradox of the known importance of, but lack of, consideration given to students' prior knowledge of word meanings, preceding an intervention (Pollard-Durodola et al., 2011). Pollard et al. (2011) suggested that future researchers attend to students' initial vocabulary knowledge so that intervention instruction can be utilized more efficiently and differentiated more effectively. Hence, purposeful grouping was a goal in this study. Heterogeneous grouping was used in this study so that instruction would be delivered similarly to all groups of students. Homogeneous grouping, although ideal for many grouping systems in classrooms, such as guided reading and math instruction, was not selected for this study because in order to compare small group and read aloud versus read aloud only participants' target word knowledge gains, instruction had to be similar in nature. The benefit evident from the heterogeneous grouping was that students with more

sophisticated vocabulary knowledge could often contribute high-quality and child-friendly definitions and examples of target words during student discussions. Students with lower levels of target word knowledge were able to discuss the words with their peers, whom they felt comfortable with, and learn from one another.

To group systematically, students' standard scores on the PPVT-IV (Dunn & Dunn, 2007) and EVT-2 (Williams, 2007) were averaged to create a composite score. Wise et al., (2007) suggested that when multiple vocabulary assessments are to be administered, many researchers compute composite scores, in which the scores on the two tests are combined into one score, so that general vocabulary knowledge of the participants can be represented. Condouris, Meyer, and Tager-Flusberg (2003) used this approach to obtain an average vocabulary score after administering the PPVT and EVT.

Students were then rank ordered from highest to lowest composite score and randomly assignment to a small group using stratified randomly sampling. In order to group students in a manner that would benefit their learning, four groups were created, consisting of three to four students each.

Small group instruction. First, the researcher introduced and reviewed the two words discussed during the read aloud. Word cards were shown so that students could see the written words. Students had an opportunity to state the definition of the words with a partner, during a think-pair-share. The researcher gave the prompt, "Tell your partner one of the words learned during your read aloud and share the definition. Then your partner will tell you the other word and explain the definition. If one of you is unsure about the definition, ask your partner to help you out. We will discuss the word definitions together after you have a chance to share." Second, the researcher clarified

the word meanings for students by reviewing the definitions used during the whole group session. During the small group instruction, however, the researcher expanded on the definitions by providing examples and personal experiences related to the words. Students were asked to also join in this conversation by recalling connections to their own experiences and stating examples of the words. Third, the researcher asked probing questions so that students had opportunities to expressively use the words and think about and reflect on the relationships between words. Students then engaged in a variety of activities throughout the sessions, such as: differentiating between pictures that represented examples and non-examples of the target words in a picture sort, comparing and contrasting words using Venn diagrams, completing semantic feature analysis charts, completing a T-chart, and matching words with corresponding pictures. Finally, students were asked to write each word of the day and draw a corresponding picture on an index card. The cards were kept in individual envelopes, which students were able to keep at the end of the intervention period. After writing the word and drawing a sketch, students explained their sketches to a partner, which provided yet another opportunity for students to expressively use the target words.

The following matrix contains abbreviated descriptions of activities that were used for the small group instruction. This protocol was used during each small group session in effort to maintain consistency between the sessions. Many activities were adapted from the intervention implemented during a recent study conducted by Maynard, Pullen, and Coyne (2010) as well as from a large number of other research studies cited throughout this document.

Small Group Instruction	
Introduce/Review words	<ul style="list-style-type: none"> • Show the two word cards used during the whole group instruction and state each word one time. • Think-Pair-Share: Each student in the pair should have the opportunity to state the definition of one of the words.
Instruction-teacher led	<ul style="list-style-type: none"> • Provide clarification of word meanings for the entire group using the definitions provided during the whole group session (see whole group matrix for example). • Expand upon the definition by providing personal examples and related experiences for each word.
Instruction/Group discussion	<ul style="list-style-type: none"> • Ask probing questions about the word meanings to extend students' responses (e.g. Have you ever been to a rainforest? How is a rainforest different from a forest we have near our homes? Why? Both forests have trees, how are the trees in the rainforest different?) • Restate, clarify and reinforce student responses • Activities (e.g.): <ul style="list-style-type: none"> ○ Show the pictures of each word. ○ Model and engage students in a picture sort of examples and non-examples of the word ○ Model and collaboratively compare words using a Venn diagram or T-chart
Closure	<ul style="list-style-type: none"> • Students write each of the target words on an index card and draw a related sketch on the back • Students explain their picture and definition to a partner. • Students will keep a binder ring of all the word cards learned during the intervention.

A similar form of this protocol was developed to reflect a more traditional lesson plan format that was used to plan the small group lessons. A copy of the lesson plan form of this protocol can be found in Appendix D.

General study procedures.

The intervention was designed to last for eight weeks. Although, 48 initial words were selected for instruction, pilot data on the reliability of the words highlighted results that led to elimination of six words and replacement of one word. The words were

grouped into eight clusters, one for each week of instruction. Before the intervention period began, IRB approval from the affiliated university was obtained. Approval had previously been obtained from the county superintendent, school principal, and classroom teachers. Informed consent forms were developed and approved as part of the IRB application. Forms were sent home with all students in both the control and treatment classes and parents were asked to sign and return the forms within one week. During this week, the researcher conducted teacher training on the whole group read aloud component of the intervention. The training sessions lasted one hour each and took place after school.

Following return of the consent forms, students were individually assessed using the Peabody Picture Vocabulary Test, PPVT-IV, (Dunn & Dunn 2007) and the Expressive Vocabulary Test, EVT-2 (Williams, 2007). Testing took place during student's typical language arts block and students were assessed individually during their independent center time so as not to miss guided reading or whole group instructional time. Students were administered the PPVT first, followed by the EVT. The purpose for this sequence was that the EVT was a more taxing assessment that could have potentially fatigued students if taken prior to the PPVT. Upon conclusion of administration of the standardized assessments, students were administered an expressive target word assessment (TWA) developed by the researcher. The TWA was administered on a separate day from administration of the standardized assessments. As previously mentioned, after assessments were completed, students were grouped using stratified random sampling.

Students in the control classroom received no instruction on the target words. Students in both of the treatment classrooms received either whole group read aloud instruction or whole group read aloud and small group instruction. A rotation schedule of when each of the four small groups was pulled was created collaboratively by the researcher and classroom teachers.

Given that the school schedule reflected a two week winter break, which fell in the middle of the intervention period, students were reassessed on half of the TWA after the end of four weeks of intervention. The purpose was to overcome the potentially negative effects of a two week break in the middle of the intervention, which could have potentially skewed results since the posttest would have occurred much longer after the first four weeks of intervention than after the second four weeks. To be consistent in the amount of time between intervention and posttest, the final four weeks of intervention were administered after the winter break and followed by the posttest on the second half of the TWA. This schedule allowed for the posttest to be administered one week after each phase, or half, of intervention. The purpose of re-administering the TWA was to determine growth in target word knowledge. Standardized assessments were not re-administered, as the purpose in obtaining the standardized data was simply to establish baselines of students' initial vocabulary knowledge used for grouping prior to the intervention and as a covariate (EVT) during analysis.

Fidelity of Implementation

The matrices of whole and small group instructional activities served as standardized protocols for the teachers and researcher. These matrices were designed by the researcher and used to maintain fidelity during the intervention. The protocols were

developed based on research-based evidence of effective strategies; however, efforts were made to minimize the complexity of the protocols so that teachers would be more likely to implement the instruction with integrity, as it was intended to be delivered. As previously mentioned, fidelity checks were conducted three times during the study, during the second, fourth and sixth weeks of intervention, to ensure that teachers were consistent in the whole group instruction delivered. The researcher created and used a fidelity checklist, reflecting the activities described on the whole group read aloud protocol. A copy of the fidelity checklist can be found in Appendix C. On any occasion that teacher instruction did not match the protocol, teachers met with the researcher to review the protocol and discuss the implementation of changes needed to align instruction with the protocol.

Internal validity.

Examples of threats to internal validity explained in McMillan (2008) included: history, selection, maturation, pretesting, instrumentation, diffusion of treatment, and experimenter effects. Efforts were made to overcome many of these potential threats and strength the internal validity of the study.

First, the threat of history involved the influence of uncontrolled events on the dependent variable. The pretest, posttest experimental design of this study helped to alleviate this threat, as students were tested at the beginning and end of the study, rather than solely at the end of the intervention. The pretest, posttest design also allowed for growth of target word knowledge to be determined, which was a strength of the study. This approach was more effective at overcoming the threat of history than studies which implement interventions and only measure outcomes in target word knowledge.

Additionally, the systematic process used to select words in this study eliminated the likelihood that students received exposure to the words outside of the intervention, during regular classroom instruction by classroom teachers.

Second, the threat of selection, the likelihood of differences between subjects related to the dependent variable, was inherently controlled for in a within-subjects design with stratified random assignment to groups. Stratified random assignment to groups reflected the assumption that the students who participated in each group had attributes that could potentially affect the dependent variable in similar ways (McMillan, 2008).

Third, McMillan (2008) also proposed that a within-subjects design with random assignment naturally controls for the threat of maturation, the changes within subjects over time that may influence results. However, the duration of the intervention and number of target words taught were carefully considered, as the intervention was designed to last eight weeks and target words were carefully selected from outside the second grade curriculum. These considerations were intended to eliminate the chance that students would be taught the target words by the classroom teacher, as part of their regular curriculum, during the intervention timeframe.

Fourth, pretesting was another possible threat that involved administering a measure of the dependent variable prior to the onset of the study (McMillan, 2008). Students were exposed to the words and pictures on the pretest of target words and therefore had potential to learn information related to the pictures and words. However, in this particular study, only an expressive measure containing target word concepts was administered, limiting the likelihood that students would learn words they did not already

know from the pretest. The reason for this notion is based on the fact that the expressive vocabulary assessment required students to state words based on pictorial representations, rather than pointing to pictures of words spoken orally by the researcher. Therefore, students did not have exposure to any words other than those they already knew and could identify. Hence, if a receptive measure had also been administered, students would have had opportunities to learn words spoken by the researcher. Also, McMillan (2008) proposed that this threat is more potent in short studies where only one group is administered a pretest on factual information.

The threat of group rivalry also has potential to affect the methodology and results of this study. McMillan (2008) suggested that pretesting posed the threat of influencing participants to learn information about the test items on their own, by feeling motivated to learn what other participants, in the treatment for instance, were learning. This also had potential to occur as students within alternative treatment groups (only receiving whole group instruction) wanted to learn the information to a greater degree, as are the students in the small groups. The threat of group rivalry was minimized by administering the pretests to students separately and away from the classroom. Small group instruction also took place outside of the classroom so that students who received the alternative treatment of whole group instruction only, were not aware of the instructional activities students in the small group treatment were engaged in. Rivalry also had potential to exist between teachers, as those leading the control or alternative groups could have tried to mimic activities the researcher implemented into the small group treatment sessions. To overcome this threat, the researcher explained to the teachers during training, that their instruction should occur as usual and that nothing new or different should be

implemented because of the study. Also, although the researcher shared generally what the small group instruction would entail, details on the activities were not discussed thoroughly. Teachers were also told that results would be shared with them at the end of the study and analysis period.

To overcome the threat of instrumentation, which refers to changes or unreliability in measurement, a pilot study of both researcher-designed assessments was conducted. The data were analyzed to evaluate the reliability of the overall assessments and individual items. More information regarding specifics of the pilot study can be found in the measures section, as well as in Appendix E. The expressive measure was deemed reliable and was therefore used in the present study.

Lastly, diffusion of treatment, in which a treatment received by one group affects another group, had potential to occur in this study (McMillan, 2008). However, to overcome this threat, the small group instruction (treatment) was administered outside of the participating classrooms. Small groups were held at another location within the school building so that students not receiving the treatment were not exposed to the additional instruction.

External validity.

External validity, the likelihood in which study results could be generalized to other populations, was strengthened by using a between-subjects design with randomized grouping and within subjects variable of pretest/posttest administration. Given that extraneous variables are often controlled in experimental research studies, external validity was more difficult to achieve than internal validity. McMillan (2008) described

five factors that should be considered when generalizing results from research, including: subjects, situation, time, interventions (treatments), and measures.

Characteristics of the subjects, such as socioeconomic status, gender, age, race, and ability (McMillan, 2008) should be carefully considered as well as how subjects were selected for the study. The appropriateness for which group conclusions could be generalized to other populations was carefully considered. The setting in which the educational intervention, for instance, was implemented contributed to the determination of whether or not the results could be generalized to students in other educational settings. The effectiveness found from the results of the intervention also may or may not be able to be generalized based on the time when deemed effective. The intervention treatment was also be carefully conceptualized and administered (McMillan, 2008). Fidelity checks were developed to help maintain consistency within the instruction delivered by classroom teachers. The researcher administered all small group instruction so that the small group instruction that was delivered was consistent across all groups and to lessen the potentially negative effects of having teachers administer their own small group instruction. Finally, the types of measures used have potential to affect external validity; however, the researcher-designed measure that was administered in this study was piloted and deemed reliable. The PPVT and EVT standardized measures were only administered at pretest to determine students' initial vocabulary knowledge.

Measures

Standardized measures.

There were two purposes for administering standardized assessments in this study. The first purpose was to create a composite score of the two measures that was

used to group students using stratified random sampling. The second purpose was to use the standardized expressive vocabulary measure (EVT) as a covariate during analysis. The EVT served as a covariate because the study was centered on gains in expressive word knowledge. Therefore, it was important to assess students' preexisting word knowledge and to control for differences in that knowledge when analyzed the effects of treatment group on gains in word knowledge.

The two standardized assessment measures included the Peabody Picture Vocabulary Test, PPVT-IV, (Dunn & Dunn, 2007) and the EVT-2 (Williams, 2007). The PPVT-IV is a norm-referenced test that serves as a screening for verbal ability and receptive vocabulary knowledge. The assessment was administered to children individually. The format of the PPVT-IV required an administrator to present individual students with a 4-box grid, containing one picture each. The researcher read a prompt that directed students to point to a particular picture. This assessment measured receptive knowledge, as the researcher stated the target words and students differentiated between four pictorial representations of the word by pointing.

The Expressive Vocabulary Test, EVT-2, (Williams, 2007) is a norm-referenced test that was co-normed with the PPVT-IV. This assessment was also administered to students individually. This test required an administrator to ask students to state what they saw in a given picture, following a series of scripted prompts. The EVT-2 measured expressive vocabulary knowledge, as the students were given the responsibility of stating the words.

Researcher-designed measures.

Although standardized assessment measures are critical for comparison purposes in an experimental study, the National Reading Panel (2000) suggested that when assessing specific vocabulary growth, researcher-designed measures are most effective in measuring the gains achieved through instructional interventions. For this reason, researcher-designed receptive and expressive vocabulary measures were developed and piloted prior to the onset of the study. The receptive measure was not deemed reliable to a degree that was preferential for this particular type of research. However, with summed score Pearson's correlation coefficient from pretest to posttest of .912 ($p < .001$), the expressive measure was deemed reliable. Additional information on the pilot study, including analyses of both assessment measures, can be found in Appendix E.

The researcher-designed, expressive target word assessment (TWA) mirrored the EVT-2, as each target word was represented by a single picture. Students were asked what they saw in the picture, eliciting their own knowledge and understanding of the pictorial representation and fostering expressive language use. The researcher-designed TWA was administered as a pretest and posttest. A copy of the TWA can be found in Appendix F.

Data Collection

Data collection for the present study took place during designated weeks for pretests and posttests. All participants were administered the standardized, PPVT-IV and EVT-2, and researcher-designed, TWA, assessments. The researcher assessed all students individually outside of the classrooms, in a quiet location. Standardized measures and the researcher-designed measure were administered on separate days, so as not to fatigue students. Pretest data were collected approximately two weeks prior to the

start of the intervention to allow time for preliminary analysis and grouping of students. All three measures were administered at pretest, including the PPVT-IV, EVT-2, and researcher-designed TWA. As previously discussed, TWA posttest data was collected in two phases, at the end of each four week intervention period. Only the researcher-designed TWA was re-administered at posttest. Each standardized assessment session took approximately 15-20 minutes and each TWA assessment session took approximately 10 minutes.

Word Selection

Many vocabulary researchers and vocabulary assessment developers fail to adequately describe the processes by which words are selected. Pearson, Hiebert, and Kamil (2007) expressed this concern during an analysis of several vocabulary assessments. These researchers implied that rarely are guidelines, frameworks, or theories provided in research reports, which describe the selection of words used for various tests or interventions. Unfortunately, many test developers have selected words based simply on what teaching professionals suggest students in particular grade levels should know, rather than using research to support the selection of words (Pearson, Hiebert, & Kamil, 2007). There are many negative implications of inadequate and non-systematic word selection processes that impede the ability for researchers to generalize the results derived from studies using vocabulary assessments. This missing and essential element of vocabulary research can also hinder understandings of why certain words are included in an intervention, potentially creating disconnect between theory and practice. Given this important critique of much recent vocabulary research, a goal within

this study was to systematically and purposefully select words that were appropriate for a second grade instructional intervention.

The first step in selecting words for the present intervention was to obtain and thoroughly review a list of kindergarten through fifth grade state standards. Once science and social studies words were deemed the foci for this study, those lists of standards were closely examined. To strengthen the internal validity of the study, the majority of words was chosen from third, fourth, and fifth grade standards. However, some words, such as: hibernation and migration, for instance, were words found to be embedded in the curriculum frameworks of multiple grade levels beginning in kindergarten. These words were included in this study and when piloted, found to be reliable words that second-grade students did not reach a ceiling on at the end of the school year. Using second grade words in the study would have increased the difficulty in discerning between the amount of word learning that could be attributed to the intervention as opposed to science and social studies instruction provided by classroom teachers.

Once the state standards had been thoroughly reviewed, the second step in the word selection process involved developing a list of all content area (science and social studies) words. At first the list contained a variety of adjectives, nouns, and verbs related to the content areas. The word list was then compared with lists found in Biemiller's (2010) *Words Worth Teaching* text that is based on Dale and O'Rourke's (1981) *Living Word Vocabulary*. The words found in Biemiller's book were grouped into tiers, similar to the work of Beck, McKeown, and Kucan (2002); however, Biemiller divided primary level words and upper-elementary words into their own sets of three tiers each. A table of the words selected for this study and their corresponding classification from

Biemiller's book can be found in Appendix G. After comparing the initial list based on state standards to Biemiller's lists, it was decided that primarily, nouns would be focus for words included in this study. In regard to Biemiller's list of words, of the 48 total words initially selected, words such as *pollen* (T2) and *prairie* (T6) were among the 50% considered T2 or T6, indicating that they are words advanced children should know at the end of second and sixth grades and words that children at-risk for language delay would likely not know. *Sepal* and *rainforest*, for instance, were words among the 14% not found on any list in Biemiller's text. Words such as, *coast* and *summit* were within the 16% of words on the D list (Difficult Elementary Words) indicating that they were rare words known by fewer than 40% of students at the end of sixth grade. Finally, 18% of the words were found on the E list (Easy Words) which represented words that most children know by the end of second grade. The range in difficulty of words selected for this study likely matched the varying levels of word knowledge of students in any given second grade class. After conducting a pilot study and completing extensive analysis of the reliability of the receptive and expressive vocabulary measures and individual items, it was evident that six of the words should be excluded from the measures: *root*, *desert*, *planet*, *insect*, *blossom* and *antenna*. The words within each semantically-based cluster were carefully considered as the aforementioned six words were removed and the potential for replacement words arose. One word, *desert*, was replaced with the word, *estuary*. After reviewing the state standards for each grade level again, it was decided that if additional semantically related words were included, the likelihood that they would come from K-2 standards was great. As previously discussed, inclusion of words in the K-2 curriculum could potentially skew results, as students may reach a ceiling on pretest

and therefore, not show gains in word knowledge over the course of the intervention or receive instruction on the words in their own classrooms. Therefore, the final number of words was 43, allowing for eight clusters of four to six words each.

