Analysis of Two Randomized Field Trials Testing the Effects of Online Vocabulary Instruction on Vocabulary Test Scores

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

CHARLES NORMAN FEHR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

Dr. Mark L. Davison, Dr. S. Jay Samuels, Advisors

September, 2011

Acknowledgements

When I entered this graduate program five years ago, I did so equipped with the knowledge that the quality of my experience would be largely dependent upon the quality of the people on my committee. That knowledge served me well, and I'd like to thank the four generous and thoughtful scholars who served as my committee. Kristen McMaster and Karen Mesce, both of whom I actively sought out as committee members, were nothing but helpful and professional, and it has been a pleasure getting to know them a bit over the last few years. Thanks to Jay Samuels, who routinely offered direct, practical advice, and whose wealth of experience provided a source of knowledge and big-picture perspective that was so refreshing to me when I was buried in the day-to-day challenges inherent in any doctoral program. Most notably, I would like to thank Mark Davison, who took his position as Director of the MITER program seriously enough to notice that I needed an occasional push in the right direction. Mark was instrumental in helping me find my research focus, and then came along for the ride. His thoughtful mentoring, consistently professional disposition, and good humor are largely responsible for my success in this endeavor. I'd like to thank the people at Seward Inc., Greg Sales and Matt Finholt in particular, who invited me to center my research on *The First 4000 Words* and were always so responsive to my requests for data and access to their work. I wish them well with their ongoing efforts. Thanks also to MITER coordinator Peggy Ferdinand, Sharon Sawyer and the professionals who staff the Educational Psychology offices.

On a more personal note, I'd like to thank my family and friends, who have been unwavering in their support of my wandering ways and academic forays for so many years. They probably think that this is finally it....

Most importantly, thank you to my beautiful wife Tina, who has smoothly endured the clutter and roller coaster during the past five years and has been my inspiration. You are the one I live with, laugh with, and love! Let's hop a train.

Abstract

Learning to read requires knowledge of word meanings for those words most commonly encountered in basic reading materials. Many young students lack the basic vocabulary knowledge needed to facilitate learning to read. Two randomized studies were conducted to test the effects of an online, computer-adaptive vocabulary instruction program designed to provide remedial instruction on word meanings for high frequency words. Study 1 was small in scope (N = 43) and tested whether the program could improve word knowledge on a corpus of 100 target words taught to all students in the treatment group. Study 2 was larger in scope (N = 192) and tested whether more extensive use of the computer-adaptive program, which teaches students individualized sets of words from a 4000 word corpus, could improve vocabulary test scores. Scaling up from 100 words in Study 1 to 4000 words in Study 2 necessarily corresponded to a proportionally equivalent decrease in posttest sensitivity to changes in students' vocabulary knowledge. It is argued that such a decrease in standardized test sensitivity requires post-intervention analyses to be conducted at the item-level rather than the posttest total score level. These studies suggest that computer-delivered vocabulary instruction may be an efficient mechanism for remediation of vocabulary deficits. Assessment of post-intervention results at the item-level may be appropriate in other attempts to scale up curricula from pilot studies to classroom use.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

Learning to read and comprehend written materials is a primary challenge children face during their early years in educational settings. Successfully meeting this challenge is prerequisite for success in virtually all subsequent areas of academic endeavor. As such, tremendous emphasis has been placed on teaching children to read during the elementary school years. Despite these efforts many children still fail to gain the reading comprehension skills required to support academic success. The importance of addressing this problem is reflected in the money and time spent on reading research and intervention approaches by both private and public institutions. The National Reading Panel (2000) and the RAND Reading Study Group (2002) represent two significant efforts to gather together leading scholars, researchers, and practitioners to provide comprehensive assessments of this problem and identify mechanisms that might yield better outcomes for students struggling to learn to read. Both of these groups concluded that focus on vocabulary knowledge development should be a component of any comprehensive plan to increase students' reading comprehension achievement. This manuscript details the results of two experiments designed to test whether a novel, online vocabulary instruction program might effectively increase the vocabulary performance of students with low vocabulary knowledge.

The intrinsic importance of vocabulary knowledge for successful reading comprehension is intuitive; a reader must have an understanding of at least most of the words in a reading passage to gain meaning from it. Reading comprehension requires overlapping and concurrent engagement in a number of component sub-skills (e.g., both low- level processes such as decoding and higher-level processes such as integration of meaning). Reading comprehension is facilitated when lower level processing is accomplished in a relatively effortless, automatic fashion (Laberge & Samuels, 1974). For example, if a student's efforts are directed primarily toward decoding (mapping letters to sounds), this will detract from cognitive resources that might be utilized for sentence level processing (e.g., syntax), or passage comprehension. When, through practice and repetition, decoding becomes automatic, more cognitive and attentional resources are available for higher-level processing. Vocabulary knowledge is central to this process. If a reader is faced with a large number of unfamiliar words in a sentence, the attentional and cognitive demands of determining meaning for those words will undermine construction of a coherent semantic representation of the passage. Conversely, if a reader possesses vocabulary knowledge of all or most words in a sentence he or she is more able to automatically activate meaning for those words and incorporate them into a semantic representation of the text, leading to greater comprehension. It is therefore reasonable to infer that one important strategy for increasing reading achievement is to help children increase their functional reading vocabularies. The National Reading Panel (2000) identified vocabulary as a fundamental component of reading comprehension and suggested that vocabulary instruction is necessary to improve students' reading comprehension achievement. As U.S. students are increasingly outperformed by students in other developed countries on reading measures, the RAND Reading Study Group (2002) also argued that vocabulary instruction should be a focus from pre-school through the elementary years. Moreover, the RAND group identified vocabulary knowledge as a "critically important determinant" (p. 43) of reading comprehension.

It is not surprising that children enter educational settings with variability in the extent of their vocabulary development. However, the degree to which particular identifiable groups (e.g., children from low socioeconomic status (SES) backgrounds) enter schooling behind their peers in vocabulary knowledge is cause for serious concern. In their landmark study, Hart & Risley (1995; 2003) found that children in high SES homes acquired expressive vocabularies of greater than a 1000 words by age three, while the low SES children acquired less than half that amount. A follow-up study showed that differences in vocabulary between these higher and lower SES children continued to increase. Hart & Risley estimated a 6000 word gap by the age of six.

Another important and rapidly growing group in the U.S. that can be identified as at-risk in English vocabulary development is comprised of those students learning English as a second language (ELL). Approximately one in nine K-12 students in the U.S. is ELL, and this ratio is predicted to grow to nearly one in three by 2015 (Francis & Vaughn, 2009). Children from homes wherein English is not the primary language spoken in the home are likely to lag in English vocabulary development, relative to peers exposed regularly to English at home. This puts ELL students at risk for failure in educational settings. ELL students' lack of adequate vocabulary may contribute to below grade-level reading comprehension achievement and lower standardized test scores, thus making placement in special education more likely (August, Carlo, Dressler & Snow, 2005). Compounding the problem for many ELL students is that many are also lower

SES students, and therefore face challenges in that regard as well. ELL learners encompass a tremendous variety of students. Most ELL students in the U.S. have Hispanic base languages, but ELL students in the U.S. also come from a wide variety of other linguistic backgrounds. The different orthographies, phonemic constraints, and lexical compositions of some native languages place greater challenges on some groups relative to others in English vocabulary development.

Many young students, particularly those who are English language learners and/or come from impoverished environments, are lacking in English vocabulary knowledge relative to their grade level peers (Hart & Risley, 2003; Lee & Burkam, 2002). Effective remediation of vocabulary deficits can be confounded by the wide range and varying profiles of children's oral and/or reading vocabularies. Thus, individualized training may be necessary to get all children to grade level proficiency. Because individualized vocabulary instruction may not be feasible for single classroom teachers, computer-adaptive technology may be a useful strategy for efficiently addressing individual differences in students' need for vocabulary instruction. The National Reading Panel Report (2000) states, "The use of computers in vocabulary instruction was found to be more effective than some traditional methods… It is clearly emerging as a potentially valuable aid to classroom teachers in the area of vocabulary instruction" (p. 14) and "computer technology must be examined for its ability to deliver instruction, for example, in vocabulary…" (p.17).

The two experiments described here tested the effects of Seward Inc.'s *The First 4000 Words* (*4KW*), an online vocabulary instruction program designed to deliver instruction on approximately 4000 of the highest frequency words found in written English. A novel approach to providing remedial vocabulary instruction created by Greg Sales and Michael Graves, the *4KW* program uses computer-adaptive software programming to individualize students' instruction based on their performance while using the program. This individualized approach brings an efficiency that allows students to work within a much larger corpus of target words than is realistic using more traditional vocabulary instruction strategies.

Attendant to such an individualized approach to teaching a large number of words is a challenge to measurement of student learning; any given student will study a unique set of new words, such that any single test is likely to lack the sensitivity necessary to detect growth and will randomly favor some students over others. That is, any given student may have encountered only a small fraction of the words that are tested on a typical standardized test of vocabulary. This is consistent with the National Reading Panel assertion that standardized tests lack the sensitivity required to detect changes in vocabulary knowledge within typical experimental timeframes (National Reading Panel, 2000). To address this challenge and determine whether, and to what extent, word learning is occurring, it is therefore logical to analyze word learning at the item level rather than the total test score level. The *4KW* software program, which tracks each student's progression through the instructional lessons in the program, makes such an analysis feasible. This constitutes a novel and more sensitive approach to measurement of vocabulary learning that might inform assessment of future vocabulary (or other) interventions.

The overarching hypothesis motivating this research is that direct instruction on the words that students are most likely to encounter in their reading will lead to better reading comprehension achievement. Testing that hypothesis will ultimately require larger scale and more sustained experiments than were conducted here. A logical and more fundamental prerequisite question is whether use of this vocabulary instruction program results in improved word knowledge. To this end, this research centers on the hypothesis that students using the *4KW* vocabulary program will exhibit better vocabulary test scores than students engaged in standard classroom practices. Two randomized field trials have been conducted to test the effects of using the *4KW* vocabulary instruction program on vocabulary test scores. A novel, item-level analysis of word learning was conducted in the second experiment to test for word learning not detected by standardized tests.

Chapter II

Review of the Literature

This chapter provides a review of the research literature in three broad topic areas relevant to the theoretical rationale supporting the two studies described in chapters 3-5. First is a review of work conducted to describe and define the relationship between vocabulary knowledge and reading comprehension. An important assumption underlying the research described in subsequent chapters is that growth in vocabulary knowledge plays a *causal* role in aiding reading comprehension achievement; this assumption provides the rationale for the strong interest in helping students build larger working vocabularies. Second is a review of research that documents how word learning begins in young children, and how group differences in vocabulary knowledge arise from environmental circumstances. Knowledge of these dynamics is important to ongoing efforts to identify those at risk for entering educational settings with vocabulary deficits, and to aid in proactive efforts to help families, teachers and students avoid the consequences attendant to poor vocabulary knowledge. Last is a review of what researchers have found to be effective strategies for teaching new words to students. A broad consensus has formed regarding the instructional components necessary for effective vocabulary instruction. Knowledge of these instructional features should inform curricula aimed at explicit vocabulary instruction, and many of these instructional strategies have been incorporated into the *4KW* vocabulary program tested here.

Theoretical and Empirical Importance of Vocabulary for Reading Comprehension

Theoretical Perspectives. This section of the paper is intended to provide theoretical reasoning in support of the supposition that vocabulary development is fundamentally important to reading comprehension.