The rationale behind selecting words from tiers two and three was to expose students to words that they may not encounter in common oral language. Most of the words selected also coincided with Beck, McKeown, and Kucan's (2002) Tier Three words. Although 18% of the words selected were on Biemiller's (2010) E list (Easy Words), the words still encompassed concepts that were potentially advanced for second graders, determined by their inclusion on state standard lists for third, fourth and fifth grades. Children can and do acquire thousands of words incidentally throughout their lifetime; however, words that are content or domain specific are less likely to be encountered during conversations and warrant teaching. Furthermore, in the era of standardized testing, it is essential that at least some rare and domain specific words be explicitly taught so that students have greater potential for understanding these words when encountered in the context of various reading passages.

Third, words were clustered by semantic relatedness. The rationale behind teaching words by semantic relatedness stemmed from the notion that students need to develop strategies for figuring out word meanings on their own, which can be enhanced by learning to make connections between words. Understanding relationships among and connections between words can lead to more efficient and accurate word retrieval (Marzano & Marzano, 1988). Neuman and Dwyer (2009) proposed that to truly know a word, one must understand what the word represents and how to apply the word in multiple contexts. This hypothesis reflected that once students develop a clear

understanding of how certain words are related, the information can be applied to understanding new words within and outside of the cluster. It was expected that students would know or have at least heard of one word within each cluster and could use the information at hand to understand a network of related words. This strategy, when used appropriately, removes some of the guess-work in understanding word meanings and can potentially serve as a trigger for word meaning recognition when encountering words in various contexts.

Finally, the 48 words, reduced to 43 after analysis of the data obtained in the pilot study, were grouped into eight clusters of four to six words each: water cycle, simple machines, animals and adaptations, earth processes, insects and life cycles, parts of a plants, habitats, and landforms. A list of all target words grouped by cluster can be found in Appendix B. Both appendices contain all 48 original words and words deleted or added. Deleted words are indicated by a slash through the word and added words are indicated by a star.

Limitations

There are several limitations to the present study. First, the dependence on a researcher-designed assessment as the sole means of assessing the intervention was a limitation of this study. Given that a measure of the specific target words used in this study was needed, the only option for an assessment was a researcher-designed measure. However, this measure, the TWA, was piloted prior to the study and deemed reliable after pretest-posttest and item-level analyses. Efforts were also made to enhance the researcher-designed measure by mirroring it after the EVT-2, a standardized, norm-referenced test. A second limitation of the study was the sole administration of the two

standardized measures, EVT-2 and PPVT-IV. Although gains in standardized vocabulary test scores would have been minimal across eight weeks, two administrations would have provided a basis for gains in general vocabulary knowledge to be examined. Third, the researcher-led small group instruction may be viewed as a limitation of the study. Future researchers may find it important to train teachers or teaching assistants to administer the small group intervention, as classroom teachers in this study viewed their inability to administer the small groups as a challenge in implementing similar instruction in the future. However, for the purposes of this study, the small group vocabulary instruction took place during classroom literacy rotations, which prevented classroom teachers from administering the small group sessions due to their commitment to administer guided reading lessons during that time. Finally, a small sample size in a predominantly white, rural school was a limitation that would prevent a broad generalization of the findings. This study will need to be replicated in a more diverse setting with a larger sample size in order to generalize the results large-scale.

Primary Data Analysis

Analysis of covariance (ANCOVA) was selected as the most desirable statistical technique for analyzing the data obtained in the present study as it allowed for mean differences across three samples to be compared while controlling for preexisting knowledge, as well as, the interactions between variables (Howell, 2010). Two covariates were used in this study: the EVT, a standardized assessment measure, and the TWA pretest. Both were administered to participants at the onset of the study. EVT and TWA pretest scores were used as covariates so that participants' preexisting knowledge could be extracted from the analysis, allowing gains in participants' vocabulary

knowledge to be compared evenly, even if some participants began the study with greater word knowledge than others. Furthermore, ANCOVA was an appropriate method of analysis given that random assignment of participants to treatment conditions was used (Green & Salkind, 2011). Important to note is that adjusted means, rather than raw means, in participants' scores were used throughout the following analyses. Adjusted means are raw posttest means that have been altered by taking into account the covariate. Hence, posttest means are adjusted so that differences in scores across groups can be analyzed equally, highlighting differences attributable to treatment condition rather than preexisting knowledge.

First, the assumptions of ANCOVA were tested and met. Second, descriptive statistics were calculated and analyzed to determine original means and standard deviations. Third, the ANCOVA output was examined to determine differences in vocabulary knowledge gained across treatment groups, while controlling for preexisting vocabulary knowledge. Partial eta squared was used as the effect size measurement during analysis. As cited in Olejnik and Algina (2000), according to Cohen (1988) .2, .5, and .8 indicate small, medium, and large effects. These guidelines were used to determine the size of effects during analysis. After examining main effects from the ANCOVA, pairwise comparisons were conducted to determine if statistically significant differences existed between pairs of groups (e.g. small group vs. read aloud; small group vs. control; read aloud vs. control).

During pairwise comparisons, Bonferroni confidence interval adjustment was made to control the overall alpha at .05 and the overall confident interval at .95. This adjustment was critical because when more than three groups are compared, the statistical

equation assumes the cube root of .95, which would have yielded a much smaller alpha. This occurrence would have increased the likelihood of Type I error: the rejection of a true null hypothesis (Howell, 2007). Hence, the Bonferroni adjustment controlled the alpha at .05.

Chapter 4: Results

The purpose of this study was to determine the effects of small group vocabulary instruction on second grade students' expressive vocabularies. A researcher-designed, expressive, target word assessment (TWA) was given to students in both treatment groups and control group at the start of the study. As previously mentioned, the TWA was readministered in two increments, after each four week phase of intervention, or no intervention (C-control). The pre/posttest design was used to determine growth in target word knowledge over the course of the intervention. This information is critical in furthering the body of literature on effective vocabulary instruction, especially in the primary grades. Also, much recent vocabulary research has focused on receptive language, rather than on expressive language and this study measures growth in students' expressive vocabularies over the course of an intervention. To statistically test the hypothesis that small group vocabulary instruction would yield greater gains in second grade students' expressive vocabulary knowledge than read aloud-based vocabulary instruction or no instruction, univariate analysis of covariance was conducted.

Assumptions

Four assumptions underlie ANCOVA. As with ANOVA, assumptions of normality and homogeneity of variance were tested. Additional assumptions specific to ANCOVA included a linear relationship between the covariate and dependent variable and homogeneity of regression, which requires that regression coefficients across treatment groups are equal (Howell, 2007).

First, normality was tested. Skewness and Kurtosis statistics were analyzed and both fell within normal ranges of +/- 1. Second, a Levene statistic was calculated to test

the assumption of homogeneity of variance (homoscedasticity). With an $F = .665$, $p = .520$, Levene's test of equality of error variances was not significantly different, which indicated that the assumption was met and the variances of each population were equal. Third, the assumption of a linear relationship between covariates and the dependent variable were tested. Univariate tests are based on linear independent pairwise comparisons of adjusted means, suggesting that all lines should be parallel. The table of parameter estimates contained t-test scores, which were used to determine the linear relationship. T-test scores for covariates were $p = .000$ for pretest and $p = .001$ for EVT. The significant t-tests indicated a linear relationship with the covariates and posttest.

Finally, the assumption of homogeneity of regression, or homogeneity of slopes was tested. Under this assumption, it was assumed that the population slopes were homogeneous (Green & Salkind, 2011). Three interactions with the two covariates (EVT and pretest scores) were analyzed: treatment group and pretest scores, treatment group and standardized vocabulary scores (EVT), and EVT and pretest. First, with an $F = .474$, $p = .627$, the interaction between pretest and group was not significant, meaning that pretest scores did not vary differently across groups, but were distributed similarly around the mean within each group. Second, with an $F = .194$, $p = .824$, the interaction between EVT and group was not significant, meaning that EVT scores did not vary differently across groups. Third, with an $F = .530$, $p = .471$, the interaction between covariates was also not significant. Obtaining a non-significant F value for each interaction meant that the homogeneity of regression assumption held true, given that the slopes for each group did not differ on the dependent variable (posttest) as a function of the covariate (Green & Salkind, 2011). Simply stated, the slopes in target word

knowledge gained, across groups, from pretest to posttest were similar, even when controlling for entering knowledge. Therefore, both EVT and pretest stood as sufficient covariates to be used in the ANCOVA.

Descriptive Statistics

After ensuring that all assumptions for ANCOVA were met, descriptive statistics were analyzed. First, baseline data were reported. The baseline data highlighted participants' scores from two standardized assessments administered at the beginning of the study, the PPVT-IV and EVT-2. The PPVT-IV measured receptive vocabulary knowledge and the EVT-2 measured expressive vocabulary knowledge. As suggested in Wise et al., (2007), when administering multiple vocabulary assessments, many researchers have developed composite scores, where the scores on the two tests are combined into one score, so that general vocabulary knowledge of the participants can be represented. For example, Condouris, Meyer, & Tager-Flusberg (2003) used the approach of averaging PPVT and EVT scores to compare children's vocabulary knowledge, as indicated by standardized measures of vocabulary and spontaneous speech. A similar format was used in this case to determine overall, general vocabulary knowledge of the participants. For the purpose of the present study, a composite of each participants' standard scores from the two assessments were calculated simply by averaging the standard score of the PPVT and EVT. The central tendency of EVT scores was 100.3. The composite scores were used to randomly assign participants to treatment groups, using the stratified random sampling method. A table of standard PPVT, EVT and composite scores can be found in Appendix H.

Although composite scores of the PPVT and EVT were used in the stratified random sampling procedure, please note that only standard EVT scores was used as a covariate during the remainder of analysis because the study was centered on expressive vocabulary knowledge (measured by the EVT) rather than receptive knowledge (measured by the PPVT).

Descriptive statistics were then calculated to determine original means for pretest and posttest scores by treatment group, as well as standardized EVT scores, which was used as a covariate in the ANCOVA below. Although adjusted means were not produced during descriptive statistics but were later computed, they have also been included in this table for clarification purposes. These means are critical to report, as the adjusted means will be used in the ANCOVA, rather than raw means. Data on original and adjusted means are presented in Table 2.

Table 1

Descriptive statistics for pretest, posttest, EVT standard scores, and adjusted posttest means

Group	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	EVT Mean	EVT SD	Adjusted Means	Std. Error
Control (C)	10.5	4.3	10.42	5.8	100.7	10.7	8.96	1.02
Read Aloud (RA)	6.85	5.08	18.07	8.7	99.71	11.7	19.73	1.03
Small Group (SG)	9.07	4.39	32.14	5.97	100.1	11.4	31.96	.99

The descriptive statistics highlighted that raw mean differences from pretest to posttest were -.08 for control (C), 11.22 for read aloud (RA), and 23.07 for small group (SG).

This indicated that SG participants learned 23 words, on average, between pretest and posttest, while C group participants learned less than 1 word on average and RA participants learned 11 words. Table 3 also showcases that the C group and SG were nearly equal on EVT scores but made distinctly different gains in target word knowledge over time. Additionally, although C group scored slightly higher on pretest than SG, SG scored more than three times as high as control group at posttest. The data also showed that RA group had the lowest pretest and EVT scores at the onset of the study, although EVT scores were within one point of C and SG scores. Therefore, based on the descriptive statistics, the hypothesis that students who received small group instruction would make greater gains in target word knowledge from pretest and posttest, than did students who received read aloud instruction or no instruction, still appeared at this point in analysis, as probable, although not yet statistically confirmed.

3-Way ANCOVA

ANCOVA was conducted to determine if statistically significant differences existed between treatment groups (e.g. C, RA, and SG) from pretest to posttest, after controlling for preexisting vocabulary knowledge. Preexisting knowledge was controlled for by two covariates including pretest and EVT scores, which were deemed reliable covariates during assumption testing.

Between subjects main effects were first examined to determine the effects of treatment group on target word knowledge. These data can be found in Table 3. With an $F(2, 3643.651) = 132.77, p < .0001, \eta^2 = .878$, group was significant, indicating that treatment made a difference in students' target word knowledge on posttest (SG vs. RA vs. C). Moreover, with a large effect size (Cohen, 1988) of .878, 87% of the variability

in test scores could be accounted for by treatment. With an $F(1, 281.869) = 20.54$, $p < .0001$, $\eta^2 = .357$, pretest was significant and a small to medium effect size of 35.7% highlighted the variability in posttest scores from pretest scores. With an $F(1, 181.289) = 13.21$, $p < .01$, $\eta^2 = .263$, EVT was also significant and a small effect size of 26.3% reflected the variability in posttest scores that could be attributed to EVT scores. These data showcased that even after controlling for preexisting vocabulary knowledge, differences in scores across treatment groups remained significant and that treatment (C, RA, and SG) contributed to statistically different gains in participant's target word knowledge from pretest to posttest.

Table 2

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4783.396 ^a	4	1195.849	87.155	.000	.904
Intercept	54.815	1	54.815	3.995	.053	.097
Group	3643.651	2	1821.826	132.777	.000	.878
Pretest	281.869	1	281.869	20.543	.000	.357
StandardEVT	181.289	1	181.289	13.213	.001	.263
Error	507.675	37	13.721			
Total	22453.000	42				
Corrected Total	5291.071	41				

a. R Squared = .904 (Adjusted R Squared = .894)

Once main effects were examined, pairwise comparisons were then calculated to determine significant mean differences between each pair of treatment groups (e.g. control vs. read aloud, read aloud vs. small group, and small group vs. control). All comparisons between treatment groups were significant. First, with a mean difference of 10.77, $p = .000$, there was a statistically significant difference in gains in target word knowledge between C and RA groups. Second, with a mean difference of 23, $p = .000$, there was a statistically significant difference in gains in target word knowledge between C and SG. Third, with a mean difference of 12.22, $p = .000$ there was a statistically significant difference in gains in target word knowledge between RA and SG. The significance of all three comparison groups proved that treatment group did make a difference in gains in target word knowledge, as all groups were significantly different from one another. Specifically, SG participants outperformed the RA and C group participants, and RA participants outperformed the C group participants. These data indicated that at this level of analysis, with a single posttest administered, small group and read aloud-based vocabulary instruction can be deemed effective methods of enhancing target word knowledge, whereas, no instruction cannot be deemed an effective method. However, it is critical to note that although both small group and read aloud seemed to be effective in enhancing TWA scores, small group had a more powerful effect.

Table 3

Pairwise Comparisons

Group	Group	Adjusted Mean Difference	Std. Error	Sig	95% Confidence Interval for Difference ^b
					Lower Bound Upper Bound

Control vs.	read aloud	-10.776	1.517	.000	-14.580	-6.973
	small group	-23.000	1.417	.000	-26.554	-19.446
read aloud vs.	Control	10.776	1.517	.000	6.973	14.580
	small group	-12.224	1.447	.000	-15.851	-8.596
small group vs.	Control	23.000	1.417	.000	19.446	26.554
	read aloud	12.224	1.447	.000	8.596	15.851

The statistically significant differences between each pair of groups were based on adjusted means. The adjusted means were original means that were adjusted during ANCOVA to take into account preexisting vocabulary knowledge, using the covariates pretest and EVT. Adjusted means are unbiased means that allow the outcome variable to be analyzed equally across all groups of participants, by extracting preexisting knowledge from the equation, placing all participants on somewhat of an even playing field. Although reported previously, with the original means in Table 2, adjusted means by group can be found in Table 5 below.

Table 4

Adjusted means

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	8.956 ^a	1.026	6.878	11.033
Read aloud	19.732 ^a	1.039	17.626	21.837
Small group	31.955 ^a	.991	29.947	33.964

Ancillary Analyses

Although not originally planned, two forms of ancillary data were obtained and analyzed. First, a delayed posttest was administered with the intent to analyze data across three time points. The purpose was to determine which treatment group retained the most target word knowledge over time. To measure word retention, a delayed posttest of the TWA was administered one month after the end of the eight-week intervention period and posttest. The purpose of this ancillary analysis was based on the notion that more information was needed to determine not only the immediate effects of vocabulary instruction on a given posttest, but also the lasting effects of instruction (treatment group) on target word knowledge.

The delayed TWA was administered to all students in the control, read aloud and small groups, with the exception of one participant. This participant moved from the participating school between the time of posttest and delayed posttest. This participant's score was included for the ANCOVA analysis, as there were two scores to analyze from both pretest and posttest. However, this participant was not administered a delayed posttest and therefore, all scores for this participant were removed from the data set during the repeated measures ANCOVA. With three time points: pretest, posttest, and delayed posttest, now available for analysis, repeated measures ANCOVA was used to statistically determine which group sustained the greatest amount of word knowledge one month after the close of the intervention.

In preparation for the repeated measures ANCOVA, Bonferroni confidence interval adjustment was once again used to control the overall alpha at .05 and the overall confidence interval at .95. Additionally, as in the primary ANCOVA analysis previously

discussed, Partial eta squared was used as the effective size measurement during analysis, with .2, .5, and .8 indicating small, medium and large effects (Cohen, 1988; Olejnik & Algina, 2000)

In the following sections, the ancillary data to be discussed was based on repeated measures ANCOVA. Assumptions, descriptive statistics, and results from the statistical tests were reported and described.

Assumptions. Tests of assumptions for repeated measures ANCOVA were consistent with the assumptions from ANCOVA, with the addition of sphericity. The assumption of sphericity is that the variances of the differences between all combinations of related groups are equal. This assumption was tested using Mauchly's Test of Sphericity, which produced a Mauchly's W of .966, $p = .566$. Given that the test was not significant, sphericity can be assumed.

Descriptive Statistics. Descriptive statistics were calculated to determine original (raw) means for pretest, posttest, and delayed posttest scores by treatment group, as well as adjusted means for the delayed posttest. Table 6 and Figure 2 below contain data on pretest, posttest, and delayed posttest original means by group, as well as the adjusted means for delayed posttest scores. Although raw scores are important in gaining general knowledge about changes in group means across time, only the adjusted means were entered into the model when analyzing group differences statistically. Adjusted means were therefore, included in the table and were discussed more in-depth during reporting of the repeated measures ANCOVA.

Table 5

Descriptive Statistics from Repeated Measure ANCOVA

Group	Pretest Raw Mean	Posttest Raw Mean	Delayed Posttest Raw Mean	Adjusted Delayed Posttest Mean
Control	10.50	10.43	10.21	10.64
Read Aloud	6.86	18.07	13.50	13.75
Small Group	9.00	31.62	25.46	21.86

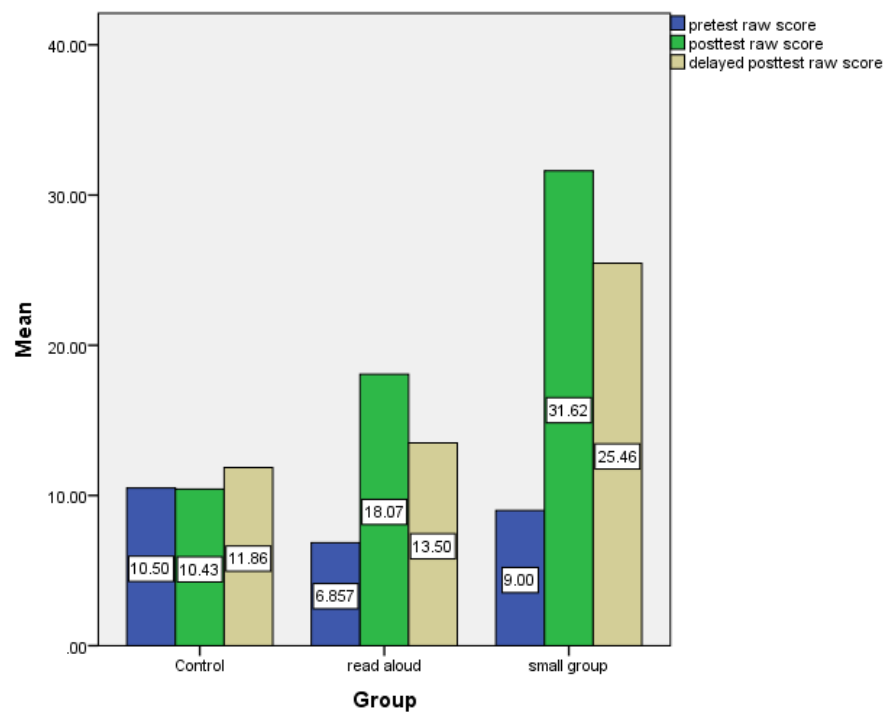


Figure 2. Raw scores across time. This figure illustrates pretest, posttest, and delayed posttest scores by treatment group.

Evident from these data, was that C group had the greatest mean at pretest (10.5) and RA had the smallest mean (7.0). At posttest, SG, which had been in between C and RA at pretest, had the greatest mean (31.6), while the C group dropped slightly, having the smallest mean (10.4). These data corroborate the analysis from the Univariate

ANCOVA, in that SG made the greatest gains from pretest to posttest. When adding the third variable, delayed posttest, SG and RA group scores decreased, while the C group scores increased slightly. The increase in C group scores may simply be a learning effect from general, second grade curriculum instruction over the course of time from pretest to delayed posttest or a learning effect from three administrations of the assessment. This increase cannot be attributed to any instruction during intervention. Descriptively, although SG and RA scores decreased from posttest to delayed posttest, SG participants seemingly retained the most target word knowledge out of all treatment groups.

Repeated Measures ANCOVA. Repeated measures ANCOVA was conducted to determine which treatment group retained the most target words over time. First, within subjects effects were analyzed, followed by pairwise comparisons. Within subjects effects were analyzed to determine the significance of the within subjects factor: time. Three levels of the within subjects factor: pretest, posttest, and delayed posttest, were included in the analysis. Time was first examined, followed by interactions between time and group, time and covariate (EVT) and time, covariate and group. Results showed that with an $F(2, 55.32) = 5.168, p = .008, \text{Power} = .811, \eta^2 = .132$, time was significant, meaning that participants' scores, no matter which treatment group they were assigned to, were statistically different from one another across all three time points. There were also statistically significant two-way interactions for time and group ($F(4, 63.18) = 2.951, p = .026, \text{Power} = .763, \eta^2 = .148$) and time and EVT ($F(2, 160.33) = 14.978, p = .000, \text{Power} = .999, \eta^2 = .306$). These data suggested that mean TWA scores by treatment groups were significantly different across the three time points and that the covariate, EVT, was significantly different across all three time points. However, there

was not a statistically significant three-way interaction for time, EVT, and group ($F(4, 46.30) = 2.163, p = .082, \text{Power} = .609, \eta^2 = .113$). This finding revealed that over the three time points, when controlling for the covariate EVT, groups were not statistically different from one another. Pairwise comparison data provided more details on which specific groups were not statistically different.

Partial eta squared was used as the measure of effect size. For time by group 14.8% of the variance was accounted for by group differences, meaning that treatment group accounted for 14% of the gains in target word knowledge across time. For time by EVT, 30.6% of the total variance in gains in target word knowledge was accounted for by preexisting knowledge indicated by the covariate, EVT. Table 7 below shows the table of within-subject effects reported in this section.

Table 6

Tests of Within Subjects Effects

Source	Sum of Squares	Df	Mean Square	F	Significance Level	Partial Eta Squared	Power
Time	55.329	2	27.665	5.168	.008	.132	.811
Time * Group	63.188	4	15.797	2.951	.026	.148	.763
Time * EVT	160.337	2	80.169	14.978	.000	.306	.999
Time * EVT * group	46.301	4	11.575	2.163	.082	.113	.609

Next, pairwise comparisons were examined to determine mean differences between each pair of treatment groups (e.g. C vs. RA, RA vs. SG, and SG vs. C) across three time points: pretest, posttest, and delayed posttest. As previously mentioned, adjusted means were used in the analysis and reflected means of delayed posttest scores

after controlling for the covariate, EVT. These means were included with the descriptive statistics raw scores in Table 6.

Although there was a statistically significant difference between C and RA groups when analyzing differences in pretest and posttest, when entering the third time factor, delayed posttest, into the model, there was no longer a statistically significant difference between the two groups. From the pairwise comparison table, found in Table 10 below, the mean difference between C and RA groups was 3.104, $p = .091$. This suggested that participation in the C group was no different from participation in the RA group, in terms of words gained and retained across three time points. There was no statistically significant difference in the gains made between these two treatment groups, as there was only a difference of 3 words retained between the RA and C groups. However, there was a statistically significant difference between the C group and SG, with a mean difference of 11.217, $p = .000$. This suggested that participation in the small group instruction resulted in statistically greater gains in target word knowledge than participation in the control group. In this case, there was a difference of approximately 11 words gained between the SG and C group, with SG having retained the greater number of words. Finally, there was also a statistically significant difference between SG and RA group, with a mean difference of 8.113, $p = .000$. This final comparison highlights the finding that there was a difference of 8 words learned and retained between SG and RA group, with SG having retained the greatest number of words. Table 9 below showcases statistical differences in pairwise comparisons.

Table 7

Pairwise Comparisons

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Control	read aloud	-3.104	1.373	.091	-6.561	.354
	small group	-11.217*	1.370	.000	-14.667	-7.767
read aloud	Control	3.104	1.373	.091	-.354	6.561
	small group	-8.113*	1.397	.000	-11.632	-4.594
small group	Control	11.217*	1.370	.000	7.767	14.667
	read aloud	8.113*	1.397	.000	4.594	11.632

*Based on adjusted means

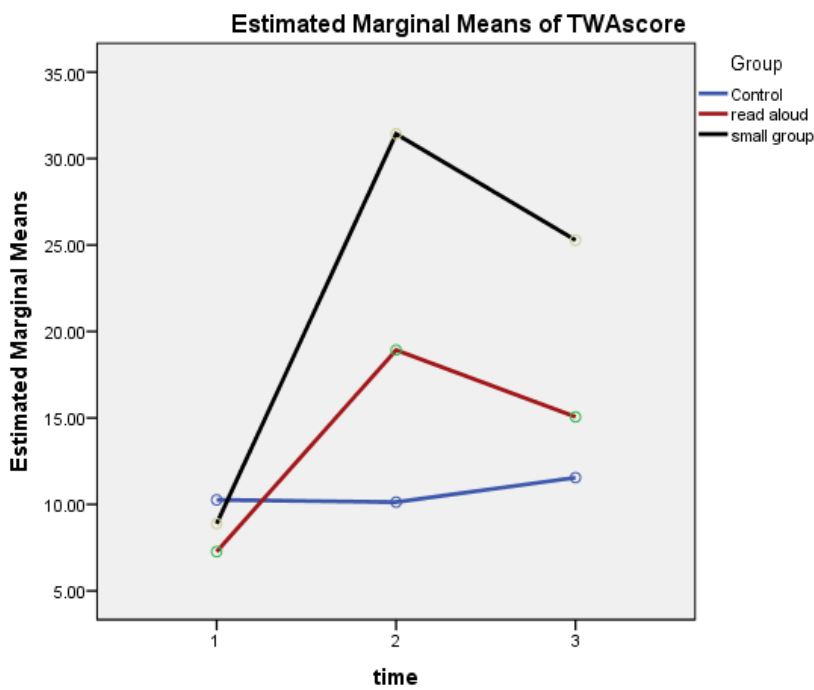
*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Overall, these data indicated that small group vocabulary instruction resulted in higher gains in target word knowledge. Students in the RA group and SG both made gains in target word knowledge from pretest to posttest, but both groups decreased in target word knowledge from posttest to delayed posttest. However, SG participants gained the most word knowledge overall and retained the greatest number of word knowledge over time. SG participants scored on average, 20% on pretest and 59% on delayed posttest. RA participants scored on average, 16% at pretest and 33.2% at delayed posttest. Students in the C group made virtually no gains in target word knowledge across the three time points. Control group participants scored on average, 24.4% on pretest and 27.6% on delayed posttest. Given these data, small group vocabulary was deemed an effective method of teaching target vocabulary words to

second grade students and led to greater retention of words than read aloud-based instruction alone or no specific vocabulary instruction, which the C group received.

Finally, Figure 3 below provides further evidence of the gains made in target word knowledge between groups across three time periods. Important to notice is the difference between slopes of the three groups. The slopes of the RA and SG are both steeper than the slope of the C group; however, it is evident from the figure below that the slope of RA participants' gains is not as steep as those for SG participants. This indicates the heightened word learning that occurred for SG participants. This line graph illustrates the changes in TWA scores by treatment group. Using adjusted means, evident from the graph is that although C group began the study with the highest TWA mean, the group failed to make significant gains at posttest or delayed posttest. The RA and SG both made gains in target word knowledge across the three time points, but SG outperformed all groups.



Covariates appearing in the model are evaluated at the following values: EVT Standard Score = 100.1500

Figure 3. Adjusted Means of TWA scores. This figure illustrates changes in target word knowledge between three treatment groups: control, read aloud, and small group, across three time periods of TWA administration.

Participant Flow

The total number of participants for this study was 41. Initial recruitment was extended to all second grade students at a designated school. The entire second grade consisted of 60 students at the time of recruitment; therefore, 60 consent forms were sent home to students. A total of 46 consent forms were returned, with the control group class having returned 14 forms and the two designated treatment group classes having returned 16 forms each. Four students were excluded from participation, even though consent forms were returned, because of their participation in pull-out Title I instruction. Due to time overlap, Title I instruction at the school prevented these four students from participating in the read aloud condition of the study. Given that the read aloud component was an essential element of both read aloud and small group intervention, missing the read aloud would have skewed the results of the study, as students would have missed the opportunity to learn target words during the read aloud. The final participant count was 41 students. All 41 students completed the intervention and assessments, which were entered into the model during the primary analysis. As earlier discussed, one participant moved between the time of posttest and delayed posttest and was therefore removed from the ancillary analysis of repeated measures ANCOVA. In addition to student participants, three teachers also participated in the study. One teacher served as the control group teacher, while the other two conducted the read aloud portion of the intervention in their classrooms.

Intervention Fidelity

As discussed in the methodology, Chapter 3, a fidelity checklist was developed based on the developed routine for whole-group instruction, administered by classroom teachers. This checklist can be found in Appendix F. The researcher conducted three fidelity checks during weeks 2, 4, and 6. Results were shared with teachers after each fidelity check and suggestions were provided as needed to enhance the implementation of the intervention and maintain consistency between the two classrooms.

The first fidelity check revealed that teachers were generally following the routine, with one minor error. After stopping to discuss the target word in the book, teachers both forgot to re-read the sentence after explaining the definition and showing the picture card. The purpose of rereading the sentence was to provide students with an additional exposure to the word meaning in context. The researcher met with teachers to discuss the fidelity checklist and reviewed this portion of the routine. The second and third fidelity checks resulted in mastery of the routine, as teachers followed the routine without error and were consistent with one another in the administration of the read aloud instruction.

Baseline Data

Baseline data consisted of participants' standardized test scores on two measures: PPVT-IV and EVT-2. Standard scores from these assessments were included above in the descriptive statistics section of the primary analysis. Inclusion of this data in that section was important so that composite scores of standardized test scores could be viewed and understood in terms of how scores were used for grouping purposes.

Statistics and Data Analysis

During the study, all students that returned consent forms were included, with the exception of the four students that attended Title I during the daily intervention time. No participant was excluded from analysis based on unsatisfactory performance during the intervention period. Again, only one student was removed from the study during ancillary analysis due to a move away from the school.

Adverse Events

No adverse events took place in any treatment group. The only effects on participants were gains, or lack of gains, in target word knowledge over time. Students did not miss out on critical classroom instruction or guided reading time with their classroom teachers. Students were not penalized in any manner when absent or scoring incorrectly on assessments.

Discussion

This study represented an innovative and fresh approach to vocabulary instruction. One of the greatest gaps in literacy research involving vocabulary instruction is the void of small group, systematic vocabulary instruction beyond that of a read aloud, specifically in the primary elementary grades. Read aloud interventions have proven to be effective in bolstering the target word knowledge of students (e.g. Biemiller & Boote, 2005, 2006; Coyne et al., 2009; Penno et al., 2003); however, few researchers have explored the benefits of taking vocabulary instruction beyond the read aloud or testing the lasting effects of word learning attributable to read aloud-based instruction. Although the read aloud is a very critical component of a balanced literacy program and can supply elementary students with exposure to high quality literature and vocabulary, it is not sufficient as a one-stop-shop for vocabulary instruction. This study yielded results that break through this prominent gap by providing preliminary evidence for the effectiveness of systematic small group instruction that supplements read aloud-based vocabulary instruction, in the primary elementary grades.

Four themes emerged from the interpretation of the results generated from this study, including: (a) the use of best practices during read alouds yield better than average target words gains; (b) clear and concise vocabulary instruction requires consistency, clustering, and repetition; (c) expressive language demands and interactive activities lead to greater gains in vocabulary acquisition; and (d) a read aloud is only a place to start. Each theme has been described below, followed by implications for teachers, implications for future research, and general conclusions.

The Use of Best Practices During Read Alouds Can Yield Better Than Average

Words Gains

As previously stated, and corroborated by other researchers, read alouds can be an effective context for teaching words to primary grade students. However, the effectiveness of this context for vocabulary instruction depends on the quality and intensity of the read aloud. Read alouds can vary from a “plain old” read aloud, where a teacher simply reads a book aloud for the enjoyment of listening to a story, to an interactive read aloud, where students become engaged in discussions around concepts or words with the teacher. A “plain old” read aloud, while having the ability to provide enjoyment of reading and listening is simply is not an effective context for bolstering vocabulary knowledge, especially expressive knowledge of words. Expressive demands for students are very low during a “plain old” read aloud, as the teacher is reading and students are listening. However, when students can be interactive in the read aloud through teacher-provided extended and focused instruction that expands the content of the book or highlights specific vocabulary words, the setting can evolve into a very rich context for word learning.

An interactive and focused read aloud context is an appealing avenue for vocabulary instruction because teachers can discuss novel word meanings to students in a whole group setting and provide context for word learning from the books being read aloud. Survey data has shown that teachers use lots of narrative fiction texts when reading aloud and incorporate spontaneous word discussion when encountering words with which students may not be familiar (Lester, 2012). This notion is in conflict with research-based best practices. The read aloud has been explored by numerous

researchers and various approaches to the instruction that accompanies it have been tested. From these studies, a set of research-based, best practices have emerged that set the framework for the present study. Neuman and Dwyer (2009) suggested a set of best practices found within various reviews of vocabulary research (a) systematic and explicit instruction; (b) opportunities for students to use words (Pressley, 2001; Wharton-McDonald, Pressley, & Hampton, 1998); (c) interactive practice guided by teachers (Beck, McKeown, & Kucan, 2002; Hoffman, 1991); (d) review of vocabulary words (Brophy & Good, 1986; Rosenshine, 1986); and (e) progress monitoring of students' word knowledge (NICHD, 2000). Additional strategies that sprouted from these general guidelines included (a) providing extended definitions to students during read aloud-based vocabulary instruction; (b) allowing students opportunities to use words expressively; and (c) using informational texts that correlated with the target words.

Systematic and explicit instruction. Systematic and explicit vocabulary instruction was an overarching goal of this study and contributed to the successes evident in students' gains in target word knowledge. A specific protocol was followed to maintain consistency between classroom teachers and to provide students with a routine to which they could adapt and feel prepared for each day. The systematic approach allowed for structure that prepped students for what was coming next. During implementation of the protocol, teachers provided extended definition and in-depth information about the target words. For instance, examples, nonexamples, superordinate terms, pictorial representations and discussions of features related to the pictures were used to support the extended definitions. This level of intensity with word definitions allowed for greater discussion around words and went beyond surface level instruction

that commonly accompanies an embedded approach to vocabulary instruction or a spontaneous stating of a synonym or vague definition.

Expressive vocabulary use. A large gap in the research is the void of instruction that supports elementary students' expressive vocabulary uses. Most studies have centered on receptive language use, which is common and logistically sensible when teaching vocabulary in the context of a read aloud. In this study, I was able to embed several opportunities for students to expressively say and use the target words in each read aloud session. Specifically, during the read aloud, teachers asked students to repeat the target words of the day five different times throughout the discussion surrounding each target word. For example, teachers asked students to repeat the word the first time it was stated, before reading the book, a second time after it was encountered for the first time in the book, a third time after showing it written on an index card, a fourth time after discussing the extended definition of the word, and a fifth time after reading the book, as students engaged in a think-pair-share about the word meanings. Giving students opportunities to say and use the target words was unique to much vocabulary research. Few, if any studies were found that were centered on student-talk related to the vocabulary instruction during a read aloud. Although repeating words evoked only a low level of expressive demand, the inclusion of this element gave students a chance to be somewhat active in the read aloud, rather than simply listening passively.

Book Selection. Third, book selection was a key component of this study. All books used for the read alouds were informational. The purpose behind using informational books was to correlate the content of the books with the selection of Tier 2 and Tier 3 target words. Most vocabulary researchers have used narrative fictional texts

as the basis for read alouds (e.g. Biemiller & Boote, 2006; Coyne, McCoach, & Kapp, 2007; Maynard, Pullen, & Coyne, 2010), but few studies have used informational texts, with the exception of Fien et al., (2011), who used a combination of both narrative and informational texts. One of the values in using informational texts for vocabulary instruction of T2 and T3 words was that teachers could effectively teach content area words in conjunction with a read aloud. Future development of this vocabulary protocol could potentially be expanded to correlate with science or social studies instruction more specifically, given the nature of the words selected and books used.