An intuitively compelling theoretical model of reading that suggests the importance of well-developed vocabulary knowledge is that of automaticity theory (Laberge & Samuels, 1974). The theory stems from an understanding that cognitive and attentional capacities are limited and limiting in comprehension processes. Because reading is an endeavor that requires overlapping and concurrent performance of a number of component sub-skills, the corollaries of this principle can be applied to processing at different levels of cognition during reading. A student's efforts directed toward decoding letters into sounds, for example, necessarily detract from cognitive resources that might be utilized for word, sentence, or passage comprehension. When, through practice and repetition, decoding becomes more automatic, more cognitive and attentional resources are available for higher level processing. The importance of vocabulary knowledge in this regard is clear. If a reader is faced with a large number of unfamiliar words in a sentence, the attentional and cognitive demands of determining meaning for those words, whether through use of context or a dictionary, will undermine maintenance of a coherent semantic representation of the sentence. Conversely, a reader possessing vocabulary knowledge of all or most words in a sentence is more able to automatically activate meaning for those words and incorporate them into semantic processing, leading to greater comprehension. Laberge and Samuels (1974) noted that automaticity arises from practice and repetition, and from doing so with words in different contexts. This has strategic implications for the instruction of vocabulary, which are addressed later in this paper.

Constructivist models of text comprehension, such as Kintsch's Construction-Integration Model (Kintsch, 1988) emphasize the activation of word meanings in a "bottom-up" construction of comprehension. In this model, propositions, inferences, and elaborations lead to a network of spreading activation of related items, towards the construction of a coherent representation of text, the "situation model". Prior vocabulary knowledge is fundamental to both the initial activation that is elicited by a word and appropriate activation of relevant connections to related items/concepts. An incomplete or inaccurate situation model is the logical result if the reader lacks the ability to activate meaning from words in the text, or if activation fails to spread to related items/concepts.

An associative learning model that incorporates both "bottom-up" and "top-down" processing, the Latent Semantic Analysis Model (Landauer & Dumais, 1997) also suggests the importance of vocabulary knowledge. This model posits the importance of extensive inferences generated from limited explicit information as key to gaining knowledge from text, and suggests that inferences account for large amounts of what is learned from direct experience. Fundamental to generation of accurate inferences, however, is gaining appropriate and rich information from the stimulus that is explicitly present, i.e. the text. Knowledge of vocabulary allows for appropriate activation of related concepts and generation of a richer set of inferences from the words provided in the reading material, leading to better comprehension.

Rumelhart (1980) suggested the importance of schemata in text comprehension. Inherent in this and related models is the idea that readers generate text representations that are largely based on background knowledge or previous experience. Vocabulary knowledge is a most fundamental aspect of background knowledge as it relates to text, so it follows that schemata generated from text wherein all words are known to the reader will necessarily be more in keeping with the intended meaning of the text. A lack of relevant vocabulary knowledge is likely to result in schemata, and therefore comprehension, that is incomplete and/or inaccurate.

A model of reading comprehension that argues for both verbal and non-verbal components, Dual Coding Theory (Pavio, 1991, Sadoski, 2005), suggests that comprehension results from an interplay between and within language-based and imagery processing. Concrete words in a text can activate both verbal and non-verbal representations which prime other, related concepts for activation as words and/or images. Sadoski (2005) suggests that vocabulary development is important for this process, because a word can elicit different activation networks depending on context. A strong vocabulary base facilitates activation of networks that are appropriate given the specific context. Ehri et al. (2007) extends this argument to the orthography (spelling) of words, suggesting that spelling plays an underappreciated role in comprehension by eliciting network activation as hypothesized by Dual Coding Theory. Ehri et al. argue that vocabulary development is fundamental to building cognitive connections between orthography, phonology and semantic representations.

Support for the importance of vocabulary knowledge to successful reading achievement also comes from theories of motivation. The amount of reading in which children engage is a strong predictor of reading achievement (Stanovich & Cunningham, 1992). The amount of reading children do is, in turn, related to their motivation to do so. The dynamics that stem from differing levels of motivation among successful and unsuccessful readers contribute to Matthew effects in reading (Stanovich, 1986). Poor vocabulary skills can lead to failures in reading comprehension and attendant academic failure and frustration. The lowered self-efficacy poor readers experience can result in even less reading, while successful readers are motivated to read more, improving their subsequent performance. The result is an increasing chasm between the vocabulary achievement and comprehension abilities of good and poor readers. Successful vocabulary development, which would be predicted to aid in reading comprehension and achievement for a variety of theoretical reasons suggested above, might increase motivation to read by promoting self-efficacy, success in classroom reading, and reading enjoyment.

Empirical Support. While it is apparent that vocabulary development is explicitly or implicitly important to any coherent theoretical model of reading or reading comprehension, empirical evidence also points to the importance of vocabulary to reading achievement and other indicators of reading ability. Correlations between performance on vocabulary and reading comprehension measures typically range from .8 to.9 (Stahl & Nagy, 2006), comparable to the correlations between different measures of reading comprehension itself (Ehri et al., 2007).

In one of the earliest documented attempts at assessing the effects of intensive vocabulary instruction, Draper & Moeller (1971) found that a year-long vocabulary program of 1800 words delivered to 24,000 4^{th}- 6^{th} grade students in the St. Louis, MO area public schools resulted in gains on vocabulary and reading comprehension subtests of the Iowa Tests of Basic Skills that were significantly larger than those predicted based on traditional growth estimates.

Addressing the consistently high correlation between vocabulary knowledge and reading comprehension, Anderson & Freebody (1981) posited three potential explanations. The first, “instrumental” hypothesis, is that vocabulary knowledge plays a causal role in comprehension; vocabulary knowledge leads directly to comprehension of text. A second, “aptitude” hypothesis, suggests that the correlations between vocabulary knowledge and reading comprehension reflect some underlying general aptitude or verbal intelligence. A third, “knowledge” hypothesis, posits a construct of world knowledge or background knowledge that underlies performance on both vocabulary and comprehension tests. While it seems likely that these hypotheses are not mutually exclusive, several studies subsequent to the Anderson & Freebody (1981) article appear to have been specifically designed to test the plausibility of the “instrumental” hypothesis.

Kameenui et al. (1982) tested the importance of vocabulary knowledge for reading comprehension by asking whether replacement of easier vocabulary words with more difficult synonyms would decrease passage comprehension, and whether vocabulary instruction on difficult words in a passage would improve comprehension. Anderson & Freebody (1981) had suggested that affirmative results to these two experimental questions would provide direct evidence in support of the "instrumental" hypothesis. In a series of randomized trials with 4th- 6th grade students, Kameenui et al. (1982) found that replacing easier vocabulary terms with more difficult ones decreased passage comprehension significantly, as measured by both inferential questioning and passage recall. This research also showed that vocabulary instruction on more difficult words in a text led to better comprehension scores, as defined above. While cautious in interpreting their results, these researchers suggest that the results are consistent with a view positing the importance of vocabulary knowledge for reading comprehension.

In another study, 4th grade students received long term vocabulary instruction and were compared to a matched control group on a variety of vocabulary and reading-related measures in a quasi-experiment conducted by Beck et al. (1982). The instructed group received vocabulary instruction on 105 words over a five-month period. After controlling for prior achievement on standardized tests of vocabulary and reading, the instructed group out-performed the control group on vocabulary and semantic decision tests of the target words. In addition, the instructed group not only showed better comprehension of custom reading passages but scored significantly higher on standardized tests of vocabulary and reading comprehension. Though they were cognizant of potential confounds due to their quasi-experimental research design, these authors suggest that appropriate vocabulary instruction can result in faster and more accurate semantic processing of words, and that effects of instruction can transfer to both words not explicitly taught and general reading comprehension (Beck et al., 1982). After identifying and correcting potential confounds in their methodology, this group replicated their experiment with similar results (McKeown et al., 1983). Though quasi-experimental, these two studies remain the most widely cited research studies supporting a causal role for vocabulary instruction on improved reading comprehension.

Furthermore, a study conducted with 1st - 5th grade students, Roser & Juel (1981) found that students receiving direct vocabulary instruction on words to be encountered in subsequent story reading performed significantly better on a follow-up comprehension assessment. Notably, the largest improvements came from those students in the lowest reading ability groups.

In a meta-analysis of more than 50 studies relating vocabulary instruction to reading comprehension, Stahl & Fairbanks (1986) argue that the data support a causal role for vocabulary instruction in improving reading comprehension, consistent with the instrumental hypothesis of Anderson and Freebody referenced above. They estimate an effect size of .97 for the effect that vocabulary instruction has on comprehension of reading passages that include taught words, and an effect size of .30 for comprehension of passages that do not include explicitly taught words. In discussing the possible mechanisms by which vocabulary instruction might result in better comprehension even if passages do not include target words, Stahl & Fairbanks (1986) draw no conclusions but suggest that one possibility is increased interest from students toward learning new words. Because at least some of the vocabulary instruction methods included within this meta-analysis use contextual strategies for learning new words, it is possible that students take advantage of such strategies in their subsequent reading, which would result in better comprehension even if explicitly taught words are not involved.

In a more recent meta-analysis, Elleman et al. (2009) estimated smaller effects of vocabulary instruction on reading comprehension. They estimated comprehension effect sizes of .50 for custom reading passages and .10 for standardized reading passages. They also found that struggling readers were the most likely to benefit from vocabulary instruction, consistent with Roser & Juel (1981). On custom reading passages, struggling readers demonstrated an average effect size of 1.23, compared to an effect size of .39 for average and good readers. It is possible that the small effects demonstrated on global reading measures reflect, in part, the limited number of words taught in almost all of the studies included in these meta-analyses (none of the included studies reported instruction on > 105 words).

Failure to generate appropriate inferences during reading can result in poor comprehension (Cain & Oakhill, 1991). Calvo (2005) reasoned that inferences are, in part, generated on the basis of appropriate vocabulary knowledge of words integral to the inference. In an eye-tracking study, Calvo found that vocabulary knowledge facilitated the generation of appropriate elaborative inferences, as detected by fewer regressive eye movements from the target of an implicit inference source.

The theoretical views of reading and empirical studies cited here support the hypothesis that vocabulary knowledge is important for reading comprehension, and that direct vocabulary instruction should be expected to improve reading comprehension. There are, of course, a great variety of approaches to vocabulary instruction that have been used in educational settings. After a review of early vocabulary acquisition, a subsequent section of this chapter will provide a description of the components that are widely considered to be part of effective vocabulary instruction, keeping in mind that no single instructional strategy is appropriate in every educational setting.

Early Vocabulary Development and Group Differences

Early Vocabulary Acquisition. Children typically begin producing words late in the first year or early in the second year of life, at roughly 8-14 months. During this time, a common trend is movement from babbling toward articulation of performants or "speech acts" and finally production of referential words, usually nominative in nature (Caselli et al., 1999). Production of new words characteristically proceeds in a linear fashion until the infant has a productive vocabulary of around 50 words, at which point a well-documented acceleration in vocabulary acquisition, known as the vocabulary spurt, oftentimes occurs (Goldfield & Reznick, 1990; Dapretto & Bjork, 2000; Fernald et al., 2006). The vocabulary spurt typically occurs late in the second year, and results in the production of many new words, as well as words of different classes such as verbs and modifiers (Caselli et al., 1999). In addition to a rapid increase in production of novel words, there appears to be concurrent increase in the vocabulary comprehension abilities of infants, as measured by picture fixation performance in "looking while listening" paradigms (Fernald et al., 2006). Performance in these tasks becomes both more accurate and efficient. While 15-month-old infants respond only upon hearing an entire word, 25-month-old infants respond to far more words and fixate *during* word presentation, apparently using phonetic onsets to identify vocabulary items. These changes in performance, followed longitudinally, consistently correlate with lexical development (Fernald et al., 2006).