In sum, much previous research has typically reflected an embedded approach to teaching vocabulary, in which words were discussed briefly when encountered during read alouds or content area instruction (e.g. Biemiller & Boote, 2006; Coyne et al., 2009). However, when a purely embedded approach was used, gains made in vocabulary knowledge were low. The average percent of words gained by participants in the present study were much higher due to the integration of an embedded approach that included extended instruction of target word meanings. For instance, Biemiller and Boote (2006) reported that average gains from vocabulary interventions prior to their own, ranged from 20%-25%. In the study conducted by Biemiller and Boote (2006), the greatest gains obtained were 41% when 7-10 word meanings were briefly explained in an embedded fashion during each reading and an additional day of review without reading was provided. Although the gains yielded in the research by Biemiller and Boote (2006) seem greater than those from other research, results from the present study showcased that students outperformed those in any previous research reviewed. Specifically, the read aloud group on average, gained 26% from pretest and posttest and 16% from pretest

to delayed posttest. Small group on average gained 52.6% from pretest to posttest and 29.9% from pretest to delayed posttest. If a review day had been included each week, such as in Biemiller and Boote (2006), target word gains would likely have been even higher. Hence, the uncommon increase in words learned in the present study can be partially attributed to the effective, research-based strategies used to create the read aloud protocol, but also to the small group instruction and semantic clustering of words.

Clear and Cohesive Vocabulary Instruction Requires Consistency, Clustering, and Repetition

In order to develop and implement high-quality, clear and cohesive vocabulary instruction in the primary elementary grades, three important features are necessary. These features include using consistent instructional techniques, clustering words semantically, and using repetition to increase word exposure. Using research-based strategies to teach vocabulary is one of the most powerful attributes of successful vocabulary instruction; however, using a variety of strategies haphazardly will not yield sufficient gains in vocabulary knowledge. Vocabulary instruction must be clear, which is reflected by following a systematic protocol, as described above. Vocabulary instruction must also be cohesive, as words are selected carefully and purposefully so that word learning builds in a way that cumulatively adds to a student's knowledge. Cohesive instruction leads to cumulative learning by gradually teaching students words that are related to one another within the context of broader concepts. For example, when teaching the cluster of habitats, the words *prairie* and *coast*, were taught for instance. Once these habitats were distinguished as habitats and features of habitats in general and of these particular examples were discussed, students were able to more quickly add the

words *marsh*, *swamp*, and *rainforest* to their repertoire of habitat-related words. Hence, knowledge of the cluster or concept of habitats was able to be slowly acquired by students as they gradually were introduced to words that were all related to one another. The cohesiveness of this approach was a strong element of the study that can potentially be linked to students' gains in target word knowledge. Further, repetition of words must be provided during instruction, as research has shown it takes many exposures to words to truly develop sufficient knowledge (e.g. Nagy & Scott, 2000). In a clustered approach to word learning, words in the cluster are discussed and naturally reviewed each day, providing students with repetition of words.

Using consistent instructional techniques. The read aloud protocol developed for the present study was based on research-based strategies, which several previous researchers have used and deemed successful (e.g. Fien et al., 2011; Maynard, Pullen, & Coyne, 2010; Neuman & Dwyer, 2009). In addition to being focused and interactive, in order for read alouds to be effective, teachers must also use consistent teaching techniques so that students can adapt to the routine and become familiar with the word learning process. During this study, students were not only learning words, but also learning how to learn words. Given that clear, cohesive, systematic, and focused vocabulary instruction occurred minimally in the classrooms at the targeted school prior to the onset of this study, students were unfamiliar with how to discuss words and how to acquire sophisticated knowledge of word meanings. During my own observations of the read alouds during the first week of the study, students seemed slightly unsure of how to engage in vocabulary instruction, given that the routine was new in their classrooms; however, through participation in the daily read aloud routine, students quickly began to

chime in when stating the target words and thoroughly enjoyed the think-pair-share time. This was also the case when working with students in small groups; the students had to adapt to the routine of discussing and interacting with words and word meanings. In both contexts, students learned routines and knew what was coming next, which allowed them to be mentally prepared for the demands and expectations of expressive language use.

This study was clearly designed to reflect high-quality and systematic small group instruction in addition to the read aloud instruction. This added instructional component took vocabulary learning to a level not commonly seen in other vocabulary research. The small group protocol also followed a systematic format that was based on research findings supporting certain vocabulary strategies and activities. Again, students quickly adapted to the routine and became active participants in their learning. This small group vocabulary was the most important element of the study, as the student participants made greater gains in target word knowledge than students in the read aloud only or control groups and retained the greatest numbers of target words over time. No previous study was able to be located that reflected the same results as the present study. Overall, the systematic approach taken with both the read aloud and small group instruction was what made the true difference in students' word learning and retention.

Semantically clustered words. In addition to using a systematic protocol for instruction, words were clustered semantically. The notion of semantic clustering arose first, from the lack of meaning-based word selection described in much research. For example, although some research reflects systematic word selection, in terms of selecting from a list (e.g. Beck and McKeown's Tier 2 words), few researchers followed word selection criteria based on meaning, especially semantic relatedness. For example,

Neuman and Dwyer (2009) analyzed a variety of vocabulary curriculums and found that some curriculums included words for instruction that although based on thematic unit were not related. For instance, the word *cooperating* was selected from a unit called “working together”, while other words were selected from books based on the topic, such as animals. These same researchers found that in other curriculums, the words *move* and *ride*, for instance, were suggested for instruction (Neuman & Dwyer, 2009). Neuman and Dwyer (2009) reported that the words *move* and *ride* are words that children typically learn by age 30 months, negating the importance of explicit teaching in elementary school. In another study, Kessler (2011) selected the words: *paddle-shaped*, *flippers*, *hind*, *rudders*, *clumsy*, *drag*, *energy*, *shore*, *exhausted*, and *effort* for instruction, which were selected from a book about sea turtles. Although this selection of words was representative of challenging words for which comprehension of the text would be improved upon by knowledge of these words, the semantic relatedness of the words is low. For instance, the words *hind* and *energy* do not appear to have a great deal of semantic relatedness beyond inclusion in the same book. Although this researcher went on to explain the words to students during instruction by linking meanings to known words and concepts, the words themselves were not semantically related and therefore, students could not make connections between them. In sum, most curriculums explored by Neuman and Dwyer (2009) did not reflect word selection that was based on guidelines set forth by researchers, such as Beck and McKeown (2007) or Biemiller (2010). Selecting words haphazardly from word lists is no better than randomly selecting words without referencing word frequency or difficulty tiers. Hence, semantic clustering has

potential to serve as a potent ingredient in primary students' word learning and should be added to the list of effective and systematic word selection criteria.

Although researchers have selected words systematically in the past, the selection criteria are often limited to the corpus from which words are chosen, such as T2 lists or words simply found in a selection of children's narratives. Additionally, many researchers have failed to thoroughly describe their selection processes, creating further ambiguity about the approach taken when selecting words. Semantic clustering was also based on the notion highlighted by Nagy, Anderson, and Herman (1987). These researchers suggested that conceptual difficulty, the relationship between a new word meaning and existing knowledge regarding the broader concept of the word, had the most influence on a students' ability to learn novel words from context. Therefore, having a strong basis for a concept can aid in the acquisition of related words. Based on this notion, semantic clustering of words was used to support word learning by allowing for cumulative learning to occur, as knowledge was able to be built gradually through the instruction and addition of related words each day.

To help fill the gap in word selection processes, teachers must consider students' schemata. Semantic clustering of words plays a role that directly fits within schema theory, as students have opportunities to connect new word learning with existing word knowledge and continue to build upon that knowledge as new words are added to their schema. In the present study, a large amount of effort was placed on grouping words into semantic clusters in order to foster the building of schema, in hopes of helping students realize their ability to acquire new words from known concepts. Additionally, efforts were made to follow guidelines set forth by prominent researchers in the field (e.g. Beck

& McKeown, 2007; Biemiller, 2006). The goal was to select words that truly reflected concepts second grade students needed to know in order to be successful in their content area classes and on state assessments. Semantically clustering words for instruction bolstered this study as students were able to see the relationships between words they were learning each day during a week's cluster. Relationships between the words were discussed as more words were added to the cluster each day. For example, one week's semantic cluster was related to animals and animal adaptations. The cluster included the words: *amphibian, mammal, reptile, hibernation, migration, and camouflage*. As students discussed the types of animal groups in a small group setting, they were also able to make connections between the animals and compare and contrast the types of adaptations various animals within different groups had. In general, word selection was one of the most important attributes of this study and aided students' abilities to hear and be a part of successful discussions around words, while acquiring more in-depth knowledge about specific words and related concepts. This result may not have been found if the corpus of words selected had been random and unrelated to one another. This notion speaks volumes about the small successes other researchers have had in developing interventions that yield sufficient gains in vocabulary knowledge. Word selection potentially holds the key for bolstering language acquisition, as students seem more likely to gain knowledge of words when clear relationships between concepts can be formed.

Repetition of words. Providing repetition of words was another important element in providing clear and cohesive instruction. Teaching words in semantic clusters was the perfect avenue for providing repetition of words because each day new words

were added to a cluster of words, such as animal adaptations or geographical features. Although originally the read aloud protocol included a rereading segment, which was included to provide students to two exposures to a book each day, this instructional component added approximately 20 minutes to the read aloud block, which teachers at the school simply could not fit into their literacy schedule. Rereading has been proven an effective method for bolstering word knowledge and many researchers have conducted studies which yielded stronger results for rereading during interventions than only administering a sole reading of a book (e.g. Brabham & Lynch-Brown, 2002; Brett, Rothlein, & Hurley, 1996; Elley, 1989; Hargrave & Senechal, 2000; Robbins & Ehri, 1994; Senechal, 1997; Senechal, Thomas, & Monker, 1995). In future research, especially action research, teachers could flexibly alter their instructional timeframes to support repeated readings. However, being the researcher and entering into classrooms that were not my own, posed particular logistical difficulties; one being the need for greater time allowance in order to fit in two readings of a book each day.

Given that the logistical constraints of this study prevented repeated readings from being implemented, other measures were taken to ensure that students had repeated exposures to words throughout the read aloud and small instruction. In the first part of each read aloud session, teachers engaged students in chorally reading all of the words in the cluster of the week. Each target word was written on a large index cards and placed in a pocket chart, in the read aloud area, with corresponding pictures. Students were able to see the entire cluster of words and pictures each day, although only two words were taught thoroughly each day. This component alone gave students varied opportunities to see the words as they walked by the pocket chart during the day and referenced the

pictures and cards during their small group sessions. Additionally, the selection of books for each week related to the semantic cluster of words and although each day's book was different, frequently a book contained multiple target words. Therefore, even on a day when the words *hibernate* and *migrate* were being taught, for instance, the words *mammals*, *amphibians*, or *reptiles* may have also been read during the read aloud. Semantic clustering of words contributed to further exposure to words within clusters, increasing the repetition that students had to hear and say words. Neuman and Dwyer (2009) suggested that review of words was an essential element of systematic and successful vocabulary instruction. Biemiller and Boote (2006) also found that students learned and retained more target words when a review day was implemented during each week of vocabulary instruction. Hence, teachers reviewed the entire cluster of words at the start of each read aloud session, which gave students additional opportunities to hear the words read aloud in the context of a particular semantic concept.

Expressive and Interactive Activities Lead to Greater Gains in Word Knowledge

One of the greatest gaps in vocabulary research in the primary elementary grades to date is the void of instruction centered on fostering expressive vocabulary knowledge. Most research interventions have fostered receptive acquisition of language, given that most interventions occur in the context of a read aloud, and only two recent studies included an expressive measure of word knowledge (Coyne et al., 2009; Fien et al., 2011). One exception to this statement would be the robust line of work on dialogic reading in preschool classes, conducted by Whitehurst and colleagues (e.g. Lonigan & Whitehurst, 1998; Whitehurst, Arnold, Epstein, Angell, Smith, & Fischel, 1994; Zevenbergen & Whitehurst, 2003). Dialogic reading is an interactive approach to shared

reading, where students take an active lead in telling the stories and teachers provide feedback. This format for reading aloud provides students with ample opportunities to expressively use language through interactions with adults and peers. However, this read aloud format and allowance for expressive language use is not supported in the same manner in elementary school, where the gap in expressive language use becomes more prominent.

The main goal of the present study was to incorporate systematic vocabulary instruction beyond that of a read aloud and additionally, to supply students with opportunities to expressively use language. Read alouds lend themselves well to fostering receptive language, as students have opportunities to hear many words used in the context of books read aloud by a teacher. The drawback to providing vocabulary instruction only through a read aloud is that although receptive language is strongly fostered, the promotion of expressive language takes somewhat of a back seat. Read aloud settings typically reflect large groups of approximately 18-25 students sitting on the carpet in front of a reader. When reading aloud to this many students, it is nearly impossible to provide each child with one or more opportunities to expressively use the words being taught. Although the present study aimed to provide students with approximately five opportunities to repeat the target words during each read aloud session, as described in theme one of the discussion section, including one think-pair-share, the majority of the expressive uses were choral statements of the words and only one opportunity to also expressively discuss the meaning of each word. The read aloud in this study provided greater opportunities for students to use words in a read aloud context than previous studies (e.g. Biemiller & Boote, 2006; Robbins & Ehri, 1994;

Senechal, LaFevre, Hudson, & Lawson, 1996); however, logistically, teachers simply could not allow students much more individual time to say and use novel words in a whole group setting.

Hence, the inclusion of the small group vocabulary instructional component was designed and implemented with the intent of allowing students multiple opportunities to say, use, and discuss target words. The small group instruction was interactive in nature, which Neuman and Dwyer (2009) suggested as another guideline that should be followed when implementing vocabulary instruction. In the small group setting, students had many opportunities to expressively use the target words. Student-led discussions around the words took place daily, where each student had something important to share, whether it was an example, related experience, question, or comment. This format for discussion and level of interaction between students regarding the target words went above and beyond that of the read alouds, where teachers were doing the reading and most of the explaining of words. Additionally, in the small groups, students were asked to share definitions of the words and include features of the words used by the teacher during the read aloud. Often times, students referenced the read aloud books they had listened to when providing examples, nonexamples or related concepts to a particular word. Although the read aloud set students up nicely for having some background about the words when coming to the small group sessions, without the small group instruction, students would not have had a context in which they could each share their ideas and information freely. Nor would students have been able to receive specific and individualized feedback on word uses from the teachers.

Students also participated in interactive activities about the words, beyond the discussions of words. For example, semantic feature analysis charts were one of the students' favorite activities. This activity required students to think about the similarities and differences between the words, as well as, specific features about the words. In order for students to be successful in content area subjects, features of words must be known as well as relationships between words. Although teachers can discuss features and relationships during read alouds, research shows that student-led discussions and student involvement in activities related to the words aids in greater gains in word knowledge (Stahl & Clarke, 1987). Venn diagrams and pictures sorts also supported the interactive framework. These activities were centered on the discussions and gave students opportunities to apply and demonstrate their depth of knowledge about the target words. These types of activities, although doable in a whole group setting, would not prove as successful in a read aloud context, given that not every child could be involved nor have opportunities to share their personal ideas and receive specific feedback.

A “Plain-old” Read Aloud is Only a Place to Start

Beck and McKeown (2007) referenced multiple studies centered on the effects of read alouds that yielded what they considered to be “nonexistent to unimpressive” gains (Biemiller & Boote, 2006; Elley, 1989; Nicholson & Whyte, 1992; Penno, Wilkinson, & Moore, 2002; Robbins & Ehri, 1994; Se´ne´chal, Thomas, & Monker, 1995). Evident from the themes presented in this discussion is that the elements used in the present study did in fact contribute to the success students had in learning and retaining target words. However, critical to note is that although a systematic read aloud was included and yielded strong initial results in word knowledge gained, my data and that of others

suggested that read alouds in and of themselves are simply not enough to provide students with the exposure to target words or opportunities to use words necessary for complete acquisition, retention and future usage. A read aloud is, however, a great place to start and can serve as a powerful context for introducing words to students.

Evident from this study is that a systematic read aloud that provides instruction on a selection of words based on standards and classroom content subject matter, extended word meanings, pictures that represent words and expressive opportunities for students to say and use words is the best approach to introducing vocabulary words during a read aloud. This introduction to word learning can give many students the boost they need to enhance overall word knowledge; however, for most students additional instruction will be needed to make the learning effects last. Simply reading aloud and discussing words spontaneously as many teachers have reported doing (Lester, 2012), will not bring about ample word learning and especially word retention.

After the conclusion of this study and data analysis, the notion of small group instruction as the active ingredient in the successes students had in gaining target word knowledge became more complex. The discussion themes that emerged from this analysis pinpoint more than simply the context of word learning: read aloud vs. small group. It was the elements and features of the small group instruction that surfaced as noteworthy, not solely the small group physical environment as opposed to a read-aloud environment. Specifically, a “plain-old” read aloud can provide great exposure to words, while a systematic and focused read aloud can provide effective introductions to target words. However, features specific to a small group setting made the difference in students’ word learning in this study.

The small group instruction provided in this study gave students additional time to interact with and discuss target word meanings. Students also had opportunities to not only use the words, but confuse the words, which is one of the greatest take-aways from the study. In a read aloud setting, teachers are simply unable to listen to each child discuss the words or provide specific feedback when words are used incorrectly. However, in a small group setting, students can try out the words by sharing examples, nonexamples, experiences related to words and concepts, related words, etc. It is not important that students use the words right every single time, but that when they do confuse a word meaning, they receive feedback on their expressive use through scaffolded instruction to acquire a deeper and more fine-tuned understanding of the word. Additionally, students in the primary grades especially, need opportunities to sit with a teacher and engage in interactive activities that support their learning. Read aloud settings cannot provide enough individual participation from students to foster the acquisition of nuanced understandings of words, but in a small group setting, all students can be active participants with their peers and with a teacher. Small group instruction provides a feedback loop that is invaluable and can rarely be provided in a larger setting. In order for kids to use, confuse, and solidify their knowledge of words, this feedback is critical and was one of the strongest active ingredients in the study.

Hence, from this study, it is my goal that all readers will take away this one critical point: a read aloud must be supplemented with meaningful, focused, small group instruction that incorporates all of the above attributes, so that students have greater opportunities to use words themselves and interact with words while receiving feedback. Hearing target words during a read aloud provides a basis for the word learning and

presents words in the context of a book; however, without opportunities to use words expressively, students will not be able to acquire a sufficient amount of vocabulary and will likely not retain enough knowledge to apply when comprehending text or learning related information. For teachers that feel small group instruction is unattainable, their goal should at least be to move away from paper-pencil activities and being to realign instructional time with expressive language activities. Students must be able to discuss words and engage in deep-processing of word meanings by comparing, contrasting, discussing features of words, and making connections and building relationships between words and concepts.

Implications for Teachers

Evident from the present study are several implications for teachers. Implications deemed important for instruction of vocabulary are as follows: words should be selected systematically and clustered systematically, read alouds should be supplemented by small group instruction, and students need opportunities to use target words expressively, not just receptively.

Words should be selected systematically and clustered semantically. In professional resource books for teachers, word selection for vocabulary instruction is often unclear and non-systematic. Most research studies do not reveal the method used for selecting words for interventions beyond selection from Tier 2 or 3 lists, which is systematic but not solely enough. This lack of clarity can potentially leave teachers unsure of the best guidelines to follow when selecting words for instruction. Teachers should take a multifaceted approach when selecting words for instruction. First, teachers need to consider the grade level standards set forth by state and national guidelines.

Second, teachers need to align words deemed important on these standards lists with words that are considered high-priority words. Two texts useful in aiding in this decision making process are the works of Biemiller (2010) and Beck, McKeown, and Kucan (2002). These texts provide information on the various tiers of words and words that students need to know and frequently don't know by certain grade levels, which emphasize the importance of instruction of the words. Third, once standards and research-based guides for tiers of words have been considered and aligned, teachers need to align the words with the scope and sequence of their curriculums to ensure that words being taught during vocabulary instruction coincide with the other content area subject matter being taught throughout the school day. Fourth, once the curriculum has been considered, teachers should cluster words by semantic relatedness. At this point in the selection process, grouping words should be somewhat easy to do, given that words are now aligned with curriculum subject matter. In sum, word selection can be a tedious and lengthy endeavor, but nonetheless, is of utmost importance and truly makes a difference in the relationships students make between words and the discussions and meaningful interactive activities that can be used to bolster word learning.

Read alouds should be supplemented by small group instruction. Once words are systematically selected, instruction of the words should be carefully planned. As discussed earlier in this chapter, a read aloud is a great place to start vocabulary instruction, but should not be the only context for word learning. Too often is the read aloud being used as the sum of vocabulary instruction, when in fact, it should be one dimension of a very multifaceted and intricate system of instruction.

The read aloud can provide all students in a class with a strong introduction to words, as teachers have opportunities to say and define words in a large group setting, while discussing and reading about the word in the context of a book. This receptive dimension of vocabulary instruction is not enough as a standalone element to bolster true word learning for students, but is an important piece that can set the groundwork for future and more in-depth learning. During a read aloud, there are several steps that teachers should take to ensure the instruction provided is effective.

First, books chosen for read alouds should reflect the semantic cluster of words being taught. Informational texts are a great resource for tying together literacy, science, and social studies and most commonly reflect words that students will need to know for state assessments and in their content area subjects. Second, teachers need to ensure that they provide students with extended word meanings during read alouds. An embedded approach to instruction simply is not enough to produce gains in word knowledge among students. Predetermining and constructing meaningful and extended definitions for will help teachers feel prepared for thoroughly explaining words when encountering them in text. Third, teachers must preview books prior to reading to ensure that the target words being taught are included in the book and that the cluster of words being focused on is reflected in the text. Survey data from primary grade teachers on read aloud practices involving vocabulary instruction, have shown that teachers too frequently do not preview texts prior to reading or preselect words to teach during the read aloud (Lester, 2012). Spontaneous explanation of words in texts can result in the offering of superficial definitions that do not accurately or thoroughly portray features or dimensions of words. The culmination of careful word selection, purposeful book selection, and a systematic

read aloud protocol can truly bolster word learning for primary grade students. However, important to keep in mind is that although a read aloud is a great context for introducing words to students, it does simply that: provide an introduction. This method for vocabulary instruction is not the end-all, be-all method, but a starting place.

Students need opportunities to use target words expressively, not just receptively. Listening to read alouds is not enough to solely bolster word learning. Teachers need to provide students opportunities to expressively use words being taught during a read aloud. Expressively using words may involve giving students opportunities to say the words multiple times, discuss the words and word meanings with peers, make comparisons between words, share examples and share related experiences related to the words. The level of knowledge required to use words expressively is greater and holds students to a higher standard of word learning. It is simply not enough for students to sit and listen to the words read aloud by adults. The method used in the present study, that was effective in giving students these opportunities, provided several times for students to see and repeat the target words as well as engage in think-pair-share discussions with a peer, where each student was asked to state one of the target words and explain the corresponding definition.

When planning vocabulary instruction in a whole group setting, teachers should at least build in opportunities for students to state target words, although simply stating words is a low level of expressive word use demand. When possible, opportunities for student discussion around the word meanings should be included, but as previously discussed, logistically a read aloud setting provides few opportunities for all students to be active participants in a discussion. A think-pair-share is an effective and quick

activity that can give students opportunities to expressively use words during read alouds. When teachers are able to expand word learning beyond that of the read aloud, students should be involved in activities that allow them to discuss words with peers and problem solve the similarities and differences between words. The continuum below showcases varying levels of expressive demand required by certain word-related activities. In order to foster expressive language and guide students into using words independently, teachers should incorporate opportunities for students to participate in activities that reflect a variety of these dimensions and gradually incorporate more demanding activities over time. For example, after modeling how to complete Venn diagrams, semantic feature analysis charts, and picture concept sorts, for instance, students could attend a center with peers to collaborate on completing these graphic organizers, which would naturally foster language use in a real setting, unprompted by teacher input. Teachers can also hold small group sessions during content area times or during literacy rotations to listen to students use words and provide necessary feedback.

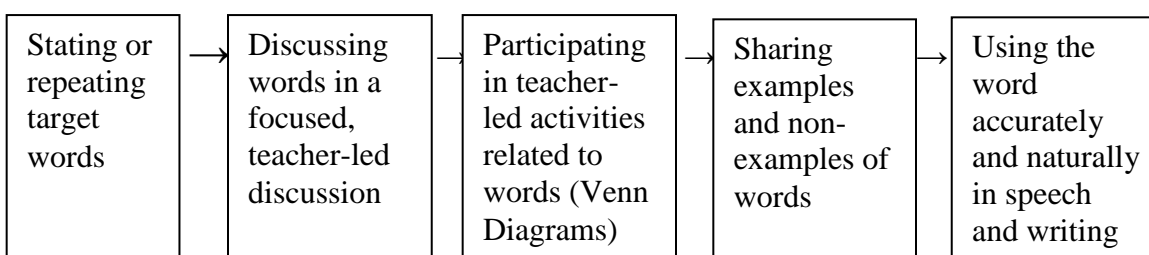


Figure 4. Continuum of expressive demand. This figure illustrates a continuum of vocabulary activities of varying expressive demand.

Implications for Future Research

Several implications for future research emerged from this study. First, the notion of heterogeneous versus homogeneous grouping for small group instruction needs further exploration. Heterogeneous grouping was selected for this study so that students could

not only learn from the teacher, but also share ideas and learn from one another during peer discussions. Also, given that similar instruction was necessary across all small groups for the purpose of equal comparison in this study, future researchers should compare across different forms of small group instruction, including the differences in homogeneous grouping versus heterogeneous grouping. From such a study, the level of benefit of small group instruction for students with varying entering vocabulary knowledge may emerge and could provide teachers with knowledge of which students may benefit most from small group instruction.

Second, more research needs to be conducted related to the effects of selecting words based on semantic relatedness. Although semantic relatedness was an element of the selection criteria in this study and believed to be a powerful ingredient in word learning, the effects on word learning were not statistically tested. In the future, researchers could compare the word learning gains made by students participating in instruction based on semantically related words versus non-semantically related words.

Third and finally, more research is needed to test the effects of embedding vocabulary instruction into content areas and in conjunction with nonfiction texts. Longitudinal data on the increases in students' vocabularies over time and impact on classroom performance and high-stakes testing over time is needed to determine if content-specific vocabulary instruction should be more heavily focused on in classrooms.

Conclusion

The findings presented in this research set the groundwork for future small group vocabulary research. The preliminary support that this study highlighted for the effectiveness of small group vocabulary instruction should be heavily considered by

future researchers. Specifically, now clear is the notion that the effects of read aloud-based vocabulary instruction are not independently significant enough to contribute to long-term vocabulary knowledge and growth. This information is especially critical for students entering elementary school with below average vocabularies. If teachers truly want to enhance the comprehension levels of their students and to bolster overall vocabulary knowledge in the primary grades long-term, then instruction beyond that of a read aloud must take place. Students simply do not receive enough opportunities to expressively use and discuss words in a read aloud setting to yield sufficient gains in vocabulary knowledge. Small group instruction can give students opportunities to discuss words with their peers, while receiving feedback on their word uses from teachers. Without these small group experiences, students must rely on receptive language acquisition during read alouds, which is not sufficient in learning and retaining target words.

References

- Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association.
- Anderson, R. C., & Nagy, W. E. (1992). The vocabulary conundrum. *American Educator*, 14-18, 44-47.
- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58(10), 1-166.
- Ard, L. M., & Beverly, B. L. (2004). Preschool word learning during joint book reading: Effect of adult questions and comments. *Communication Disorders Quarterly*, 26(1), 17-28.
- Baker, S. K., Fien, H., & Baker, D. (2010). Robust reading instruction in the early grades: Conceptual and practical issues in the integration and evaluation of Tier 1 and Tier 2 instructional supports. *Focus on Exceptional Children*, 41, 1–20.
- Baumann, J. F., Kame'enui, E. J., & Ash, G. E. (2003). Research on vocabulary instruction: Voltaire redux. In J. Flood, J. Jensen, D. Lapp, & J. R. Squire (Eds.), *Handbook of research on teaching the English language arts* (2nd ed., pp. 752–785). New York: MacMillan.
- Baumann, J. F., & Kame'enui, E. J. (2004). *Vocabulary instruction: Research to practice* (pp. 28–40). New York: Guilford Press.
- Beck, I. L., & McKeown, M. G. (2001). Text talk: Capturing the benefits of readaloud experiences for young children. *The Reading Teacher*, 55, 10-20.
- Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral

- vocabulary repertoires through rich and focused instruction. *Elementary School Journal*, 107(3), 251-271.
- Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction. Solving problems in the teaching of literacy*. New York: Guilford Press.
- Beck, I., McKeown, M., & McCaslin, E. (1983). All contexts are not created equal. *Elementary School Journal*, 83, 177-181.
- Beck, I., McKeown, M., & Omanson, R. (1987). The effects and uses of diverse vocabulary instructional techniques. In M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 147-157). Hillsdale, NJ: Erlbaum.
- Beck, I. L., Perfetti, C., & McKeown, M. G. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506–521. doi:10.1037/0022-0663.74.4.506
- Becker, W. C. (1977). Teaching reading and language to the disadvantaged: What we have learned from field research. *Harvard Educational Review*, 47, 518-543.
- Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *American Educator*, 47, 24–28.
- Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology*, 24, 323-335.
- Biemiller, A. (2004). Teaching vocabulary in the primary grades: Vocabulary instruction needed. In J. F. Baumann & E. J. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 28–40). New York: Guilford Press.
- Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In E. H. Hiebert & M.

- L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practices* (pp. 223-242). Mahwah, NJ: Erlbaum.
- Biemiller, A. (2010). *Words worth teaching*. Columbus, OH: McGraw Hill.
- Biemiller, A. J., & Boote, C. (2006). An effective method of building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44-62.
- Biemiller, A. & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498–520.
doi:10.1037/0022-0663.93.3.498
- Blewitt, P., & Walsh, B. A. (2006). The effect of questioning style during storybook reading on novel vocabulary acquisition of preschoolers. *Early Childhood Education Journal, 33*(4), 273-278.
- Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. *Elementary School Journal, 96*(4), 415-422.
- Brabham, E. G., & Lynch-Brown, C. (2002). Effects of teachers' reading-aloud styles on vocabulary acquisition and comprehension of students in the early elementary grades. *Journal of Educational Psychology, 94*(3), 465-473.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Byrnes, J. P., & Wasik, B. A. (2009). *Language and literacy development: What educators need to know*. New York, NY: Guilford.

- Carey, S. (1978). The child as a word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Cassidy, J., & Loveless, D. J. (2011). Taking our pulse in a time of uncertainty: Results of the 2012 *What's Hot, What's Not* literacy survey. *Reading Today*, 29(2), 16-21.
- Chall, J. S., & Jacobs, V.A. (2003). Poor children's fourth-grade slump. *American Educator*, 27(44), 14-15.
- Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.
- Cohen, J. (1988). *Statistical power analysis of the behavioral sciences*. (2nd ed.). New York: Academic Press.
- Condouris, K., Meyer, E., & Tager-Flasberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology*, 12, 349-358.
- Coyne, M. D., Capozzoli-Oldham, A., & Simmons, D. C. (2012). Vocabulary instruction for young children at risk of reading difficulties: Teaching word meanings during shared storybook readings. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (pp. 51-71). New York: The Guilford Press.
- Coyne, M., McCoach, B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students. *Learning and Disabilities Quarterly*, 30, 74–88.
- Coyne, M. D., McCoach, B. D., Loftus, S., Zipoli, R., & Kapp, S. (2009). Direct

- vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, 110(1), 1-18.
- Coyne, M. D., McCoach, B. D., Loftus, S., Zipoli, R., Ruby, M., Crevecoeur, Y., et al. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, 3 93-120.
- Coyne, M., Simmons, D., Kame'enui, E., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, 12, 145–162.
- Coyne, M. D., Zipoli, R. P., Chard, D. J., Faggella-Luby, M., Ruby, M., Santoro, L. E., et al., (2009). Direct instruction of comprehension: Instructional examples from intervention research on listening and reading comprehension. *Reading and Writing Quarterly*, 25, 221-245.
- Creswell, J. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oakes, CA: Sage Publications, Inc.
- Cunningham, A.E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934-945.
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, 22(1-2), 8–15.
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, 42, 895-901, 948.
- Dale, E., & O'Rourke, J. (1981). *The living word vocabulary*. Chicago: World Book/Childcraft International.
- DeTemple, J., & Snow, C. E. (2003). Learning words from books. In A. van Kleeck, S.

- A. Stahl, & E. B. Bauer (Eds.) *On Reading Books to Children: Parents and Teachers* (pp. 16-36). Mahway, New Jersey: Larence Erlbaum Associates, Inc.
- Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly, 29*(2), 104-122.
- Douglas-Hall, A., Chau, M., & Koball, H. (2006). *Basic facts about low-income children*. New York: National Center for Children in Poverty.
- Duncan, G., Brooks-Gunn, J., & Klebanov, P. (1994). Economic deprivation and early Childhood development. *Child Development, 65*, 296- 318.
- Dunn, L. M., & Dunn, L. M. (2007). Peabody Picture Vocabulary Test (4th ed.). Circle Pines, MN: American Guidance Service.
- Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 4*, 174–187.
- Feitelson, D., Kita, B., & Goldstein, Z. (1986). Effects of listening to series stories on first-graders' comprehension and use of language. *Research in the Teaching of English, 20*, 339-356.
- Fien, H., Santoro, L., Baker, S. K., Park, Y., Chard, D. J., Williams, S., & Haria, P. (2011). Enhancing teacher read alouds with small-group vocabulary instruction for students with low vocabulary in first-grade classrooms. *School Psychology Review, 40*(2), 307-318.
- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied psycholinguistics, 24*, 369-383.
- Foorman, B. R., & Connor, C. M. (2011). Primary grade reading. In M. L. Kamil, P. D.

- Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research: Volume IV* (pp. 136-156). New York: Routledge.
- Graves, M. F., Brunetti, G. J., & Slater, W. H. (1982). The reading vocabularies of primary grade children of varying geographic and social backgrounds. In J. A. Niles & L. A. Harris (Eds.), *New inquiries in reading research and Instruction: 31st yearbook of the National Reading Conference* (pp. 99-104). Rochester, NY: National Reading Conference.
- Green, S. B., & Salkind, N. J. (2011). *Using SPSS for windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, NJ: Pearson Education Inc.
- Hargrave, A. C., & Senechal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15*, 75-95.
- Hart, B., & Risley, T. R. (Eds.). (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes
- Herman, P., Anderson, R., Pearson, P. D., & Nagy, W. (1987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly, 22*, 263-284.
- Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont, CA: Cengage Wadsworth.
- Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45*, 337-374.
- Institute for Education Sciences (n.d.). Common core of data: Search for public school

- districts. Retrieved from: <http://nces.ed.gov/ccd/districtsearch/>
- Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergarteners. *Language, Speech, and Hearing Services in Schools, 36*, 17-32.
- Kamil, M. R., & Hiebert, E. H. (2005). Teaching and learning vocabulary: Perspectives and persistent issues. In E. H. Hiebert and M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 1–23). Mahwah, NJ: Erlbaum.
- Kame'enui, E., Carnine, D., & Freshi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly, 17*(3), 17-32.
- Kessler, T. (2011). Shared reading to build vocabulary and comprehension. *The Reading Teacher, 64*(4), 272-277.
- Lester, L. M. (2011). An investigation of rural teachers' vocabulary practices, perceptions, and beliefs. A paper presented at the annual meeting of the Literacy Research Association, Jacksonville, Florida.
- Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. *Early Childhood Research Quarterly, 13*, 263-290.
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science, 11*(3), F9-F16.
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research,*

- 80(3), 300–335.
- Marzano, R. J., & Marzano, J. S. (Eds.). (1988). *A cluster approach to elementary vocabulary instruction*. Newark, Delaware: International Reading Association.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Maynard, K. L., Pullen, P. C., & Coyne, M. D. (2010). Teaching vocabulary to first grade students through repeated shared storybook reading: A comparison of rich and basic instruction to incidental exposure. *Literacy Research and Instruction, 49*, 202-242.
- McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*, 482-496.
- McKeown, M. G., & Beck, I. L. (2003). Taking advantage of read-alouds to help children make sense of decontextualized language. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 159-176). Mahwah, NJ: Lawrence Erlbaum Associates.
- McKeown, M. G., Beck, I., Omanson, R., & Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior, 15*(1), 3-18.
- McKeown, M. G., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*, 522-535.
- McKeown, M. G., Beck, I. L., & Sandora, C. (2012). Direct and rich vocabulary

- instruction needs to start early. In E. J. Kame'enui & J. R. Baumann (Eds.), *Vocabulary Instruction: Research to Practice* (2nd ed.), (pp. 17-33). New York: The Guilford Press.
- McLloyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53, 185-204.
- McMillan, J. H. (2008). *Educational research: Fundamentals for the consume* (5th ed.). Boston, MA: Allyn and Bacon.
- Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53(2), 253-279.
- Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52-77). New York: Guilford.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19, 304-330.
- Nagy, W., Anderson, R., & Herman, P. (1987). Learning words from context during normal reading. *American Educational Research Journal*, 24, 237-270.
- Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-36). Hillsdale, NJ: Erlbaum.
- Nagy, W., Herman, P., & Anderson, R. (1985a). Learning words from context. *Reading Research Quarterly*, 20, 233-253.