It is now clear that there are large individual differences which limit the degree to which infants' vocabularies can be appropriately described as undergoing a "spurt" in a uniform manner (Ganger & Brent, 2004). There appears, however, to be a very common and substantial increase in the rate of vocabulary acquisition that occurs after infants have acquired their first 50 words or so. The cognitive processes and changes that underlie this expansion in vocabulary, be they gradual or discrete, have been the subject of extensive research and supposition.

Prior to the age of six months the structural anatomy of an infant's vocal tract may preclude the motor dexterity and the range of formant production required to produce recognizable words (Kuhl & Moeltzoff, 1996). As the vocal tract develops to a more adult-like form, infants begin to actively imitate the sounds that they hear (particularly vowels), and progress from babbling towards production of meaningful speech. Once infants possess the motor abilities required to produce coherent words and are therefore not restrained in that regard, what cognitive processes govern the rate at which they produce and understand new lexical items?

Central to a number of accounts of new word production is the hypothesis that comprehension abilities outpace production abilities in the early years of life (likely in part to motor limitations suggested above). Benedict (1979) suggests that comprehension of words precedes production by 3-5 months late in the first year, and that the rate of new word comprehension is double that of production at the onset of the vocabulary surge. By the age of two the rates of oral comprehension and production are closely aligned. This view suggests that the vocabulary surge, in part, reflects an ability to express words that have been comprehended for some time. In a similar vein, Dapretto & Bjork (2000) suggest that comprehension ability precedes production significantly, and that retrieval of the appropriate sound pattern initially delays production. Their research demonstrated that infants with more productive vocabularies were able to retrieve the words for objects hidden in a box while subjects with less productive vocabulary performance could not, although both groups possessed the words in their spoken and receptive vocabulary. As the ability to retrieve appropriate sound patterns increases in efficiency, late in the second year, word production increases.

Most recent research on early vocabulary acquisition revolves around phonemic awareness and phonological memory. Six-month-old infants can discriminate between phonemes, and their ability to do so predicts productive vocabulary at age two (Tsao et al., 2004). Gathercole et al. (1992) posited a causal role for phonological memory, as measured by non-word repetition tasks, in vocabulary acquisition. Their longitudinal study suggested a causal role for phonological memory until age four to five, when vocabulary knowledge potentially provides a greater scaffold for learning new words. The trends in their data suggest that phonological memory may play an even more important role in younger children. The non-word repetition tasks these researchers used predicted vocabulary knowledge a year later, after controlling for non-verbal intelligence and previous vocabulary scores. They suggest a subsequent reciprocal relationship between phonological memory and vocabulary, such that phonological memory capacity is improved by greater experience with new words (Gathercole et al., 1992).

Similar conclusions were reached by Metsala et al. (1999), hypothesizing that initial coding of words is performed at a holistic level, with segmentation based on phonological processing occurring with experience. As phonological processing increases in efficiency, due in large part to vocabulary acquisition, cognitive capacity to encode new stimuli is increased; a reciprocal relationship between vocabulary and phonological processing is again implied (Metsala et al., 1999).

This brief survey of theoretical viewpoints related to initial vocabulary acquisition suggests that the complex nature of the tasks that infants and children face and overcome in this regard requires analysis at a number of different levels of processing. The hypotheses that are summarized above should not be viewed as competing models, but as perspectives that approach vocabulary development from different, complementary viewpoints.

In 1992 Gathercole et al. suggested that “although studies of word learning in children have documented the remarkable facility of preschool children to acquire new vocabulary…the factors underpinning the large individual differences in young children's abilities to learn new words are as yet little understood” (p. 887). Although significant advances have been made in the past 15 years, particularly in areas of phonological processing, the above quote is particularly relevant to practical application of theory to guide early vocabulary instruction. Below are a few suggestions that would seem to follow logically from what we’ve learned about early vocabulary development.

The most obvious and likely most important practical piece that comes from research on vocabulary acquisition is the necessity for infants to be exposed to rich language stimuli. Studies conducted with infants by Saffran et al., (1996) and others suggest that more language input to the infant's processing system will allow for more rapid and thorough word segmentation. If words are being appropriately and efficiently segmented from fluent speech, cognitive resources are freed to conduct phonological processing within words, which would likely increase phonological awareness. As a number of the above researchers posit, phonological awareness is a key component of vocabulary acquisition. Even if motor constraints prevent production of coherent words during the first year, the findings of Dapretto & Bjork (2000), and Fernald, et al (2006), which posit the advanced capacity for comprehension relative to production, would suggest that rich language stimuli would result in more robust vocabulary development.

The phonological-based perspectives of Metsala et al. (1999), Bowey (2001), and Gathercole et al. (1992), as well as the morpheme-based perspective of McBride-Chang, et al. (2006), all suggest that increasing the extent and quality of language input to the child would lead to better vocabulary performance. All of this is to point out the obvious: we should be aggressive about providing infants and young children with rich language experience. Previously held beliefs that vocabulary acquisition depends primarily on innate differences in ability to learn based on heredity are not supported by the evidence. Scarr & Weinberg (1978) demonstrated that the correlation between vocabulary performance by young children and their biological mothers was no greater than that between children and their adoptive mothers, in contrast to other subtests on the Wechsler Adult Intelligence Scales. Studying 14-26-month-old infants, Huttenlocher et al. (1991) found that the amount of speech directed at children by their mothers was highly predictive of the rate of vocabulary development.

Many of these important developments in lexical processing occur during the first few years of life, and the trajectory of vocabulary acquisition appears largely linear from ages three to six. Providing a rich environment to infants is therefore critical to prevent diverging trajectories and large deficits for many children. The research of Hart & Risley (1995) speaks to this directly. They found that the vocabularies of children from higher socio-economic families were double that of children from lower socio-economic groups at age three and that these differences swell up to and through the onset of formal schooling. That study and others described below strongly implicate environmental factors as causative agents in vocabulary development and the deficits that some children face in this regard.

Group Differences in Vocabulary Acquisition. It would be expected that children entering educational settings would exhibit variability in the extent of their vocabulary development, as with any other ability or psychological construct. The degree to which some groups enter schooling behind their peers in vocabulary, however, is both striking and cause for serious concern. It is striking in that vocabulary deficits are so extreme, and cause for concern due to the critical importance of a well-developed vocabulary for reading comprehension, as described above. Struggles with reading comprehension are inherently problematic, but also lead to difficulty in virtually every aspect of educational endeavor. Although research specifically examining relationships between socioeconomic/ethnic/racial groups and vocabulary development is limited, the results of studies that have been done are consistent in identifying significant differences.

One identifiable group that has been repeatedly found to enter educational settings with significant deficits in vocabulary development is defined by SES. Although it is well-established that lower SES children enter school settings with cognitive abilities less developed than more advantaged peers (Lee & Burkam, 2001), the deficit in vocabulary is particularly apparent. A stark example of this was found in the research conducted by Hart & Risley (1995, 2003). Hart and Risley conducted longitudinal research with 42 families categorized on SES as professional, working-class, or welfare. Monthly monitoring of parent-child interactions over two and a half years, beginning with children at age seven months, found that children in high SES homes acquired expressive vocabularies of greater than a 1000 words at age three, while the low SES children acquired less than half that amount. Over the course of this study, high SES parents spoke more than double the average number of words per hour to their children than low SES parents, and used an average of 2000 different words overall compared to 1000 for low SES parents. A follow-up study showed that differences in vocabulary between these higher and lower SES children continued to increase, such that by age six a 6000 word gap between SES-defined groups was estimated. Extrapolating their data, these researchers suggest that the higher SES children were exposed to 30 million more spoken words than their lower SES counterparts by the age of three (Hart & Risley, 2003).

Hart & Risley's research was conducted with families in one Midwestern town, and is vulnerable to threats to external validity. In a larger and more nationally representative sample, however, Farkus and Beron (2004) reached many of the same conclusions. They analyzed a large longitudinal data set (Children of the National Longitudinal Survey of Youth, CNLSY) that measured vocabulary scores over a ten year period, tracking children from age three to age thirteen. Their results confirmed the finding of large vocabulary gaps between high and low SES children on measures of oral vocabulary. Going beyond the results of Hart & Risley (2003), Farkus & Beron (2004) found that although large vocabulary gaps are present between SES (and racial, see below) groups as they enter formal schooling, these differences remain relatively constant after the age of six. This finding suggests that schooling may have a stabilizing effect on group differences in vocabulary achievement. However, White et al. (1990) found that an average 2nd grade middle SES student had a reading vocabulary approximately 5000 words larger than an average low SES student, and that the gap continues to widen through at least the 3rd and 4th grades.

Research specifically focused on ethnic/racial group differences is also limited, and results of some studies are confounded by a failure to control for SES. Farkus & Beron (2004) found large gaps between white and black children's oral vocabularies at age three, after controlling for SES. Farkus & Beron's analysis suggests that race and SES contribute almost equally to the difference between black and white children's vocabulary achievement. These authors noted that the average black child already lags a full year behind the average white child in expressive vocabulary by the age of three, based on differences in performance and vocabulary acquisition rates.

A wide variety of mechanisms have been suggested that might mediate SES-based differences in vocabulary achievement and cognitive development of children in general. Family income and material hardship, educational achievement of parents, access to educational materials, reading habits of parents, family structure, nutrition and wellness variables, etc. have all been suggested as mediators of SES-based differences in achievement and general cognitive performance (Duncan & Magnusun, 2005). With respect to vocabulary achievement, in particular, variables related to language use in the home have been implicated as important mediators. It is clearly the case that infants are dependent upon incoming language stimuli for acquisition of vocabulary (i.e. children who lack this input, such as feral children, fail to acquire language). Huttenlocher et al. (1991) demonstrated that the amount of parental speech was a significant predictor of rate of vocabulary acquisition in children ages 14-26 months, and that the frequency of individual words spoken by parents was a strong predictor of the order in which those words appeared in children's expressive vocabulary.

In addition to the findings of Hart & Risley (1995) referenced above, a number of studies have identified quantitative and qualitative SES-related differences in parental/maternal speech directed at children. In an observational study of maternal speech directed at two-year-old infants, Hoff (2003) found that differences in the language use of 30 high SES mothers and 30 middle SES mothers fully accounted for the differences in vocabulary growth between the respective groups of children over a two and a half month period. High SES mothers used a larger vocabulary and the total amount of words spoken was significantly greater than mid-SES mothers. In addition, utterances from high SES mothers were significantly longer than those from lower SES mothers (Hoff, 2003). These observations suggest that the stimuli received by higher SES children are richer and more complex, which might facilitate vocabulary development in a variety of ways related to phonemic and morphemic processing described above. In addition, a richer vocabulary input may be important at a contextual level.