- Nagy, W., Herman, P., & Anderson, R., (1985b). Learning word meanings from context: How broadly generalizable? Champaign, IL: University of Illinois, Center for the Study of Reading.
- Nagy, W. E. & Hiebert, E. H. (2011). Toward a theory of word selection. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research: Volume IV* (pp. 388-404). New York: Routledge.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 269-284). Mahwah, NJ: Erlbaum.
- National Assessment Governing Board. (2007). *Reading framework for the 2009 National assessment of education progress*. Washington, DC: U.S. Department of Education.
- National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction [on-line]. Retrieved from:
<http://www.nichd.nih.gov/publications/nrp/report.cfm>.
- Neuman, S. B. (2011). Enhancing the intensity of vocabulary instruction for preschoolers at-risk: The effects of group size and supplemental instruction. A paper presented at the annual meeting of the Literacy Research Association, Jacksonville, Florida.
- Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-k. *The Reading Teacher*, 62(5), 384-392.
- Nicholson, T., & Whyte, B., (1992). Matthew effects in learning new words while listening to stories. In C. K. Kinzer & D. J. Leu (Eds.) *Literacy research, theory*

- and practice: Views from many perspectives: Forty-first yearbook of the National Reading Conference* (pp. 499-503). Chicago: National Reading Conference.
- Nino A. & Snow, C. E. (1996). *Pragmatic Development*. Boulder, CO: Westview
- Olejnuk, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241-286.
- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282-296.
- Penno, J. F., Wilkinson, A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23-33.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (Published as Vol. 11 of the series *Studies in Written Language and Literacy*). Philadelphia: John Benjamin.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383.
- Pollard-Durodola, S. D., Gonzalez, J. E., Simmons, D. C., Kwok, O., Taylor, A. B., Davis, M. J., . . . Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children, 77*(2), 161-183.
- Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the

- PPVT-III. *Language, Speech and Hearing Services in Schools*, 37(1), 5–16.
- Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, 86, 54–64.
- Rossman, G. B., & Rallis, S. F. (2011). *Learning in the field: An introduction to qualitative research* (3rd ed.). Thousand Oak, CA : Sage.
- Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Child Language*, 24, 123–138.
- Se´ne´chal, M., & Cornell, E. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly*, 28, 360-375.
- Senechal, M., Thomas, E., & Monker, J. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology* 87(2), 218-229.
- Shuttleworth, M. (2009). Within Subject Design. Retrieved from:
<http://www.experiment-resources.com/within-subject-design.html>
- Silverman, R., & Crandell, J. D. (2010). Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly*, 45(3), 381-340.
- Snow, C. E., & Goldfield, B. (1983). Turn the page please: Situation-specific language acquisition. *Journal of Child Language*, 10, 551-569.
- Stahl, S. A., Burdge, J. L., Machuga, M. B., & Stecyk, S. (1992). The effects of semantic grouping on learning word meanings. *Reading Psychology*, 13(1), 19-35.
- Stahl, S. A. & Clark, C. H. (1987). The effects of participatory expectations in classroom discussion on the learning of science vocabulary. *American Educational*

- Research Journal, 24(4), 541–555.
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56(1), 72-110.
- Stahl, S. A., & Nagy, W. E. (Eds.). (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.
- Stahl, S. A., Richek, M. A., & Vandevier, R. J. (1991). Learning meaning vocabulary through listening: A sixth grade replication. In J. Zutell & S. McCormick (Eds.), *Learner factors/teacher factors: Issues in literacy research and instruction. Fortieth Yearbook of the National Reading Conference* (pp. 185–192). Chicago: National Reading Conference.
- Stanovich, K. E. (1986). Matthew effect in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.
- Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Erlbaum.
- Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, 21, 371-380.
- van Kleeck, A., Stahl, S. A., & Bauer, E. B. (Eds.). (2003). *On reading books to children: Parents and teachers*. Mahwah, NJ: Lawrence Erlbaum.
- Varma, S. (2012). Preliminary item statistics using point-biserial correlation and p-

- values. Retrieved from:
http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf
- Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Contributions of the home literacy environment to preschool-aged children's emerging literacy and language skills. *Early Child Development and Care, 176*, 357–378.
 doi:10.1080/03004430500063747.
- Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*(2), 265–279. doi:10.1037/0012-1649.37.2.265
- Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition*. Circle Pines, MN: AGS Publishing.
- Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research, 50*, 1093-1109.
- Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology, 30*, 679-689.
- Zevenbergen, A. A., & Whitehurst, G. J. (2003). Dialogic reading: A shared picture book reading intervention for preschoolers. In A. Van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.). *On reading books to children: Parents and teachers*. Mahwah, NJ: Lawrence Erlbaum.

Appendix A

Book list

Water Cycle

Branley, H. B. (1983). *Rain and Hail*. New York: Harper Collins.

Mason, A. (2006). *Change it: Solids, liquids, gases and you*. Toronto, Canada: Kids Can Press.

Williams, J. (2005). *Why is it raining?* Berkley Heights, NJ: Enslow Elementary.

Animals and Adaptations

Berger, M. & Berger, G. (1995). *What do animals do in winter?* Nashville, TN: Hambleton Hill Publishing, Inc.

Smith, J. (2006). *Animal hide and seek*. New York: DK Publishing.

Sill, C. (1997). *About mammals: A guide for children*. Atlanta, GA: Peachtree Publishers.

Life Cycle

Hariton, A. (1995). *Butterfly Story*. New York: Dutton Children's Books.

Schuh, M. C. (2003). *Butterflies*. Mankato: MN: Capstone Press.

Habitats

Brooks, F. (1999). *The Usborne first encyclopedia of our world*. Tulsa, OK: EDC Publishing.

Bussolati, E. (1990). *My picture word book of people, places, and things*. New York: Derrydale Books.

Gibbons, G. (1998). *Marshes and swamps*. New York: Holiday House.

Simple Machines

Roberts, C. (2007). *Airplanes*. Mankato, MN: The Child's World.

Thompson, G. (2006). Lever, screw, and inclined plane. Washington DC: National Geographic Society.

Welsbacher, A. (2001). Levers. Mankato, MN: Capstone Press.

Earth

Branley, F. M. (1986). What makes day and night? New York: Thomas Y. Crowell Junior Books.

Gibbons, G. (1995). The reasons for the seasons. New York: Holiday House.

Owens, L. L. (2011). Earth. Mankato, MN: Child's World.

Plants

Bauer, J. (2007). Sunflower life cycle. New York: Scholastic, Inc.

Mills, J. E. (2005). Flowers. New York: Scholastic, Inc.

Landforms

Charles, R. (2003). Land and water. Tucson, AZ: Learning Page.

Sellman, J. (n.d.). Natural wonders of the world. Reading A-Z.

Appendix B

Word List by Cluster

<p><u>Water Cycle</u> Liquid Evaporation Precipitation Gas Solid condensation</p>	<p><u>Animals and adaptations</u> Mammals Amphibians Reptiles Hibernate Migrate Camouflage</p>	<p><u>Life Cycle</u> Larva Pupa Cocoon thorax <i>Antenna</i> <i>Insect</i></p>	<p><u>Habitats</u> Prairie Coast marsh Swamp Rainforest estuary <i>Desert</i></p>
<p><u>Simple Machines</u> Propeller Lever Wedge Pulley Fulcrum latch</p>	<p><u>Earth</u> Revolve Equator Atmosphere Axis Rotate <i>Planet</i></p>	<p><u>Plants</u> Pistil Bud pollen sepal <i>blossom</i> <i>root</i></p>	<p><u>Landforms</u> Canyon Cliff Valley gorge Summit plateau</p>

Appendix C

Fidelity Checklist

First Reading	Teacher 1	Teacher 2	Teacher 3
Read through book without providing vocabulary instruction			
Introduction			
Show word cards			
State each target word twice			
Ask students to chorally say each target word twice			
Second Reading			
Stop at end of sentence containing target word			
Partners tell one another the words			
Re-state words and show cards			
Share researcher-provided definitions			
Reread sentence replacing word with definition			
Procedure followed for second word			
After Reading			
Restate words and definitions			
Think-pair-share			

Appendix D
Lesson Plan Template for Small Group Instruction

<i>Sequence</i>	<i>Description of Activities</i>	<i>Comments/Researcher Journal</i>
Introduce/Review (2 minutes) <i>Target Words:</i> _____ _____	<ul style="list-style-type: none"> • Show word cards and orally read words clearly for students • Think-pair-share: Students each share the meaning of one target word with their partner • Two students are randomly selected (switch students each day) to share one of the word meanings for the group 	
Teacher-led instruction (2 minutes)	<ul style="list-style-type: none"> • As words are shared by students, the researcher clarifies the word meaning as needed, by rereading the researcher-developed definition • Show a picture for each target word while clarifying the definition • Point out important features of the pictures 	
Group Discussion (4 minutes)	<ul style="list-style-type: none"> • The researcher leads the discussion with the following (as relevant): <ul style="list-style-type: none"> ○ Personal experience related to a target word ○ Examples of the target word (provide more than one) ○ Non-examples ○ Ask for students comments and related experiences if participation is lacking 	
Activity (5 minutes)	<ul style="list-style-type: none"> • Select from one of the following as applicable: <ul style="list-style-type: none"> ○ Venn Diagram ○ T-chart ○ Semantic feature analysis table ○ Concept sort 	
Closure (2 minutes)	<ul style="list-style-type: none"> • Students write each target word on separate index cards and draw a related sketch. • Students do a quick think-pair-share of the two target word meanings when finished with their writing and sketch. 	

Appendix E

Pilot Study of Assessment Measures

Purpose

The purpose of conducting a pilot study on the researcher-designed assessment measures was to evaluate the reliability and correlation of the assessments using a test, re-test procedure. The measures administered during the pilot included a receptive and expressive vocabulary test.

Procedure

The first measure administered to students during the pilot study, was the expressive vocabulary assessment. The purpose in administering the expressive measure first was to prevent students from hearing the target words stated by the researcher prior to having the opportunity to state the words themselves. For instance, the following protocol was followed for the expressive assessment, “I am going to show you some pictures and I want you to tell me what the picture represents. In other words, tell me what you see in the picture. For example, if I was shown this picture (show the student a picture of an ocean) I would say the word *ocean* because I see the ocean in this picture. Now I want you to tell me what you see in the pictures I show you. Let’s begin. What is this a picture of?” The researcher would continue to show the pictures one at a time and ask “What is this a picture of?”. Depending on the concept or word that the picture represented, such as the picture of a bear hibernating, one of the following questions were asked, “What do you see in this picture?” or “What is the animal doing in this picture?” During this assessment, the individual students were the only ones to state the target words; the researcher did not provide any target words or restate target words for the student during this task.

The second measure administered to students was the receptive vocabulary assessment. During this task, the researcher showed students a grid of four pictures and asked the students to

point to a particular picture. For instance, the researcher gave the following instructions, “Point to the picture of the summit” or “Point to the picture that shows condensation”. During this assessment, it was necessary for the researcher to state the target word when giving the instructions; therefore, the receptive assessment was administered after the expressive assessment so that students would not have prior exposure to the words during the assessment period.

Students were administered each of the two assessments twice, with two weeks in between testing sessions. Students were tested individually, outside of their classroom to minimize possible distractions that would impede their concentration. Each testing period lasted approximately 15 minutes per student. Upon conclusion of the pilot study, after all students had been administered both the receptive and expressive measures as a pre and post test, data were analyzed.

Data Analysis

After obtaining the basic statistic of Cronbach’s Alpha, used to determine overall internal consistency of the data from pre to post test, three levels of analysis were used to evaluate the reliability of the researcher-designed assessments. First Pearson’s correlation coefficients of the summed scores on both tests were calculated to determine if there was significant correlation between students’ overall scores on the pre and post tests. Second, Pearson’s correlation coefficients were calculated for each individual test item for the pre and post tests of both measures. Although Pearson’s correlation coefficients provided adequate information on the relationship between students’ pre and post test answers, further analysis was needed to determine the discrimination between students with varying levels of vocabulary knowledge. Specifically, students with more word knowledge should answer correctly more often than students with less word knowledge. Third, point-biserial correlations (r_{pbi}) were analyzed to

determine the correlation between students' scores on individual items in relation to their overall score. A positive r_{pbi} represents an item in which students that scored high overall also answered the item correctly. A point-biserial value of .15 is acceptable, but .25 or better is desirable (Varma, 2012).

Receptive data analysis. First, an internal consistency estimate of reliability was computed for pretest and posttest scores of the receptive measure assessment. The Cronbach's Alpha reliability coefficient for pretest scores was .755 and .566 for the posttest scores, indicating a moderately reliability assessment.

Second, a Pearson's correlation coefficient of the summed scores was calculated. The correlation of the summed scores of the receptive measure was .691 ($p < .001$). Given this moderately significant correlation coefficient, correlations were then calculated for individual items. The following table shows the correlation coefficients for pre and posttest scores on each individual test item.

Table A1

Correlation coefficients for summed scores of pretest and posttest

Test Item	Correlation Coefficients	Significance Level
1	0.311	0.048
2	-0.036	0.824
3	0.597	.000
4	no variance	
5	no variance	
6	no variance	
7	0.425	0.006
8	0.247	0.12
9	0.648	.000
10	0.218	0.17
11	0.288	0.068

12	0.198	0.214
13	no variance	
14	-0.052	0.747
15	0.477	0.002
16	0.329	0.036
17	0.304	0.054
18	0.729	.000
19	0.46	0.002
20	no variance	
21	no variance	
22	0.37	0.017
23	no variance	
24	0.041	0.801
25	0.669	.000
26	0.408	0.008
27	0.329	0.036
28	no variance	
29	-0.051	0.75
30	0.096	0.549
31	0.307	0.051
32	0.466	0.002
33	0.223	0.161
34	0.372	0.017
35	0.327	0.037
36	0.569	.000
37	0.392	0.011
38	0.415	0.007
39	0.725	.000
40	0.11	0.492
41	no variance	
42	-0.084	0.6
43	1	.000
44	0.149	0.354
45	0.569	.000
46	0.297	0.059
47	0.34	0.029
48	0.503	0.001

The highlighted boxes indicate test items that were not significantly correlated between test one and test two at the .05 level. Alarming, there were 13 out of 48 total items that were

not significantly correlated, in addition to the items that could not be calculated due to the lack of variance. The boxes stating “no variance” means that one or both of the test scores for a particular item were identical, with no variation in the scores; therefore, a correlation coefficient could not be computed for these nine items. The remaining 26 items were significantly correlated, but further analysis was needed to determine precisely why so many items were not.

After analyzing the summed scores, item-level analyses were conducted to determine the reliability and value of each particular test item. As evident from the table above, there were 13 items (numbers: 2, 8, 10, 11, 12, 14, 24, 29, 30, 33, 40, 42, and 44) that were not significantly correlated at the .05 level. Frequencies of these items were examined to determine the number of students that answered correctly versus incorrectly. When analyzing the frequencies, presented in table three below, it appeared that the two variables representing pre and posttest scores for each item were so similar, that one or both acted as a constant and therefore, skewed the Pearson’s correlation coefficient.

Table A2

Frequencies of receptive assessment items

Question	Frequency of Incorrect Responses	Frequency of Correct Responses
Pretest 2	2	39
Posttest 2	1	40
Pretest 8	6	35
Posttest 8	10	31
Pretest 10	23	18
Posttest 10	21	20
Pretest 11	4	37
Posttest 11	7	34
Pretest 12	7	34
Posttest 12	2	39
Pretest 14	4	37

Posttest 14	1	40
Pretest 24	11	30
Posttest 24	10	31
Pretest 29	2	39
Posttest 29	2	39
Pretest 30	4	37
Posttest 30	6	35
Pretest 33	4	37
Posttest 33	3	38
Pretest 40	10	31
Posttest 40	9	32
Pretest 42	5	36
Posttest 42	2	39
Pretest 44	6	35
Posttest 44	3	38

From the frequencies reported above, it became evident that more students scored correctly than incorrectly on 42 out of the 48 test items. These data indicate that students were near ceiling on many test items. In further illustration of this finding, the table below provides the summed scores for each student. As group, students were able to receptively identify the majority of words on the assessment.

Table A3

Student Test Scores on Test 1 and Test 2

Student	Test 1	Test 2
1	45	42
2	40	38
3	37	40
4	35	30
5	37	41
6	36	39
7	39	39
8	36	38
9	35	41
10	32	36
11	39	41

12	36	39
13	37	42
14	38	38
15	43	42
16	43	42
17	41	44
18	28	37
19	36	39
20	37	41
21	39	43
22	45	45
23	40	42
24	43	44
25	33	38
26	36	39
27	44	47
28	37	37
29	40	44
30	38	41
31	43	42
32	27	32
33	39	40
34	40	40
35	35	34
36	31	40
37	25	35
38	27	33
39	44	37
40	40	36
41	32	36

Interestingly, 27 of the 41 students scored higher on pretest than posttest, suggesting the occurrence of a learning effect. This may be attributable to the fact that students had exposure to both a receptive and expressive assessment during the pretest which could have influenced scores on the posttest.

At this point in analysis, the receptive measure was becoming questionable in regard to its reliability and effectiveness of showcasing students' true knowledge of words. Hence, the fourth step in the analysis was to analyze point-biserial correlations, which would reveal the

degree of discrimination of each item. The fourth column in the following tables: Corrected Item Total Correlation presents the point-biserial values for each item. Missing items are those with zero variance. Highlighted items had negative values or positive values less than .15. Table four shows pretest scores and table five shows posttest scores.