In a study of 50 mothers and their children interacting at home, Weizman & Snow (2001) identified over 150,000 total words spoken by the mothers, with the most frequent 3000 words comprising 99% of the total. The number of less-frequent, sophisticated words was a more powerful predictor of children's kindergarten and second grade vocabulary scores than was total number of words spoken. The density of sophisticated words within maternal speech when children were five years old accounted for a third of the variance in second grade vocabulary scores. Lower SES mothers' quantity and quality of verbal interactions with their infants may be affected by a lower sense of competence in parenting skills, smaller vocabulary, or parental satisfaction, all of which might act to reduce verbal interaction with children (Bornstein et al., 1998). Hoff and Tian (2005) found that higher SES mothers asked more questions of their children, while lower SES mothers' speech was more directive in nature. It could be that higher SES children are given more opportunities to talk in response to questioning, which would be expected to aid in expressive vocabulary competence. Consistent with this possibility, Walker et al. (1994) found that lower SES children were more likely to be prohibited from talking by their parents. Knowledge of principles of child development was identified as a key mediator of SES-based differences in amount of speech directed to children. Higher SES mothers are more likely to understand the importance of verbal interactions for language development and therefore engage in more of them (Rowe, 2008).

The studies referenced above strongly suggest that the differences in vocabulary performance between higher and lower SES children are at least in part due to differences in the quantity and quality of spoken interactions between infants/children and their parents. Further research is needed to identify the specific components of parent-child interactions which facilitate vocabulary growth, and to uncover other variables that mediate the vocabulary gaps that arise between these SES-defined groups prior to entry into formal schooling.

As noted above, in reference to the findings of Farkus and Bowen (2004) that race accounted for significant variance in vocabulary scores between black and white children after accounting for SES, it appears that cultural differences may also play a significant role in young children's vocabulary achievement. A number of studies have identified culturally-based differences in the ways parents and children interact with language, although research specifically relating these differences to vocabulary development are lacking. For example, in a study of shared book reading to Head Start children, Hammer et al. (2005) found that Hispanic mothers used more words and made more non-text-based comments than African-American mothers, and Hispanic children were more assertive in asking questions. In a study conducted in the southeastern U.S., low income African-American parents were found to be less likely to engage in child-directed speech than their white counterparts, reflecting a difference in belief about the role of such speech in children's verbal development (Heath, 1983). A study comparing the mother-child interactions of a sample of mothers in the U.S. with a group of mothers from the Gusii of Kenya found that the Gusii mothers avoided eye contact and discouraged verbal expression from their infants, while the American mothers attempted to elicit such expressions (Richman et al., 1992). Given the demonstrated importance of language interactions for vocabulary development, it's likely that differences in parental interactions with children play a causal role in ethnic/cultural/racial differences in vocabulary achievement. Becker (1977) suggests that insufficient vocabulary knowledge is the primary barrier to achievement for lower SES students, and argues that systematic vocabulary instruction is required to address the achievement gap.

Another important and growing group in the U.S. that can be identified as at risk for deficits in English vocabulary development is English language learners (ELL). Approximately one in nine K-12 students in the U.S. is ELL, and this ratio is predicted to grow to nearly one in three in the next five years (Francis & Vaughn, 2009). As non-English speaking immigrants or the children of families wherein English is not the primary language, it's not surprising that these children might lag behind their English-speaking peers in English vocabulary development. However, this puts them at high risk for failure in educational settings. The inherent deficits that these children often face with regard to spoken English prior to entry into English-language schools can lead to struggles with phonological and orthographic processing, and word recognition skills relative to their English-only (EO) counterparts (Verhoeven, 1990). ELL students' potential lack of adequate English vocabulary makes them less likely to comprehend grade-level reading materials; they subsequently score lower on standardized tests, and are more likely to be diagnosed as learning disabled or placed in special education (August et al., 2005). ELL students underperform relative to their EO peers on the National Assessment of Educational Progress reading measure at fourth, eighth, and twelfth grade levels (NCES, 2005). While nearly a third of native English 8th grade students are at the proficient-level on national achievement reading tests, only 4% of ELL students meet this threshold (Lee et al., 2007). Lack of academic vocabulary knowledge, which impedes comprehension of content area text, is a primary obstacle for ELL students' academic success (Vaughn, et al 2009). Compounding the problem for many ELL students is that many are also lower SES students, and therefore they face challenges in that regard as well.

ELL is a label that encompasses a tremendous variety of students. A majority of ELL students in the U.S. come from homes or backgrounds with Hispanic base languages, and tremendous variability exists in children's literacy and vocabulary knowledge in their native language within that population. These differences in native-language proficiency are likely to affect English vocabulary and other literacy-related achievement. The ELL label also applies to children from other linguistic backgrounds. The different orthographies, phonemic constraints, and lexical compositions of some native languages place greater challenges on some groups relative to others in English vocabulary development.

Whether a student's vocabulary deficit stems from a disadvantaged language background or is a result of growing up in a non-English speaking home, the advantage that a student with a stronger vocabulary has is one that has developed gradually, likely over a period of years. This suggests that remedial efforts will require sustained effort over a similar time scale (Hirsch, 2001). The negative educational (and by extension societal) impacts that result from deficiencies in vocabulary are such that aggressive and robust efforts must be made to remediate them.

Effective Vocabulary Instruction

Current Theory. Vocabulary instruction cannot be a simple matter, because vocabulary acquisition is incremental and complex (Nagy & Scott, 2000). As described by Ehri et al. (2007) vocabulary acquisition necessarily incorporates and connects orthographic, phonological, and semantic representations. In addition, both phonological and, to a greater degree, semantic aspects of a given word can vary substantially in different contexts. Even the most literate adults are likely to be continuously making subtle alterations to the semantic representations of some of the words in their lexicons, as well as adding new lexical entries. Vocabulary acquisition is an ongoing and complex process.

Much of the literature describing effective vocabulary practices reiterates and builds upon three central components suggested by Stahl (1986). Subsequent to an extensive meta-analysis which suggested positive effects of vocabulary instruction on reading comprehension (referenced above, Stahl & Fairbanks, 1986), Stahl identified key features of the instructional programs that had the strongest positive effects on comprehension. Three components were identified as parts of effective vocabulary instruction. First, instruction should provide both definitional and contextual information. To “know” a word requires that a student be able to offer some reasonable definition of it, but also be able to use the word and understand how it is used in context. Accordingly, the best instructional programs provide a user-friendly dictionary-like definition, complete with synonyms, antonyms, word roots, etc, and also present the word embedded in sentence context, preferably more than once. Second, good vocabulary instruction should promote deep processing, which Stahl suggests can be considered as connecting the new word to known words, concepts and background knowledge. Stahl posits three successive stages to deep processing; *association,* wherein the word is associated with other known words, which might be synonyms, antonyms, etc; *comprehension*, wherein the student is asked to do something with the word, such as fit it into a sentence using a cloze procedure or find antonyms/synonyms for the word; and *generation*, wherein the student does something novel with the word, such as construct her/his own definition, or use the word in a written or oral context. The third component of effective vocabulary instruction identified by Stahl is multiple exposures to the word. This requires that the word be encountered more than 1-2 times, and preferably in a variety of contexts. This allows the time necessary for deeper processing to take place, and is also likely to facilitate automaticity with the word, a key theoretical step towards fluent reading (Laberge & Samuels, 1974).

Consistent with the three fundamental aspects of effective instruction described above, Stahl also suggests that his meta-analysis revealed what doesn't work. Vocabulary instruction that provides only definitions of words, or provides only 1-2 exposures to target words was not correlated to any gains in reading comprehension. Stahl points out that these ineffective practices are exactly what *are* found in many middle and high school texts, and suggests the above guidelines be used for both creating effective vocabulary instruction and for screening/selecting educational texts when possible. These principles have been borne out in subsequent studies of the effects of vocabulary instruction on both vocabulary achievement and reading comprehension. Bos & Anders (1990) and Nash & Snowling (2005) both showed that instruction focused on contextual approaches to vocabulary learning was superior to definition-only instruction. Both studies were conducted with students identified as learning disabled or having poor vocabulary knowledge, which represents a large fraction of the population for whom remedial vocabulary instruction would be appropriate.

Building upon this basic set of principles, Stahl & Nagy (2006) argue for a comprehensive approach to vocabulary development. In addition to teaching the meaning of specific words in an explicit manner, characterized by the three principles identified above, Stahl & Nagy emphasize the importance of immersing students in extensive exposures to rich oral and written language. A preponderance of vocabulary acquisition occurs incidentally, through reading and oral interactions. This is demonstrable by comparing students' vocabulary growth, estimated as high as 5000 words per year in the early and middle grades (Nagy, et al 1985) with estimates of the number of words explicitly taught , which range from 200-400 per year (Stahl & Fairbanks, 1986). This ratio suggests that extensive exposure to rich language is a powerful means of vocabulary development. Since oral comprehension exceeds reading comprehension extensively in early grades, oral storytelling and oral reading should be an integral part of developing vocabulary. Extensive opportunities for reading texts rich in language are also important, particularly as students advance towards a balance of oral and reading comprehension abilities, typically in the middle school years. Reading facilitates the positive reciprocal relationship between vocabulary acquisition and reading comprehension (Baker et al., 1995). Stahl & Nagy suggest the importance of establishing "word consciousness" (Anderson & Nagy, 1992) in young learners, suggesting that a multi-faceted approach is required to instill children with a sense of curiosity and awareness of words and how they are used. Exercises comparing de-contextualized (written) and contextualized (oral) ways of communicating the same ideas can help students identify how words are important. Game playing and other activities that center on words can be powerful for creating word consciousness, and lead students to view new words as puzzles to be solved rather than insurmountable obstacles.

Stahl & Nagy (2006) also suggest that care must be given to identifying *which* words will be the focus of direct instruction. Since the number of potential unknown words in a particular reading often precludes teaching all of them, teachers must make judgments about the importance and utility of any potential vocabulary item. Determination of whether a given word is important to the overall understanding of a reading passage, and the degree to which the word will appear repeatedly in future reading should guide instructional focus. Decisions on which words to teach directly must also be informed by a sense of students' existing vocabulary knowledge. When learning a new word, a student may be in one of three knowledge conditions relative to the word (Graves & Prenn, 1986). She/he may already have the word in her/his oral vocabulary, may be unfamiliar with the word but have understanding of a closely related concept, or may be unfamiliar with both the word and the concept. It is impossible for a teacher to have this specific knowledge for all words and all students, but teachers should use all information available, especially the spoken or written language used by students, to make thoughtful decisions in this regard (Stahl & Nagy, 2006).

An application of many of these principles can be found in the "robust" or "rich" vocabulary instruction implemented by Beck & McKeown (2007) and Beck et al. (1982). Robust vocabulary instruction is characterized by giving definitions of vocabulary words in student-friendly language, providing the words in multiple and varied contexts, allowing students to create their own definitions and contexts, and asking students to give examples of both appropriate and inappropriate use of the word. This was found to improve reading comprehension significantly relative to conditions of no or reduced vocabulary instruction.

Another effective approach, consistent with an emphasis on the importance of younger students' oral vocabulary development, is shared book (Graves, 2008), or dialogic (Whitehurst et al., 1994), reading. These strategies modify the traditional, "adult reads to child" model to make it more interactive. In shared book reading the (adult) reader frequently pauses to ask questions about the story and target words in particular, and allows the child to ask questions as well. In this interactive process, the story is typically read more than once and target words are the focus of repeated questions and discussion, leading to deeper processing than is possible in a one-time, uninterrupted reading experience. Shared book reading, designed to expand oral vocabulary, is particularly effective and useful for younger children who are not yet reading and for those students whose oral vocabularies are deficient (Whitehurst et al., 1994).