Table A4

Point-biserial values for pretest items

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Cronbach's Alpha if Item Deleted
Ques1Pre	37.1220	24.210	.117	.755
Ques2Pre	36.3171	23.722	.462	.746
Ques3Pre	36.6829	22.472	.429	.740
Ques4Pre	36.2683	24.751	.000	.755
Ques5Pre	36.2683	24.751	.000	.755
Ques6Pre	36.2683	24.751	.000	.755
Ques7Pre	36.5366	23.455	.252	.750
Ques8Pre	36.4146	25.149	-.146	.765
Ques9Pre	36.5122	22.756	.435	.741
Ques10Pre	36.8293	22.995	.312	.746
Ques11Pre	36.3659	23.388	.438	.744
Ques12Pre	36.4390	24.602	.001	.760
Ques13Pre	36.2683	24.751	.000	.755
Ques14Pre	36.3659	24.638	.008	.758
Ques15Pre	36.5366	23.055	.347	.745
Ques16Pre	36.6829	24.422	.016	.762
Ques17Pre	36.4878	23.106	.365	.744
Ques18Pre	36.4146	24.149	.135	.754
Ques19Pre	36.4634	23.755	.214	.751
Ques20Pre	36.2927	24.162	.368	.750
Ques21Pre	36.2927	24.762	-.023	.757
Ques22Pre	36.4390	22.952	.453	.741
Ques23Pre	36.2927	24.762	-.023	.757
Ques24Pre	36.5366	24.405	.033	.760

Ques25Pre	37.0244	23.024	.369	.744
Ques26Pre	36.5122	22.506	.498	.738
Ques27Pre	36.3659	24.588	.025	.757
Ques28Pre	36.2927	24.162	.368	.750
Ques29Pre	36.3171	23.972	.343	.749
Ques30Pre	36.3659	24.388	.092	.755
Ques31Pre	36.3659	23.588	.368	.746
Ques32Pre	36.7561	24.639	-.029	.765
Ques33Pre	36.3659	23.638	.350	.747
Ques34Pre	36.5366	22.005	.605	.732
Ques35Pre	36.6585	23.530	.204	.752
Ques36Pre	36.6341	22.938	.337	.745
Ques37Pre	36.8293	24.945	-.089	.768
Ques38Pre	36.5854	22.899	.362	.744
Ques39Pre	36.3659	23.138	.527	.741
Ques40Pre	36.5122	23.656	.214	.751
Ques41Pre	36.2683	24.751	.000	.755
Ques42Pre	36.3902	23.994	.200	.752
Ques43Pre	36.3659	24.238	.143	.754
Ques44Pre	36.4146	23.399	.354	.746
Ques45Pre	36.4634	24.405	.047	.758
Ques46Pre	37.0244	23.424	.270	.749
Quest47Pre	36.5610	23.802	.164	.754
Ques48Pre	36.5366	23.455	.252	.750

Table A5

Point-biserial values for posttest items

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Cronbach's Alpha if Item Deleted
Ques1Post	39.0488	12.348	.081	.566
Ques2Post	38.3902	12.844	-.027	.568
Ques3Post	38.6829	11.472	.358	.531
Ques4Post	38.3902	12.944	-.116	.572
Ques5Post	38.4146	12.799	-.005	.568
Ques6Post	38.3902	12.794	.017	.566
Ques7Post	38.6098	11.294	.464	.521
Ques8Post	38.6098	12.944	-.094	.585

Ques9Post	38.5366	11.955	.280	.545
Ques10Post	38.8780	11.710	.252	.544
Ques11Post	38.5366	11.455	.480	.524
Ques12Post	38.4146	13.049	-.164	.577
Ques13Post	38.4390	12.702	.035	.567
Ques14Post	38.3902	12.994	-.160	.573
Ques15Post	38.5854	12.099	.193	.553
Ques16Post	38.6098	11.744	.304	.540
Ques17Post	38.4390	12.402	.197	.556
Ques18Post	38.5366	11.855	.319	.541
Ques19Post	38.4146	12.899	-.069	.572
Ques20Post	38.3659	12.838	.000	.566
Ques21Post	38.3659	12.838	.000	.566
Ques22Post	38.4390	12.402	.197	.556
Ques23Post	38.3659	12.838	.000	.566
Ques24Post	38.6098	13.044	-.126	.588
Ques25Post	39.0244	12.124	.144	.559
Ques26Post	38.5122	12.556	.061	.566
Ques27Post	38.5122	12.756	-.018	.573
Ques28Post	38.3659	12.838	.000	.566
Ques29Post	38.4146	12.799	-.005	.568
Ques30Post	38.5122	13.106	-.153	.585
Ques31Post	38.4146	12.849	-.037	.570
Ques32Post	38.7805	13.026	-.121	.592
Ques33Post	38.4390	12.752	.008	.568
Ques34Post	38.5122	11.806	.367	.537
Ques35Post	38.7317	13.451	-.238	.605
Ques36Post	38.6341	12.588	.015	.573
Ques37Post	38.6829	12.672	-.017	.578
Ques38Post	38.6341	11.438	.395	.528
Ques39Post	38.5366	11.705	.379	.535
Ques40Post	38.5854	11.449	.428	.526
Ques41Post	38.3902	12.794	.017	.566
Ques42Post	38.4146	12.499	.189	.557
Ques43Post	38.4634	12.205	.259	.550
Ques44Post	38.4390	12.802	-.018	.570
Ques45Post	38.4878	12.656	.031	.568
Ques46Post	39.0976	11.940	.225	.549
Ques47Post	38.5122	12.106	.242	.549
Ques48Post	38.6341	11.088	.518	.513

Evident from the table above, is that many items on the receptive measure were non-discriminating based on students' knowledge and overall scores. A review of the final column in the above table was completed to determine if deletion of particular items would result in an increased overall reliability score. On test one, items: 8, 12, 16, 24, 31, and 37 would all contribute to an increase in Cronbach's alpha if deleted; however, the increase in Cronbach's alpha would be less than .10. On test two, items: 2, 4, 5, 8, 12, 13, 14, 24, 27, 30, 31, 32, 33, 35, 36, 37, 44, and 45 would contribute to an increase in Cronbach's alpha if deleted; however, deleting in most of these items would result in no more than a .10 increase.

In sum, it is evident that this measure is not highly reliable under a test-retest condition and should be excluded from administration in the proposed dissertation study.

Expressive assessment data analysis. Similar data analysis was conducted for the expressive data as for the receptive data. First, internal consistency estimates of reliability were computed for the test one and test two scores of the expressive measure assessment. With a Cronbach's Alpha reliability coefficient of .799 at pretest and a Cronbach's Alpha reliability coefficient of .854 at posttest, the expressive measure is considered highly reliable. The test-retest reliability, calculated by using Pearson's correlation coefficient of the summed scores is .912 ($p < .001$), also suggesting that the pre and posttest data sets are highly correlated. Although the summed scores of pretest and posttest were significantly correlated, correlation coefficients were additionally calculated for individual test items. Table six below shows the correlation coefficients for the test items.

Table A6

Correlation Coefficients for test items

Item Number	Correlation	Significance
1	.297	.05
2	no variance	
3	-.051	.75
4	no variance	
5	.625	.000
6	.520	.000
7	.620	.000
8	no variance	
9	.470	.002
10	.437	.004
11	.563	.000
12	.563	.000
13	.460	.002
14	.584	.000
15	.169	.291
16	.371	.017
17	.506	.001
18	.464	.002
19	no variance	
20	.169	.291
21	.613	.000
22	.527	.000
23	.608	.000
24	.434	.005
25	no variance	
26	.481	.001
27	.735	.000
28	-.025	.877
29	.594	.000
30	.491	.001
31	no variance	
32	.613	.000
33	no variance	
34	.169	.291
35	.481	.001
36	.679	.000
37	.867	.000
38	.481	.000

39	.468	.002
40	.858	.000
41	.749	.000
42	no variance	
43	.223	.161
44	no variance	
45	.399	.010
46	.729	.000
47	.531	.000
48	.853	.000

In contrast to the large number of items not significantly correlated on the receptive measure, only six highlighted boxes above represent test items on the expressive measure that were not significantly correlated at the .05 level. The boxes stating “no variance” means that the pre and posttest scores were identical for these items. When one or more variables are constant, such as if all students responded correctly or incorrectly on a particular item, the correlation cannot be computed.

Overall, there were six items (numbers: 3, 15, 20, 28, 34, and 43) that were not significantly correlated at the .05 level. Frequencies were examined to determine the number of students that answered correctly versus incorrectly for these particular items. Similar to the receptive analysis, when analyzing the frequencies it appeared that the two variables representing pre and posttest scores for each item were so similar in nature, that one or both acted as a constant and therefore, skewed the Pearson’s correlation coefficient. Specifically, Table seven shows the frequencies of incorrect (0) and correct (1) responses on pre and posttests for the six items that were not significantly correlated. The below tables show the frequency counts for each item that was not significantly correlated.

Table A7

Frequencies for expressive assessment items

Question	Frequency 0 (incorrect)	Frequency 1 (correct)
Pretest 3	2	39
Posttest 3	2	39
Pretest 15	4	37
Posttest 15	4	37
Pretest 20	4	37
Posttest 20	4	37
Pretest 28	1	40
Posttest 28	1	40
Pretest 34	37	4
Posttest 34	37	4
Pretest 43	3	38
Posttest 43	4	37

More students appeared to answer correctly more frequently than incorrectly on five of the six items: 3, 15, 20, 28, and 43, indicating that most students knew these words. The words were: planet, antenna, insect, root, and desert. Word number 34, blossom, was not highly correlated between pre and post test, but the majority of students did not know this word. When analyzing the data qualitatively, it became apparent that although students did not use the word blossom, the majority responded with “bloom”, signifying that although the majority of students were unable to orally produce the word *blossom*, they were familiar with the concept of an open flower bud, which had bloomed. The data were also qualitatively analyzed by revisiting the receptive scores for these five words, to determine whether students knew the words both receptive and expressively, or just expressively. For the word *desert*, all students identified it correctly during pretest and 40/41 students identified it correctly on posttest. The word *root*, was

known receptively by all students on pretest and by 39/41 students on posttest. For the word *insect*, 35 students identified the word receptively during pretest and 34 students identified it during posttest. The word *planet* was identified by 40 students on pretest and by 41 students on posttest. Finally, 40 students identified the word *antenna* correctly on pretest and 41 students identified it correctly on posttest. Therefore, students know the words: desert, root, insect, planet, and antenna well, both receptively and expressively. These are words that may need to be deleted on the revised assessment, as students have already reached a ceiling.

Given that the frequencies of the non-correlated items indicated students had reached a ceiling on five particular words, the data analysis thus far led to the question of whether or not students knew the majority of the words on the assessment, as was the case with much of the receptive data. Hence, each student's pre and posttest summed scores were analyzed to determine if students were answering correctly or incorrectly the majority of the time. The following table shows data on each student's summed scores for pre and posttest.

Table A8

Pre and posttest summed scores by student

Student	Pretest	Posttest
1	20	26
2	14	16
3	9	17
4	15	19
5	20	23
6	16	20
7	15	16
8	17	21
9	7	12
10	11	14
11	15	24
12	8	13

13	21	25
14	15	15
15	21	27
16	23	34
17	19	27
18	12	16
19	19	24
20	13	19
21	24	27
22	20	21
23	20	26
24	19	23
25	12	15
26	12	17
27	21	24
28	13	18
29	25	32
30	17	24
31	10	11
32	14	15
33	15	18
34	17	22
35	13	15
36	14	19
37	7	7
38	7	6
39	26	30
40	14	15
41	12	8

Only 11 out of the 41 students assessed knew roughly 50% of the words on posttest. No student came close to reaching a ceiling on the expressive vocabulary test as a whole; therefore, the test contains words that most students at the end of second grade do not know expressively.

Finally, point-biserial correlations were analyzed to determine the degree of discrimination of each item. The fourth column in table nine below: Corrected Item Total Correlation represents the point-biserial values for each item. Missing items are those with zero variance. Highlighted items had negative values or positive values less than .15.

Table A9

Point-biserial values for pretest items

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Pre1	16.1220	26.060	.211	.797
Pre2	16.8537	27.228	.000	.799
Pre3	15.9024	26.840	.151	.798
Pre4	16.7317	27.801	-.195	.808
Pre5	16.4146	25.649	.261	.795
Pre6	16.2195	25.376	.329	.792
Pre7	16.7561	25.789	.442	.790
Pre8	16.8537	27.228	.000	.799
Pre9	16.4146	24.799	.435	.788
Pre10	16.5366	24.505	.536	.783
Pre11	16.8293	27.045	.098	.798
Pre12	16.8293	26.695	.315	.795
Pre13	16.8049	26.811	.164	.797
Pre14	16.5122	25.706	.265	.795
Pre15	15.9512	27.248	-.035	.803
Pre16	16.8049	26.511	.298	.795
Pre17	16.7561	25.789	.442	.790
Pre18	16.7805	26.626	.196	.797
Pre19	16.8537	27.228	.000	.799
Pre20	15.9512	26.348	.256	.795
Pre21	16.3659	25.588	.270	.795
Pre22	16.6829	25.122	.514	.786
Pre23	16.4390	25.002	.396	.789
Pre24	16.4878	25.606	.280	.794
Pre25	16.8537	27.228	.000	.799
Pre26	16.8293	27.345	-.087	.801
Pre27	16.0976	25.490	.353	.791
Pre28	15.8780	27.110	.058	.799
Pre29	16.2683	26.501	.093	.802
Pre30	16.4390	24.952	.407	.789
Pre31	16.8537	27.228	.000	.799
Pre32	16.3659	24.538	.486	.785
Pre33	16.8537	27.228	.000	.799
Pre34	16.7561	27.039	.032	.801

Pre35	16.8293	26.695	.315	.795
Pre36	16.4878	24.806	.450	.787
Pre37	16.0976	24.740	.532	.784
Pre38	16.8293	26.845	.222	.797
Pre39	16.7805	26.976	.067	.800
Pre40	16.0732	25.870	.277	.794
Pre41	16.6829	25.822	.326	.793
Pre42	15.8780	27.060	.089	.799
Pre43	15.9268	27.120	.014	.801
Pre44	16.8537	27.228	.000	.799
Pre45	16.2927	25.662	.258	.795
Pre46	16.7073	25.662	.397	.790
Pre47	16.2439	24.789	.446	.787
Pre48	16.3902	25.894	.210	.798

Table A10

Point-biserial values for posttest items

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Post1	18.7805	39.526	.529	.846
Post2	19.4878	42.506	.018	.855
Post3	18.5854	42.199	.127	.853
Post4	19.2927	41.862	.098	.856
Post5	19.0000	39.800	.400	.849
Post6	18.7317	40.251	.431	.848
Post7	19.3171	40.222	.415	.848
Post8	19.5122	41.906	.334	.852
Post9	19.0244	39.274	.485	.846
Post10	19.2195	39.876	.421	.848
Post11	19.4634	42.305	.067	.854
Post12	19.4634	40.805	.514	.848
Post13	19.3415	41.680	.147	.854
Post14	19.1463	40.278	.332	.850
Post15	18.6341	40.988	.397	.850
Post16	19.4634	41.855	.199	.853
Post17	19.3659	41.638	.167	.854
Post18	19.2683	39.901	.442	.848

Post19	19.5122	42.956	-.183	.856
Post20	18.6341	40.688	.477	.848
Post21	19.0000	40.100	.352	.850
Post22	19.2195	39.526	.482	.847
Post23	19.0732	39.770	.405	.849
Post24	18.9512	40.248	.333	.850
Post25	19.4878	41.656	.320	.851
Post26	19.4390	42.402	.029	.855
Post27	18.7805	39.326	.567	.845
Post28	18.5610	42.452	.063	.854
Post29	18.9024	40.990	.220	.853
Post30	19.1707	41.745	.099	.856
Post31	19.5366	42.605	.000	.854
Post32	19.0000	39.300	.482	.846
Post33	19.5122	42.806	-.110	.855
Post34	19.4390	42.402	.029	.855
Post35	19.4390	41.502	.262	.852
Post36	19.1220	39.510	.454	.847
Post37	18.7317	39.451	.594	.845
Post38	19.4390	41.152	.353	.850
Post39	19.4146	42.349	.034	.856
Post40	18.7561	39.439	.568	.845
Post41	19.2683	40.501	.333	.850
Post42	18.5366	42.605	.000	.854
Post43	18.6341	41.938	.148	.853
Post44	19.4878	42.506	.018	.855
Post45	18.8780	39.910	.406	.849
Post46	19.2927	39.812	.475	.847
Post47	18.8049	39.911	.440	.848
Post48	19.0976	40.390	.307	.851

Although there was evidence of non-discrimination on multiple items, the final column of the above table was also reviewed to determine whether removal of these items would increase the overall reliability. The final column in the above two tables shows the overall reliability value with each item deleted. The only two items that could be deleted and contribute to an increased Cronbach's alpha would be items four on pre and posttest test and item 15 on pretest.

These two items were qualitatively analyzed to determine if the pictures used on the assessment measure were confusing, misleading or contained features that could skew students' responses. Item number four, *rotate*, is represented by a picture of a blue circle with the axis line in the middle and arrows indicating rotation. Student responses on this particular item show that 20/41 students stated "I don't know" or gave no response, 8/41 students stated "spinning", "turning" or "circulating" and the remaining 13/41 students stated either, "night and day", "orbiting" or "solar". Hence, this word represents one that students do not seem to know expressively and that there is confusion between the terms: "*rotate*" and "orbit"; however, revising the picture may provide students with a clearer representation of what it is they should be focusing on. Word 15, antenna, was simply too easy for the group of students tested and the majority could expressively state the word represented by the picture. Only seven students were unable to expressively identify the word *antenna*. Hence, this word, along with the other "known" words (planets, insect, root, blossom, and desert) will be replaced.

Conclusions

The pilot study analyses were in-depth and thorough. The analysis progressed from an overall Cronbach's Alpha to item-level correlation analyses. The receptive data simply was not as clean as the expressive data. There was a learning effect, which took place from pre to posttest on the receptive test that was not evident to the same degree on the expressive test. The correlation at the summed score level was only .69 for the receptive test, while the correlation was .91 for the expressive test. Hence, given this critical information, the measure which will be most useful and effective in providing relevant data for the actual dissertation study is the expressive vocabulary test. An expressive vocabulary test is more demanding than a receptive test and is more accurate at truly exhibiting the depth of knowledge students have about words.

On a receptive test, students are provided four possible answers and therefore, only have to eliminate three in order to figure out the correct answer. On the expressive test, students have to use their own knowledge of what is represented by a picture and orally state the corresponding word. Given that one element of the present study is centered on fostering and measuring student growth in expressive vocabulary knowledge, the expressive measure will be important to administer and will be used as the study measure for this research.