A number of researchers (Stahl & Fairbanks, 1986; Stanovich, 1980; Nagy et al., 1985; Garcia, 1991) argue that the vocabulary deficit faced by many students often precludes the ability to gain meaning of a word from context, and suggest that direct vocabulary instruction is therefore most important and most beneficial for students with low vocabulary knowledge.

Comprehensive descriptions of effective strategies for vocabulary instruction targeted to more narrowly defined groups such as ELL are emerging. Most of the themes are built upon what has been demonstrated as effective for English-only learners described above. ELL students will enter school settings with a wide range of English language knowledge, so an awareness of students' oral language proficiency is a prerequisite for constructing activities that are appropriate and likely to result in successful learning. More extensive scaffolding, simple and clear instructional objectives, plentiful opportunities for verbal interactions, and identification and use of Spanish-English cognates (for children with Hispanic-based first languages) are all likely to be helpful in increasing the vocabulary knowledge of ELL students (Helman, 2008). Bialystok (2007) suggests that a bilingual perspective may enhance meta-linguistic awareness, which should be a goal of any comprehensive approach to vocabulary instruction. There is broad consensus among scholars and experts in this area that ELL vocabulary instruction should be both explicit, with targeted words defined, presented in multiple contexts, and actively used by the students in novel contexts, and implicit, through increased word consciousness and extensive exposure to rich oral and written language

A Role for Technology. The information presented in this chapter suggests that teachers are faced with a tremendous challenge. Success in virtually all academic areas is fully dependent upon reading comprehension, which requires strong vocabulary knowledge. A large percentage of students, however, come to the classroom with deficits in vocabulary knowledge that range from moderate to profound. Addressing the significant vocabulary deficits that a number of students in a classroom might have would be a daunting task, even if all of the deficits were the same, but they are not. Knowledge of each student's vocabulary (and therefore deficit) would be necessary to avoid either teaching words that are already possessed in the student's vocabulary or teaching words that are too advanced for the student. The critical importance of addressing vocabulary deficits and the limited time teachers can allot to this effort necessitates an efficient, individualized strategy. Even with the assistance of an aid or specialist, it is probably not reasonable to expect an individual teacher to meet this challenge, particularly in schools that have high proportions of ELL and/or low SES students.

There is reason to believe that technology may play an important role in meeting this challenge. Computer-based strategies for delivering vocabulary instruction are emerging, and the National Reading Panel Report (2000) states that "The use of computers in vocabulary instruction was found to be more effective than some traditional methods… It is clearly emerging as a potentially valuable aid to classroom teachers in the area of vocabulary instruction" (p. 14) and "computer technology must be examined for its ability to deliver instruction, for example, in vocabulary…" (p.17).

Vocabulary instruction has been a favorite target of computer-based instruction since computer use expanded in education in the 1980's (Ma & Kelly, 2006). Typical methods have been games and cloze-like designs developed to aid ELL or low-level learners, but quality and effectiveness of many of these packages has been low due to a lack of theoretical or pedagogical knowledge on the part of program designers (Ma & Kelly). A divide in this field between those driven by technological goals and those focused on theoretical and pedagogical goals has been an ongoing impediment to production of highly effective computer-based programs for vocabulary (and other) instruction (Levy, 1997).

There has been progress towards bridging the gap between pedagogically driven and technologically driven approaches to use of computers for vocabulary instruction. In a randomized study with sixty second-language learners, subjects were assigned to either a treatment wherein three hours per week were spent on computer-delivered instruction of 2000 high frequency words or a control condition of reading text and comprehension tasks. Treatment subjects performed significantly better on subsequent tests of both vocabulary and reading comprehension (Tozcu & Coady, 2004). Although the preponderance of computer-delivered vocabulary programs have targeted ELL students, a number of studies (Clements & McLoughlin, 1986; Davidson et al., 1996; Heise et al., 1991) have supported the National Reading Panel's assertion that computer-based vocabulary instruction can aid vocabulary development of native English speakers (Kamil & Hiebert, 2005).

Wood (2001) conducted an evaluation of 16 software packages designed to teach vocabulary, determining the degree to which each aligned with consensus theoretical concerns. Wood examined whether instruction attempts to connect the new word to background knowledge, promotes in depth processing (as defined by Stahl, 1986), provides multiple exposures, allows for direct and incidental word learning, or promotes further reading by the student. The software packages that focused on direct instruction were largely found to mimic traditional strategies of definition memorization, matching and fill-in-the-blank exercises. Very little activation of background knowledge, facilitation of deep processing, or encouragement to read more was found in these packages. More comprehensive software packages were better, often embedding new words in context designed to activate students' knowledge, allowing for associative and comprehension-level processing, and providing opportunities for direct and indirect word learning, but often included words too difficult for the target audience. Wood suggests that future software programs provide more opportunity for generative processing and actively encourage more reading by the student.

Many studies of online approaches to vocabulary instruction in recent years have focused on the use of computerized glosses, online dictionaries, and multimedia software to facilitate incidental second language vocabulary learning. Chun and Plass (1996), Li (2010), Yoshii (2006) and others have shown that use of computer glosses during reading lead to better vocabulary learning than print dictionaries for first and second language learners. A recent meta-analysis also demonstrated that computer glosses are highly effective in aiding both reading comprehension and vocabulary development for ELL students (Abraham, 2008). These studies, however, have been conducted with secondary and college-aged students and focused on incidental vocabulary learning during reading. Few studies have documented computerized approaches to explicit vocabulary instruction in young children. Sun and Dong (2004) used a multimedia strategy in their study with

1st and 2nd grade Chinese ELL students and found that vocabulary learning and sentence comprehension were enhanced by use of pictorial and auditory contextual cues in combination with de-contextualized word training. This is consistent with established research showing that explicit word learning is optimized by a combination of contextualized and de-contextualized target word instructions (Beck, McKeown, & Kucan, 2008; Graves, 2006).

A potentially promising avenue for computer-delivered vocabulary instruction may come from use of computer-adaptive technologies. Computer adaptive technology allows for enhanced efficiency and precision in assessment by estimation of proficiency and ongoing, individualized adaptation of instruction to match performance (Meijer & Nering, 1999). This technology might be used to identify the proper placement for an individual student in a vocabulary instruction program, and adjust the level of instruction to the student's ongoing performance. This approach addresses a fundamental obstacle that teachers face: how to individualize instruction for a group of students whose vocabulary deficiencies are not equivalent.

Computer-delivered vocabulary instruction cannot remediate significant vocabulary deficits by itself, but may be integrated with use of rich language in the classroom and exposure to appropriate level texts to help students with vocabulary deficits gain knowledge of the words they need to know to aid in reading development. It would seem that efficiencies available through computer adaptive technology, when combined with theoretically driven and pedagogically sound software programs, may be a very useful strategy for introducing students to words and word meanings as part of a comprehensive approach to vocabulary deficit remediation.

Previous research findings, described above, clearly demonstrate that explicit vocabulary instruction combining user-friendly definitions with repeated exposure to words used in context can lead to effective word learning. However, a large number of the highest frequency words must be learned by students to facilitate basic reading skills and allow for the exponential growth in vocabulary knowledge made possible by implicit word learning through exposure to written materials. The need for individualized, explicit instruction on a large corpus of high frequency words motivated the research detailed in this manuscript. In the following chapters two studies are described that tested a computer adaptive, online vocabulary instruction program, *The First 4000 Words*. This software program is designed to deliver remedial vocabulary instruction on approximately 4000 of the highest frequency words in written English. Study 1 was conducted to test whether use of a simple prototype of *The First 4000 Words* program would result in increased word knowledge on a small corpus (n = 100) of target words. Study 2 tested whether use of the complete program (n = 4000 words) would result in better performance on standardized-like tests of vocabulary knowledge.

Chapter III

Methods

Study 1

Research Design

This study was a randomized field trial. Students who were identified as potential subjects and who met parental and student consent criteria were randomly assigned to receive either the online vocabulary intervention or "business as usual" classroom instruction. The use of random assignment allows the potential for strong causal inferences to be drawn from differences in the performance of treatment and control subjects.

Setting. This study was conducted during a summer school session in 2008 at an elementary school in a large Midwestern metropolitan area in the United States. The school is suburban and serves very diverse, mostly lower income families (>70% free/reduced lunch).

Participants. Summer school students entering Grades 2-4 from a racially diverse, economically disadvantaged (>70% free or reduced lunch) first-ring, suburban elementary school of a major Midwestern city enrolled in this study. From this pool of students, 43 students (stratified by grade, approximately 14 students per grade level) were selected for inclusion in the study based on their poor performance on the pretest (see test details below). These students were randomly assigned to experimental (*n* = 22; 7 female) or control conditions (*n* = 21; 13 female).

Materials

Pretest and Posttest. The pretests and posttests consisted of the same paper-and-pencil vocabulary test with 40 multiple-choice items. Each item consisted of a black-and-white drawing with four word choices (See Appendix A for sample items.). The 40 words tested on the pre- and posttests were selected from the 100 target words taught to students in this study (See below). Testing with similar students and similar tests has yielded internal consistency reliability coefficients, Cronbach's alpha, between .7 and .9 depending on the grade.

The First 4000 Words Software. The First 4000 Words software program systematically delivered vocabulary instruction on 100 words selected from The Educator's Word Frequency Guide (Zeno et al., 1995). The 100 words were divided into 10 lessons with 10 words each. The 40 words on the paper-and-pencil pre- and posttests (see section ‘‘Pretests and Posttests’’) were embedded within the 100 words targeted in the computerized vocabulary instruction. The lessons were graded incrementally in regards to selected word frequency. The selected words that occur more frequently were generally included in the first lessons, and the words that occur less frequently were generally included in later lessons. Each lesson consisted of five components: vocabulary reading pretest, interactive oral reading, interactive vocabulary activities, repeated reading, and vocabulary reading posttest. Each component is described in detail below.

Each lesson began with a greeting from a Woodland Animal Guide (e.g. Redwood the fox), who guided the student through the lesson's activities. The vocabulary reading lesson-level pretest included the 10 selected words for the lesson. The lesson-level pretest involved speaking the word into a microphone. First, a selected word would appear on the computer screen. Next, the student would click on the record button (the students were taught how to use the record button to record their voices). Then, the student would say the selected word into the microphone on his or her individual headset. *The First 4000 Words* software included voice-recognition software to determine if the students were reading the word correctly. A 'sound board' provided visual feedback by indicating whether a word was pronounced correctly or incorrectly. If a selected word was pronounced correctly, the slider on the sound board would move up and the corresponding light would turn green. If a selected word was pronounced incorrectly, the slider on the sound board would move down and the corresponding light would turn red.

The interactive oral reading section involved listening and reading. Each story was read aloud twice to the student by the narrator. The text for each story was presented on the computer screen in a book format. Each text was approximately 200 words in length and incorporated each of the 10 selected words two to five times in the text. The text was written in approximately 20 point, black font, except for the selected words that were written in red. Each story included a title on the 'book cover' and was approximately four pages long.

As the narrator read the story aloud for the first time, the text would be underlined by sentence. The 'book' functioned like a book in which each page could be re-read. After reading the text for the first time, the students completed the interactive vocabulary activities. The interactive vocabulary activities involved practice of pronouncing all 10 selected vocabulary words and receiving immediate feedback via the voice-recognition software. First, the students would click on a selected word on the left side of the screen. A sentence from the just-read text that included the selected word would appear on the screen and the narrator would read the sentence. The selected word would appear in red as it originally did in the text. In addition, a picture would appear that illustrates the selected word. Then, the narrator would provide a verbal definition of the selected word. Next, the students would practice saying the word and record the word as they did during the vocabulary speaking pretest. At this point, the students had the option to play back their recording and re-record. After completing the interactive vocabulary activities, the students would re-read the story. During the second reading of the story by the narrator, individual sentences could be replayed by clicking on the individual sentence (see Figure 1). During the second reading, the selected words could also be reviewed by clicking on the word. Also, during the second reading, selected words that were missed on the pretest were highlighted and simple definitions for these words were provided (see Figure 2).

The vocabulary reading lesson-level posttest was identical to the lesson-level pretest and involved reading the selected vocabulary words into the microphone. If a student scored 9 of 10 or better on the pretest, he or she read the story only once and moved onto the next lesson.

Teacher Feasibility and Student Satisfaction Surveys. To assess program practicality and student impressions of the computerized vocabulary instruction, teacher feasibility surveys (Appendix C) and student satisfaction surveys (Appendix D) were developed. The teacher feasibility survey addressed questions about general impressions of the computerized vocabulary instruction and to what extent it would integrate into existing curricula. Items were rated using the following scale: 0=poor, 1=fair, 2=good, 3=excellent.

The student satisfaction survey was informal and administered verbally throughout the program administration. The survey asked students about impressions and experiences using the computerized vocabulary instruction program. Student responses were rated on the following scale: 1 = thumbs down, 2 = no opinion, 3 = thumbs up.

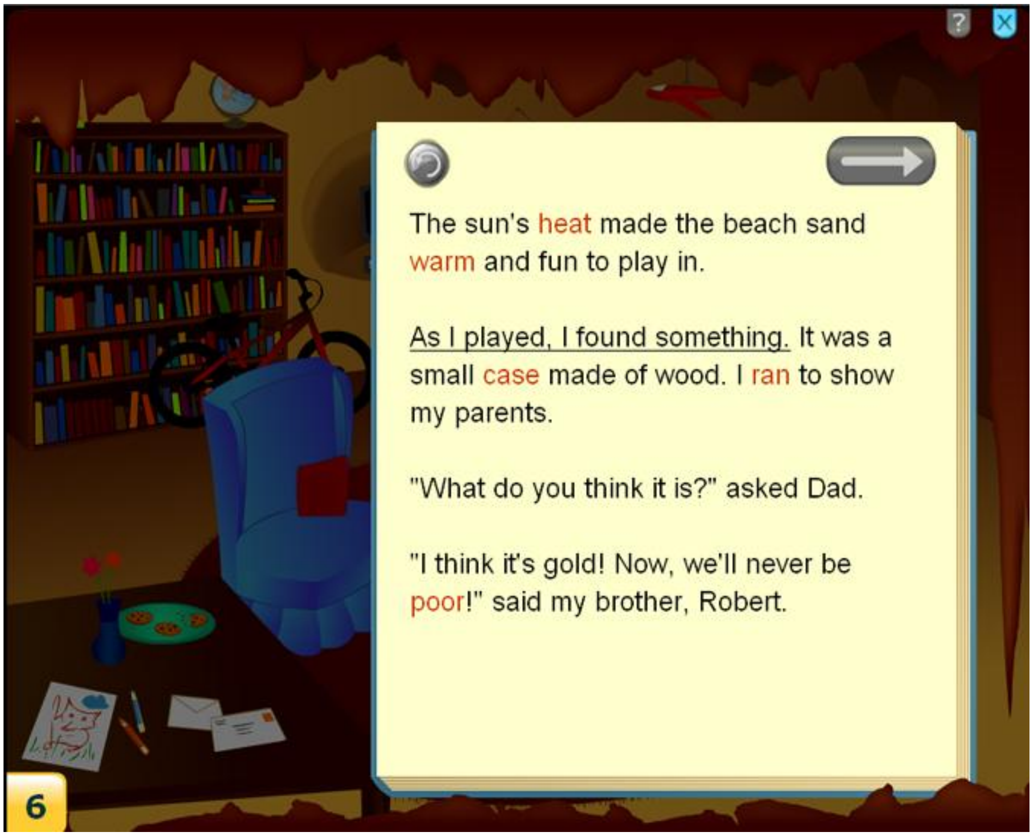


Figure 1. First reading of a story, with highlighted target words and sentence underlining.

Figure 2. Second reading of a story. Student has clicked the word "backward".

Procedure

All summer school students entering Grades 2-4 completed the 40-word vocabulary pretest. Students identified as having poor vocabulary skills were enrolled in the second phase of the study. Students who were randomly assigned to the experimental condition received the online vocabulary instruction in a pull-out session away from the regular classroom. These students were initially introduced to the vocabulary lesson procedures and computer log-in. Students typically spent approximately 15-25 minutes per session and worked independently through the lessons. The students were able to complete one to three lessons during a typical session. Two or three experimenters monitored student progress and provided technical assistance when a computer malfunctioned. Most students were able to complete the 10 lessons (100 words) within five to six sessions. These sessions occurred over the span of two weeks. After completing all 10 lessons, students completed the survey to assess their affective responses to use of the *4KW* software program. Students returned to their normal classroom for the remainder of the summer school session.

Students who were randomly assigned to the control condition remained in their normal classrooms and received the normal summer school curriculum. The typical summer school curriculum varied by grade, by classroom, and by day. However, the curriculum typically consisted of lessons geared toward remediation in reading and mathematics. Within one week after all students in the experimental condition had completed the computerized vocabulary lessons, all students in grades 2-4 were administered the posttest. The 40 words on the pre- and posttests were embedded within the 100 words targeted in the computerized vocabulary instruction. After the posttest was administered, the experimenters met with the teachers to administer the survey and participate in the brief group interview based on the survey questions.

Study 2

Research Design

This study was a randomized field trial. Students who were identified as potential study subjects and who met parental and student consent criteria were randomly assigned to receive either the online vocabulary intervention or “business as usual” classroom instruction. The use of random assignment potentially permits strong causal inferences to be drawn from differences in the performance of treatment and control subjects.

Setting. This study was conducted during the spring of 2009 in three elementary schools in a large Midwestern metropolitan area in the United States. The schools were selected to provide a wide range of socioeconomic, racial and ethnic participants. School 1 was located in a relatively affluent suburban area, serving predominantly white, middle to upper income families. School 2 was also suburban, but serving very diverse, mostly lower income families (>70% free/reduced lunch). School 3 was a charter school in an urban area, serving low income (>85% free/reduced lunch), predominantly African-American and Hispanic families.

Participants. In each of the three schools in this study, two classrooms from each of grades 1 through 4 were involved. All students in these 24 classrooms were group-administered two vocabulary pretests, described below. Students who scored lower than chance were eliminated from consideration, since previous research suggested they may not yet be ready to benefit from the *4KW* online instruction. The next eight lowest scoring students in each classroom were randomly assigned to receive either the intervention or normal classroom instruction. This random assignment resulted in 96 experimental and 96 control students, evenly distributed across classrooms, grades, and schools. Of the students in the experimental group, 57 were girls, 36 were ELL, 48 were eligible for free/reduced lunch, and 15 were identified by their teachers as students with special needs. The control group included 42 girls, 37 ELL students, 40 students eligible for free/reduced lunch and 12 identified as special needs students.

Materials

Measures. Two vocabulary tests, the Seward Listening Vocabulary Test and the Seward Reading Vocabulary Test, were administered as pretests and posttests to all students. Both of these multiple choice paper and pencil tests consist of a stratified random sample of 40 words selected from the *4KW* corpus of target words. The Listening Test asks students to select the correct image (out of four choices) that corresponds to a word read aloud by their teacher (see Appendix B). The Reading Test asks students to select the correct written word (out of four choices) that corresponds to a given image (see Appendix A). Both of these tests were found to have high reliability (Cronbach's *alpha* between .7- .9, depending on grade level).

As an additional posttest, first and second grade students were administered the Gates-MacGinite Level 2 Vocabulary Test, which is conceptually similar to the Seward Reading Vocabulary Test described above. Third and fourth grade students were administered the Gates-MacGinite Level 3 Vocabulary Test, which asks students to select the correct word (out of four choices) that best matches a short definition or synonym. The target words in these two standardized Gates-MacGinite tests share considerable overlap with the *4KW* corpus of target words (65% and 67%, respectively).

In addition to the paper and pencil tests described above, which serve as the principle dependent variables in the initial analysis of results reported here, the *4KW* software program automatically recorded students' daily performance at the word and lesson levels. These measurements allowed teachers to monitor student progress and, if needed, adjust student placement within the hierarchical framework of the program.

*The 4KW Intervention***.** The *4KW* program includes instruction on approximately 3700 words most frequently found in written English (Zeno et al., 1995). The words are arranged from most frequent to least frequent, and are divided into units, each of which consists of a number of ten-word lessons. The vocabulary instruction was administered to students individually at computer stations. Students used headsets to listen to instruction and actively respond to word tests and game-like activities that require integrating words and definitions.

The online setting is an animated world, the "Vocabitat", in which various animals guide instruction. Each ten-word lesson began with a pretest of the ten words. During the pretest students heard the target word and were asked to choose the picture (out of three choices) that best represented the meaning of the target word (see Figure 3). A green (correct) or red (incorrect) light provided feedback to the student. If a student scored 9 or 10 on the pretest she/he listened to a story that included those ten words and moved on to the next lesson. The text for each story was presented on the computer screen in a book format. Each text was approximately 200 words in length and incorporated each of the 10 selected words two to five times in the text. Each story included a title on the 'book cover' and was approximately four pages long. As the narrator read the story aloud for the first time, the text would be underlined by sentence. The ‘book’ functioned like a book in that each page could be re-read.

If two or more words were missed on the pretest, the story listening was augmented with simple definitions of the missed words, and a second reading of sentences incorporating those words. During the second reading of the story by the narrator, individual sentences could be replayed by clicking on the individual sentence (see Figure 1). During the second reading, the selected words could also be reviewed by clicking on the word. Also, during the second reading, selected words that were missed on the pretest were highlighted and simple definitions for these words were provided (see Figure 2). Following the second reading students engaged in a series of game-like exercises that focused on those words missed at the lesson pretest. The games required connecting target vocabulary items with their definitions, and “points” were awarded for speed and accuracy (see Figures 4 and 5). The story was then listened to a third time and a posttest of the ten words, identical to the pretest, concluded the lesson. An additional activity that focused on function words appeared after each fifth lesson. An optional component of the program allowed students to practice speaking target words into the headset microphone and to receive feedback mediated by voice recognition technology.