Appendix F

Expressive Target Word Assessment (TWA)



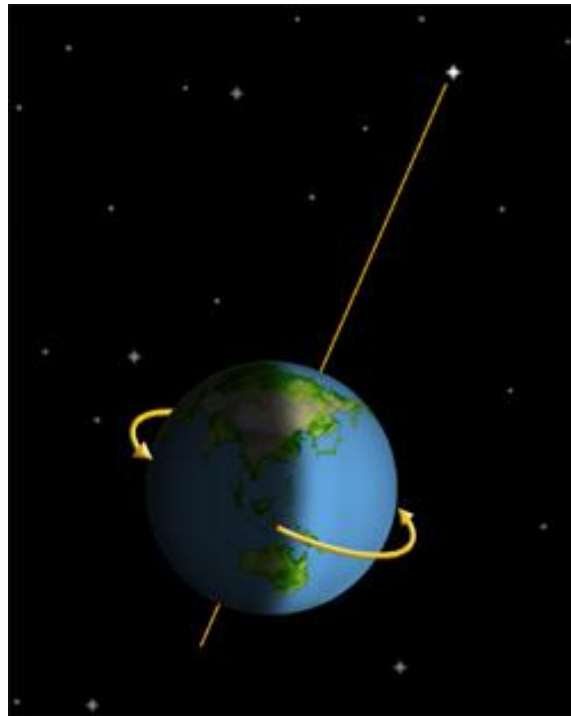
www.salmanspets.wordpress.com, www.en.wikipedia.org, www.swfsc.noaa.gov,
<http://en.wikipedia.org/wiki/Squirrel>



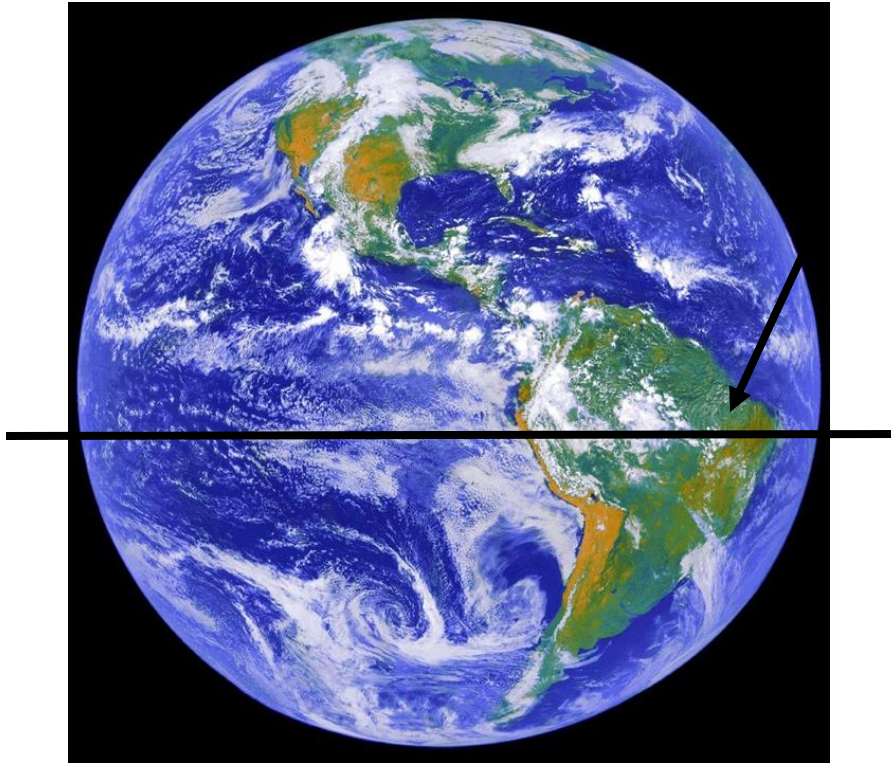
www.merriam-webster.com



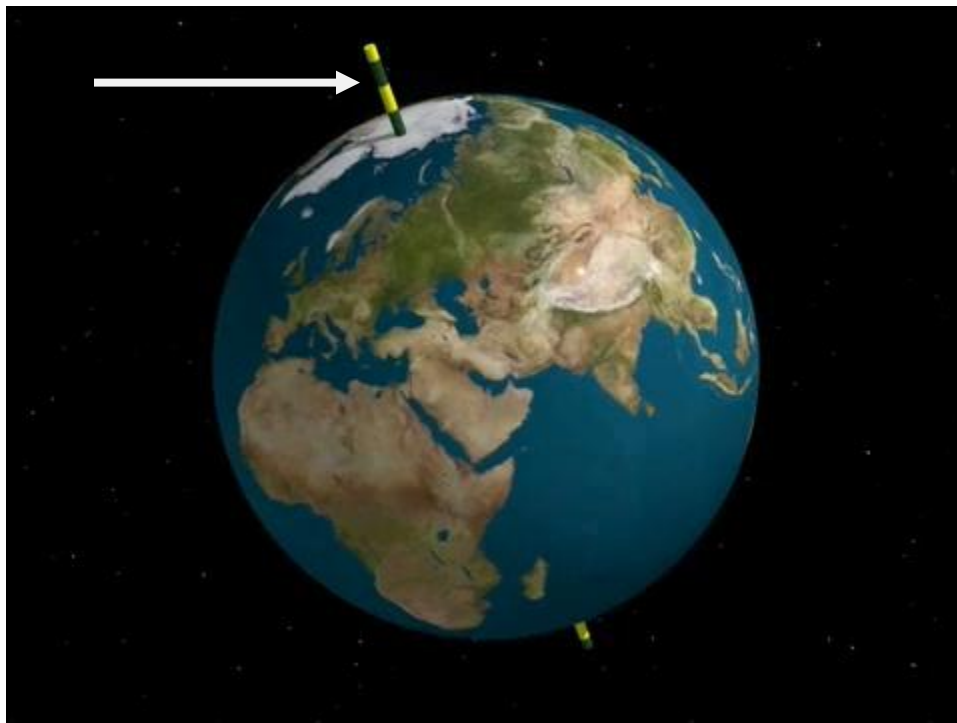
www.oceana.org



www.lpi.usra.edu



www.teacher.scholastic.com



www.commonswiki.org



www.pd.scisdragons.net



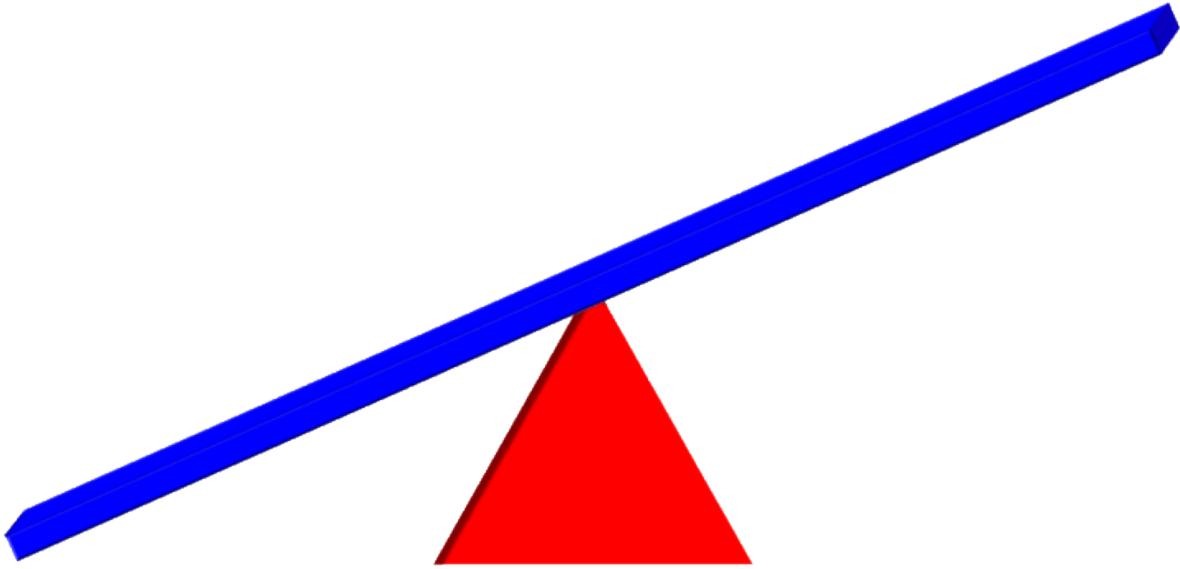
www.cimss.ssec.wisc.edu



www.commonswikimedia.com



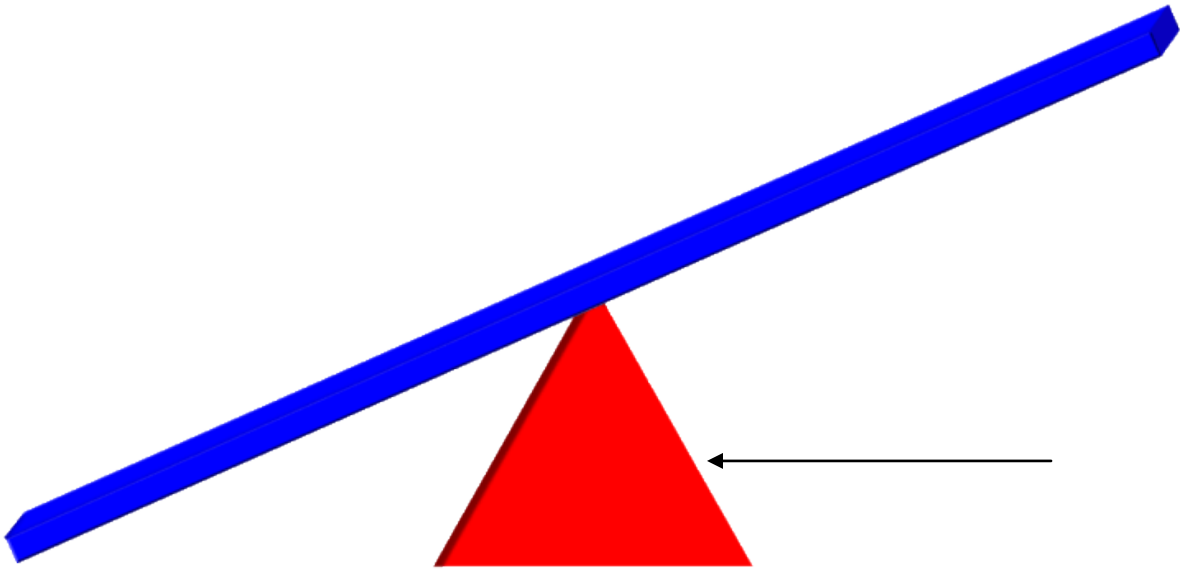
www.learnersdictionary.com



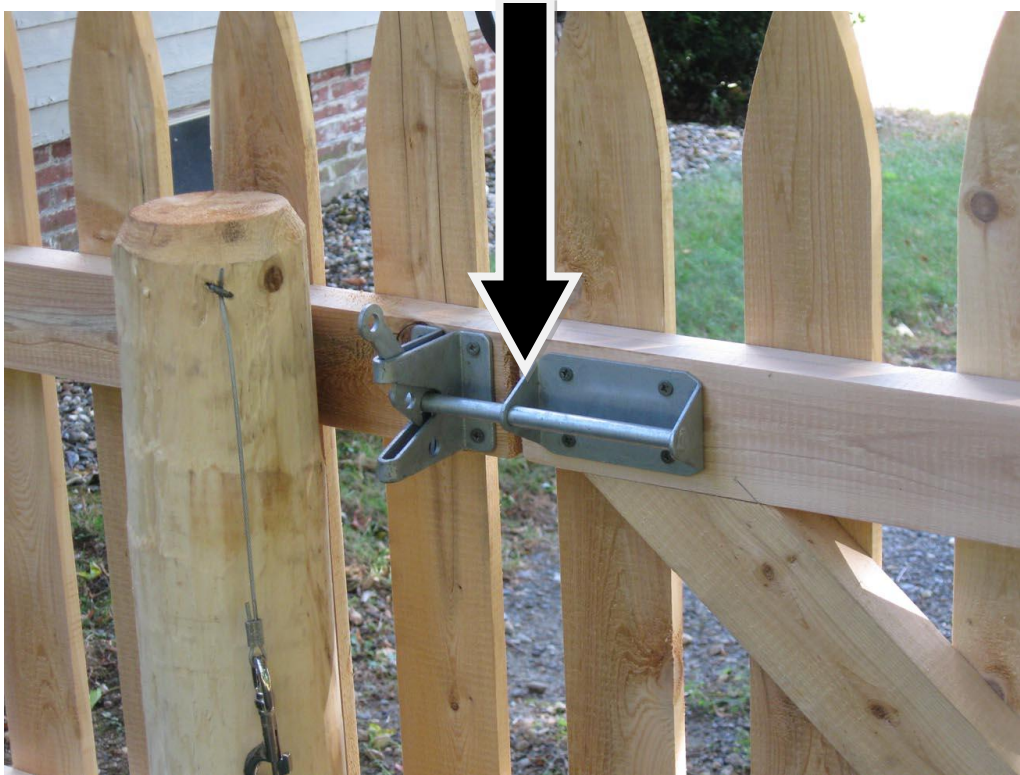
www.makeitgo.wikispaces.com



www.adf-safetytools.com



www.makeitgo.wikispaces.com



www.whitneyfence.com



www.ecolibrary.com



www.blog.growingwithscience.com



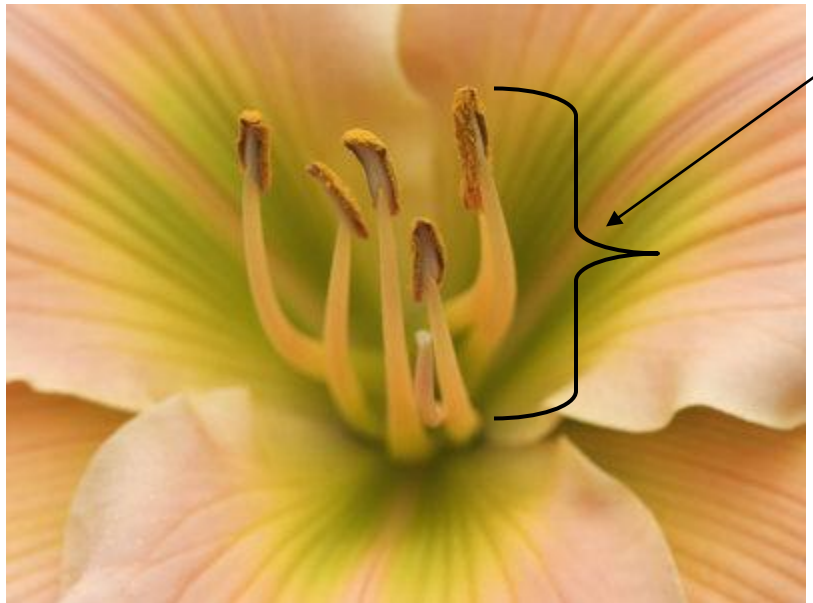
www.maria.fremlin.de



www.commonswiki.org



www.flickr.com



www.art.com



www.bio.tamu.edu



www.wakpaper.com



www.destinations360.com



www.nice-cool-pics.com



www.tulane.edu



www.science.kennesaw.edu



www.boards.elsaelse.com



www.funfantasticfacts.blogspot.com



www.alteranchopet.com



www.cedarcreekcabinrentals.com



www.theproducerperspective.com



www.sickrose.wordpress.com



www.himalman.wordpress.com



www.buzzle.com



www.gorell.com



www.layoutsparks.com
www.xweather.com



www.kids.britannica.com



www.techcrunch.com



www.123rf.com



www.poemsfromaspiritualheart.blogspot.com



www.howstuffworks.com



www.centralexterminatingco.com



www.classbb.wordpress.com

Appendix G

Word List with Biemiller's *Words Worth Teaching* Classification

Word	Classification
Liquid	L2
Evaporation	*
Precipitation	T6
Gas	E
Solid	E
Condensation	*
Propeller	T2
Lever	T6
Wedge	D
Pulley	T2
Fulcrum	T6
Latch	T2
Mammals	T2
Amphibians	T6
Reptiles	T2
Hibernate	T2
Migrate	T6
Camouflage	T2
Revolve	T6
Equator	T2
Atmosphere	D
Axis	D
Planets	E
Orbit	T6
Rotate	T6
Larva	T6
Pupa	*
Antenna	L2
Cocoon	L2
Insect	E
Thorax	*
Pistil	*
Bud	E
Pollen	T2
Sepal	*
Root	E
Blossom	E
Prairie	T6
Desert	T2
Estuary	*

Coast	D
Marsh	T6
Swamp	D
Rainforest	*
Canyon	E
Cliff	T6
Valley	E
Gorge	D
Summit	D
Plateau	D

Biemiller's (2010) classification system for words worth teaching are grouped using the following characteristics:

"T2: High-priority words known by 40% to 80% of children by the end of grade two.

T6: High-priority words known by 40% to 80% of children by the end of grade six

D: Words known by fewer than 40% of children by the end of grade six

E-Words known by most children at the end of grade two."

*Words not included in Biemiller's (2010) word lists

Appendix H

Standard Scores and Composites of Participants' Standardized Assessments

Table A11

Composite Scores from PPVT and EVT

Participants	PPVT	EVT	Composite Score
1	120	112	116
2	113	99	106
3	91	84	87.5
4	113	112	112.5
5	99	95	97
6	114	109	111.5
7	113	114	113.5
8	121	115	118
9	109	87	98
10	92	92	92
11	88	94	91
12	112	90	101
13	108	99	103.5
14	111	109	110
15	128	94	111
16	117	107	112
17	100	90	95
18	87	83	85
19	100	80	90
20	106	106	106
21	114	104	109
22	124	120	122
23	88	95	91.5
26	121	112	116.5
29	114	109	111.5
30	100	99	99.5
31	116	88	102
32	119	109	114
33	109	100	104.5
34	110	114	112
35	95	89	92
36	108	94	101
37	109	88	98.5
38	109	106	107.5
39	93	80	86.5
40	124	119	121.5
41	87	97	92

42	99	90	94.5
43	101	97	99
44	121	112	116.5
45	109	104	106.5
46	116	112	114

Appendix I

Anecdotal Information from Classroom Teachers

Both teachers who participated in the read aloud condition of the study met with the researcher one week after the delayed posttest had been administered. The purpose of the meeting was to informally discuss attributes of the study that the teachers felt were applicable to their own classrooms after the study and to also discuss attributes of the study that were potentially problematic for teachers. The discussion was very open-ended in nature, so as to allow for open discussion between the teachers on their overall feelings, opinions and perceptions of the study procedures and routines, as well as, the applicability of the study procedures to regular classroom instruction. Two very general themes emerged from the discussion: read alouds and small group instruction

Theme One: Read Alouds. The first theme that emerged during the discussion was that teachers had developed interest in certain components of the read aloud portion of the intervention. The teachers expressed their beliefs in the importance of introducing target words that related to science and social studies curriculums during the read aloud. For example, Ms. Jones expressed enthusiasm at the amount of "...exposure and background knowledge being built". Although many of the words in this particular study represented words from other grade level curriculums, teachers still deemed the target words important, as exposure to the words would help prepare students for future instruction on state standards.

Teachers also seemed to recognize the importance of the word cards and accompanying pictures and definitions used when discussing target words during read alouds. Teachers liked having the script of word definitions on the back of the word cards, so that they could hold the word card and picture card for students to see while reading and the explaining the word definitions. The teachers also referenced the picture cards as very useful during the reading of

target word explanations. Teachers could hold the word card in one hand and the picture card in the other, creating a visual link between the word and picture. Mrs. Stone stated, “Usually when we read there isn’t a focus on a word with an accompanying picture and the aspects of the pictures in the books don’t focus on the target words.” Although inclusion of the word cards with definitions and pictures in this study was consistently agreed upon as one of the most important and useful components of the read aloud, both teachers agreed that the time needed to develop and create the cards, definitions and pictures was possibly too great to be done without assistance from another individual. Teachers both agreed that if word cards, definitions, and pictures could be provided for read alouds then this element of the study would be applicable to read aloud-based instruction and manageable for teachers.

A second dimension of the read aloud theme was that both teachers recognized the importance of incorporating nonfiction books into read alouds, especially nonfiction books that supported content area curriculums. Mrs. Stone stated, “Sometimes we talk about key words, like sight words for the week that we call attention to, but not much vocabulary goes on unless it is a word that [the kids] don’t understand.” Up until the study, teachers reported using few nonfiction books during read alouds, although they reportedly read aloud nonfiction during science and social studies instruction as needed; however, after the study, both teachers expressed interest in using nonfiction books during read aloud times to further support their content curriculum instruction.

Theme 2: Small Groups. The small group vocabulary instruction for this particular study was provided by the researcher. Teachers were not responsible for implementing any portion of the small group instruction themselves, but were made aware of the level and intensity of instruction provided during the groups. Teachers felt that the small group instruction was a

beneficial aspect of vocabulary teaching, but without additional teachers to lead the groups, they agreed that this element of the study would not be manageable in the classroom setting, at least during literacy rotations when guided reading was taught. Both teachers indicated feelings of concern about managing the planning and development of materials for the small group instruction, in addition to the guided reading groups they were already holding during the literacy block. Teachers suggested that if a student teacher, blocking student, resource teacher, reading tutor, or paraprofessional were available to run the small intervention groups then it would be feasible to implement into literacy rotations as a center.