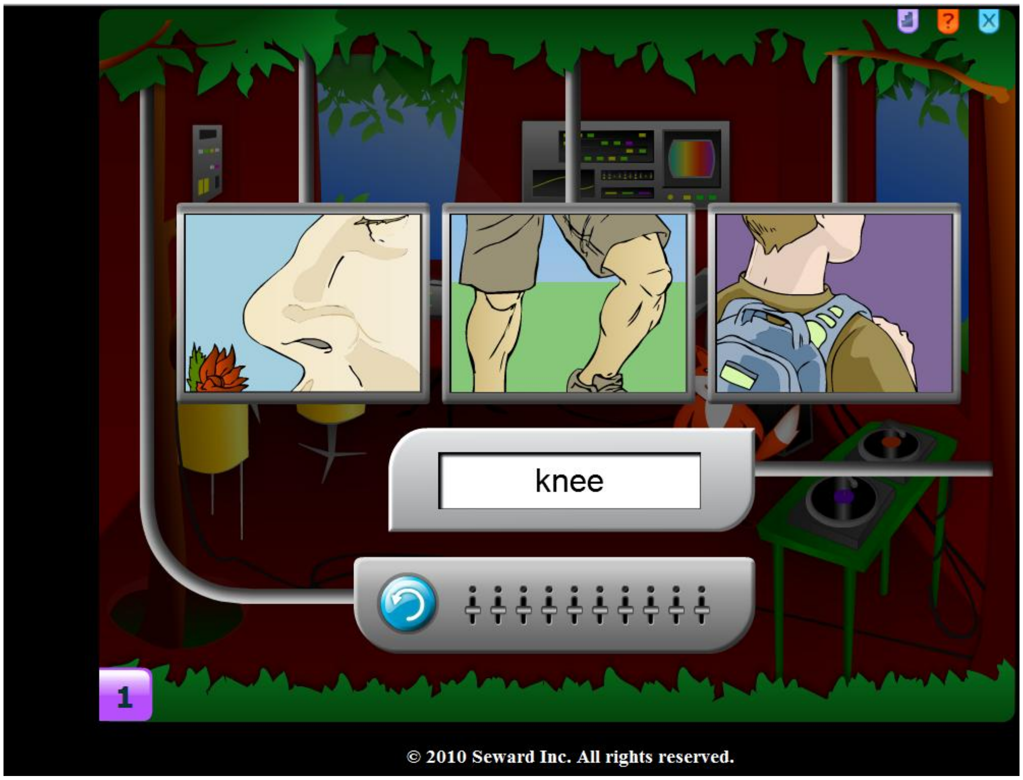


Figure 3. Lesson pretest/posttest: Student clicks on appropriate image.

Figure 4. “Submarine” vocabulary game: Student sinks sub by responding rapidly.

Figure 5. “Jigsaw” vocabulary game. Student uncovers image by recognizing definitions.

Procedure

Prior to beginning the intervention researchers met with teachers individually or in small groups to demonstrate the software program and familiarize them with the instructional lesson design. Teachers were also shown how to access student performance files and how to manipulate student placement within the overall program if necessary (i.e. students’ placement in the system could be adjusted if performance suggested words were too difficult or too easy).

The intervention plan called for treatment students to use the online vocabulary instruction for at least 30 minutes a day three times per week over the course of approximately 15 weeks. Teachers were responsible for incorporating these sessions within their classroom settings, and had discretion to do so at any point in the school day and on the days of their choice.

Teachers and researchers teamed to introduce the software program and instructional design to treatment students individually, ensuring that students became familiar with login procedures and basic tasks inherent in the lesson structure. Researchers conducted classroom observations of students on a weekly basis to assess student engagement and address teacher or student questions and concerns. These observation periods were also utilized to record activities in which control students were engaged during treatment sessions. Researchers collected data at predetermined time point intervals to assess student on-task engagement with the vocabulary program (see Appendix E for a sample observation form).

Treatment students were placed within the *4KW* unit/lesson hierarchy based on their pretest performance. Teachers monitored students' performance periodically and, if necessary, adjusted student placement within the hierarchical word corpus. The Seward and Gates-MacGinite posttests were administered within one week of the end of the 15-week intervention.

Analysis

The effect of the treatment on post-test measures of reading, listening, and the Gates- MacGinitie was determined using four separate one-tailed analyses of covariance (ANCOVA). Separate analyses were run for 1st and 2nd graders who took the Gates-MacGinitie-2 and 3rd and 4th graders who took the Gates-MacGinitie-3. Due to preliminary analyses showing significant associations between the dependent variables with grade, language background, and reading and listening pretest scores, these variables were used as covariates in all treatment vs. control analyses. In an effort to reduce teacher burden, the Gates-MacGinitie was not administered as a pretest. For Gates posttest analyses, the *Seward Reading Assessment* pre-test was used as a covariate. Approximately 60-70% of the words included on the Gate-MacGinitie are included in the *4KW* corpus, but due to individualization and limited duration of the instruction, a given student may not have encountered many of the target words. An α of .05 was used across all analyses in determining significance.

The *4KW* software program tracked all of the students' online activity. This allowed us to test, in addition to the treatment vs. control posttests described above, whether treatment students' posttest performance was better on the specific words which they missed at pretest and subsequently encountered during the intervention. A hierarchical generalized linear model (HGLM) was used to test whether student-level variables such as gender, ELL status, or grade level were significant predictors of effects of *4KW* exposure to target words. Chi-square analysis including all cases of words missed by each student on the pretest was also used to address this question.

Analysis of posttest performance by treatment and control students on words *not* encountered by treatment students was also conducted, in order to test for possible effects of "word consciousness" (Anderson & Nagy, 1992). Word consciousness has been described by various researchers as a peripheral effect of direct vocabulary instruction wherein students become more aware of words in their surroundings and therefore become more proficient in implicit word learning of non-target words.

The database created by the *4KW* software program was also used to assess fidelity of treatment. The data allowed for estimation of average time spent on individual lessons and provided specific information on the amount of time students spent using the *4KW* program.

Chapter IV

Results

Study 1

Our aim was to determine if students using the *4KW* program would gain knowledge of the meaning of words to which they were exposed, compared to students engaged in business-as-usual curricular activities. Table 1 shows mean pre- and post-test raw-scores and standard deviations for students assigned to experimental and control conditions by grade level and overall. An independent samples *t*-test shows that no statistically significant difference in pre-test scores exists between those assigned to experimental and control conditions in any of the three grade levels, or overall. This result suggests that randomization of subjects was successful in balancing prior vocabulary knowledge between the experimental and control conditions.

Table 2 reports mean gain scores (post-test score – pretest score) for each grade, boys, girls, and all students combined; *t*-test p-values and the respective effect sizes (Cohen's *d*) are also reported. Gain scores in the experimental groups were significantly higher than in control groups in grades 2 and 3, among boys, and overall. Gain score differences in 4th grade and among girls failed to reach statistical significance. The effect sizes for gain scores averaged more than one standard deviation overall.

Table 1.

Mean Pre-test and Post-test Scores by Grade and Experimental Group

Group	Test	Control		Experimental	
		Mean	SD	Mean	SD
Grade 2	Pre-test	10.6	3.2	10.9	5.5
	Post-test	10.1	3.5	18.3	9.1
Grade 3	Pre-test	21.7	7.3	23.3	6.7
	Post-test	24.3	8.6	30.5	5.6
Grade 4	Pre-test	19.5	2.5	18.3	4.1
	Post-test	24.3	9.5	27.3	9.0
Total	Pre-test	16.3	6.7	17.4	7.5
	Post-test	18.2	9.9	25.2	9.4

Note. SD = Standard Deviation

Table 2

Mean Gain Scores and Effect Sizes by Grade and Experimental Group

Group	Control		Treatment		Effect size
	Gain score	N	Gain score	N	
Grade 2	-0.4	9	7.4	8	1.7**
Grade 3	2,5	6	7.3	8	1.3*
Grade 4	4.8	6	9.0	6	0.6
Boys	0.0	8	7.9	15	1.5**
Girls	3.1	13	7.6	7	0.9
Total	1.9	21	7.8	22	1.1***

Note. *p < .05, **p < .01, ***p < .001; Effect size is Cohen's *d*.

Teachers had limited exposure to the *4KW* program, but expressed generally favorable views, and felt that the program might integrate well with existing curricula and classroom frameworks (See Table 3). Teachers' ratings of overall perception of the program yielded a mean = 2.9 (s.d. = .4) on a 0 to 3 scale. Teacher ratings of how the program would fit within existing curricula resulted in a mean = 2.8 (s.d. = .4). Teachers liked the direct interaction with words that students experienced with this instructional design, noting that providing definitions with contextual presentation of words is consistent with their currently practiced vocabulary instruction. Teachers also said that the software would integrate well into existing classroom literacy centers, especially given young students' increasing comfort with autonomous use of computers.

Table 3

Teacher ratings of 4KW instructional program

	Mean	Standard Deviation
Fits curriculum	2.8	.4
Overall perception	2.9	.4

Note. Scale*:* 0 = poor to 3 = good

Students' affective responses were also very favorable (See Table 4). Students' survey results indicated that they thought the lessons were fun (mean = 2.7), liked the stories that provided context for word learning (mean = 2.7), liked the graphics (mean = 2.8), felt that the program helped them learn new words (mean = 2.9), and expressed a desire to use the program more (mean = 2.8).

Table 4

Student ratings of the 4KW instructional program on 1 = poor to 3 = good scale

	Mean	Standard Deviation
Lessons were fun	2.7	.7
Liked the stories	2.7	.6
Liked the graphics	2.8	.6
Helped learn new words	2.9	.5
Would like to use more	2.8	.6

Study 2

Our original aim was to test whether use of the *4KW* computer program resulted in improved performance on standardized tests of vocabulary knowledge when compared to business-as-usual classroom instruction. Table 3 presents descriptive statistics of the unadjusted means and standard deviations of scores on the pre-test and post-test across all measures. In general, no differences emerged on overall posttest scores between treatment and control students. A number of factors may have contributed to the lack of a treatment effect on posttest scores. Treatment fidelity was inconsistent at best, as described below. The *4KW* lesson-level pretest design also resulted in students not receiving explicit instruction on approximately a third of the words they didn't know, since they would guess correctly one third of the time. These factors lead us to ask whether treatment group students performed better on the posttest on words which they actually did encounter during use of the *4KW* program. The overall posttest score results are reported below, followed by a description of the word-level results.

Seward Reading Vocabulary Post-Test. An ANCOVA was conducted to determine the effect of treatment on the reading vocabulary post-test. Language background and reading pre-test both accounted for a significant amount of variance in the dependent variable. After adjusting for the effects of the covariates, the results of the analysis indicated no significant difference between the treatment and control groups on the reading post-test $F(1,178) = 1.40$, $p = .123$. The adjusted means for the reading vocabulary post-test were 29.0 and 29.7 for the control and treatment group, respectively.

Seward Listening Vocabulary Post-Test. An ANCOVA revealed that a significant proportion of variance in listening vocabulary post-test scores could be accounted for by listening vocabulary pre-test, language background and grade level. However, after adjusting for the covariates, there was no main effect of condition on listening post-test scores, $F(1,178) = 1.59$, $p = .145$. The adjusted means for the listening post-test were 30. 2 and 29.6 in the control and treatment group, respectively.

Gates MacGinitie-2. Results of the ANCOVA showed that grade and pretests accounted for a significant proportion of variance in scores on the Gates MacGinitie-2. After adjusting for the effects of the covariates, there was no significant difference favoring the treatment for first and second grade students on the Gates MacGinitie-3, $F(1,84) = 4.61$, NS. The adjusted post-test means were 20.8 and 18.8 for the control and treatment condition, respectively.

Gates MacGinitie-3. The results of the ANCOVA show that language background and pretest scores accounted for a significant proportion of variance in third and fourth grade scores on the Gates MacGinitie-3. The results of the ANCOVA show that after adjusting for the effects of the covariates, there was a significant effect of condition favoring the treatment group for third and fourth grade students on the Gates MacGinitie-3, $F(1,82) = 3.043$, $p = .043$. Adjusted post-test means were 23.9 and 26.0 for the control and treatment condition, respectively.

Word-level analysis. A post-intervention analysis of the data collected by the *4KW* software program showed that during the intervention the average treatment student was exposed to only 3.4 words that she/he missed on the paper and pencil pretests, a number too low to expect significant differences on posttest total scores. This was in part due to poor treatment fidelity (described below). This finding motivated analyses at the word-level to determine if treatment students performed better on posttest items if they had encountered them during the intervention. A Chi-square analysis using all cases of words missed by treatment students on the paper-and-pencil pretests (n = 2561) was performed, crossing exposure to the word during the intervention with posttest performance. The results showed that exposure led to improved posttest performance (See Table 6). When students did not encounter missed words during the intervention

Table 5

Unadjusted means and standard deviations by group

Measure	Control		Treatment	
	Mean	S.D.	Mean	S.D.
Reading vocabulary				
Pretest	25.1	9.3	26.0	8.9
Posttest	28.6	7.6	30.0	7.9
Listening vocabulary				
Pretest	26.6	4.7	26.3	4.7
Posttest	30.0	4.5	28.8	6.9
Gates-MacGinitie 2				
Pretest	-	-	-	-
Posttest	20.4	9.3	19.2	8.7
Gates-MacGinitie 3				
Pretest	-	-	-	-
Posttest	23.7	9.8	26.2	10.9

they were successful on those words less than half of the time at posttest. Students who were exposed to missed words during the intervention were successful at posttest nearly two thirds of the time, and this difference was highly significant (p < .001).

“Exposure”, as it is used here and in subsequent analyses, means that the student completed the lesson that contained that word. The student was pretested on the word and listened to the word used in story context at least once, and may or may not have had further explicit instruction on the word.

Table 6

Chi Square results crossing posttest item outcomes with 4KW word exposure

		Posttest item outcome	
		Items correct (%)	Items incorrect (%)
4KW exposure	Yes	215 (64%)	119 (36%)
	No	1052 (47%)	1175 (53%)

The Chi-square results described above motivated an HGLM approach to modeling student performance at the word level. That is, a model with word items nested within students. Treatment and control students’ performance did not differ significantly on posttest items to which treatment students were not exposed (*t* = -1.20, df = 4771, p = .23). Due to the low frequency with which treatment students were exposed to words found on the paper and pencil posttests, Seward posttest results were combined. Gates-MacGinitie posttest results were not included in the word-level analyses, since they were not administered as pretests and analyses were based on items that students missed on the paper-and-pencil pretests.

Dummy coding for the 80 words found on the integrated Seward posttests was used to create an unconditional Level 1 model with posttest item outcomes from every student as the dependent variables. This Item Response Theory (IRT) approach provided item difficulty estimates and an estimate of the mean vocabulary ability of the posttest respondents. The item difficulty estimates suggest that although the *4KW* word corpus is arranged in hierarchical fashion by word frequency, a consistently linear progression of easier to more difficult items does not follow from that hierarchy. This is likely due to a variety of factors including variability in the ease with which some words are visually represented in drawings compared to others, the degree to which drawings accurately represent word meanings, and the fact that the word frequency hierarchy is based on written English samples that are not limited to reading materials that children would likely encounter (Zeno et al., 1995).

The HGLM models included all cases wherein, for a given item, the paper and pencil pretest outcome was incorrect and a paper and pencil posttest item outcome was available. The HGLM model here is defined at two levels, with item and treatment variables incorporated at Level 1 and random effects related to students at Level 2.

If π_{ij} is defined as the probability of person *i* correctly responding to item j, then the logit $\eta_{ij} = ln\,[\frac{\pi_{ij}}{1-\pi_{ij}}]$, and the equation for Level 1 in the model becomes

$$\eta_{ij} = \beta_{0i} + \beta_{1j} + \beta_{2j}y_{ij} \qquad (1)$$

where β_{0i} represents a random effect of student vocabulary ability, β_{1j} represents a fixed effect of difficulty for item j, and β_{2j} represents a fixed effect for exposure to item *j* with y_{ij}= 0 if student *i* was not exposed to item *j* and y_{ij} = 1 if student *i* was exposed to item *j* during the intervention. Consistent with the Chi-square findings referenced above, when entered as a fixed effect, "exposed" was a highly significant positive predictor of posttest item success (z = 3.87, p < .001). The fixed effect, .52, suggests that students exposed to a word during the 4KW intervention were 1.68 [exp(.52) = 1.68] times more likely to answer correctly to that item on the paper and pencil posttest than if they were not exposed. When entered as a random effect, "exposed" was similarly positive and significant (z = 3.89, p < .001). The fit of the models with "exposed" as either a fixed or random effect, did not differ significantly, Chi square (2 degrees of freedom) = 1.82, p = .40. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were lower for the fixed effect model than for the random effect model suggesting that the fixed effects model was a better fit, and the lack of significant difference between the models confirms the conclusion that the fixed effect model is the better model (See Table 5). As such, the Level 2 models were limited to modeling of the random effect of student ability, β_{0i}. Student (Level 2) variables, including treatment vs. control assignment, grade, gender, ELL status, Special Education status, and SES status were then incorporated into various HGLM models; results are described below.

ELL status ($z = -4.42$, $p < .001$), SES status ($z = -2.97$, $p < .01$), Special Education status ($z = -2.43$, $p < .05$) and grade levels 1 and 2 ($z = -5.20$, $p < .001$ for grade 1; $z = -2.57$, $p = .01$ for grade 2; $z = -.84$, $p = .40$ for grade 3, with grade 4 as a referent) were demographic variables that reached significance as individual predictors of item-level outcomes. When entered into the model together, only ELL and grades 1 and 2 remained significant. The Level 2 model was then defined as

$$\beta_{0i} = w_{00} + w_{ell}z_{1i} + w_{grd1}z_{2i} + w_{grd2}z_{3i} + u_{ij} \quad (2)$$

where w_{00} is an intercept representing the overall mean student ability level, w_{ell} represents the fixed effect of ELL status, $z_i = 0$ if student *i* is not identified as an ELL student, $z_i = 1$ if student *i* is identified as an ELL student, w_{grd1} represents the fixed effect of being in grade 1 with $z_{2i} = 1$ if student *i* is in first grade and 0 otherwise, w_{grd2} represents the fixed effect of associated with being in grade 2 with $z_{3i} = 1$ if student *i* is in grade 2 and 0 otherwise, and u_{ij} is a residual term. The negative fixed effect associated with identification as an ELL student, -.47, suggests that ELL students were 1.58 times less likely to respond correctly to posttest items to which they had not been exposed. In effect, exposure to a word during the intervention would raise an ELL student's odds of answering that item correctly on the posttest to the level of a non-ELL student who had not been exposed to the word during the intervention. This result is consistent with well-documented research described earlier suggesting that ELL students often face vocabulary knowledge deficits. The negative fixed effects associated with being in grades 1 and 2 suggest that those students were less likely to respond correctly on the posttest to items they had missed on the pretest, in comparison to students in grade 4.

When incorporated as a student-level variable, school was also a significant predictor of posttest item outcome. With the relatively affluent School 1 as the referent, the less affluent urban School 3 ($z = -4.70$, $p < .001$) was associated with a negative effect on posttest item performance. “School” as a Level 2 correlates with SES and ELL status in the model; when entered together the ELL effect was no longer significant and the *p*-value of the school effect was greatly reduced but was still significant. “School” was not incorporated into the final model, however, as it correlates with SES (.76) and ELL (.58) highly, and is intuitively less descriptive than those demographic variables. Consistent with the lack of overall differences on posttest total scores, assignment to treatment or control groups was not a significant Level 2 predictor ($z = .71$, $p = .48$). That is, being in the treatment group led to no advantage or disadvantage on posttest vocabulary items to which treatment group students were not exposed during the intervention. Gender was not a significant Level 2 predictor ($z = .18$, $p = .89$). The full model thus included “exposure” as a Level 1 predictor (Equation 1), and ELL, and Grades 1 and 2 as Level 2 predictors (Equation 2) explaining variance in the random effect of theta from Level 1 (see Table 7 for fit statistics and Table 8 for parameter estimates).

Table 7

Fit Statistics for Unconditional, Fixed Effects, Random Effects and Full Models

	AIC	BIC	degrees freedom
Unconditional	6499.8	7030.3	81
Fixed effects	6487.0	7024.1	82
Random effect	6489.5	7039.7	84
Full model	6470.3	7027.1	85

Note. AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.

Student Engagement. Assessment of student engagement while using the *4KW* program was conducted over the course of Study 2. Working individually, two researchers conducted observations in classrooms on a rotating basis before and during student use of the *4KW* program on teacher-scheduled days. Observers posed as classroom aides, moving around classrooms engaging and observing all students, so as to remain unobtrusive to students using the *4KW* program. At predetermined time points (approximately five minute intervals) observers turned attention to *4KW* users for 3-5 seconds and judged whether students were appropriately engaged in program tasks. Researchers also noted the specific *4KW* program tasks with which students were engaged during these time points.

Table 8

Unconditional and Full Model Parameter Estimates

Unconditional model		
Variable	β	Standard error β
Theta/intercept	-.37	.16

Full model		
Variable	β	Standard error β
Theta/intercept	-.09	.18
Exposure	.52***	.14
ELL	-.47***	.11
Grade 1	-.70***	.18
Grade 2	-.34**	.18
Assignment to treatment	.08	.11

Note. **$p < .01$, ***$p < .001$.

Students were judged to be properly engaged at 91% of the pre-determined time points (N = 687) collected by two researchers. Inter-rater reliability was .94. The high level of student engagement was consistent across schools and throughout the course of the 14 week intervention (89% on-task during the last four weeks of the study). Of the eight researcher-specified activities within the instructional program (pretest/posttest, story-listening I, II, III, puzzle game, matching game, submarine game) students were most likely to be rated as off-task while listening to the lesson stories. Although students were highly engaged during all lesson tasks, 90% of time points wherein students were judged to be off-task occurred during the more passive, story-listening components of the program (See Table 7). Discriminating between the three story-listening tasks in this regard was not feasible, since during many lessons students' pretest performance dictated that they hear the story only once. Only one instance of student off-task behavior was recorded during the game-like instructional activities in the program, and students were rarely judged to be off-task during lesson pretests or posttests.

Table 9

Student task engagement during use of the 4KW program

Task	Time points observed	On task %
Lesson pre/posttest	241	96%
First story read along	152	82%
Second/third read-along	132	80%
Game-like tasks	162	99%
Overall	687	91%

Fidelity of Treatment. The intervention schedule called for treatment students to spend 30 minutes per day, three times per week using the *4KW* program, over the course of 15 weeks. Specific times of use were at the teachers' discretion. Analysis of average student progress through *4KW* lessons suggests that, conservatively, students should have completed 130-140 lessons given the prescribed schedule. In fact, students completed less than half of that number of lessons, approximately 58 on average. Significant technical difficulties at Schools 1 and 2 contributed to this low treatment fidelity. School 1 students averaged 117 lessons completed, School 2 students averaged 46 lessons completed, and School 3 students averaged 57 lessons completed. The intervention schedule called for students to use the 4KW program on approximately 45 days during study, but students' use of the program averaged only 17-18 days in all 3 schools. As a result of these fidelity issues treatment students were, on average, exposed to only 3.5 words that they had missed on the paper and pencil pretests, which was approximately 13% of the words the average student missed on the pretest.

Chapter V

Discussion

Study 1

The purpose of this study was to test whether students' use of the *4KW* vocabulary program resulted in improved knowledge of word meanings sufficient to affect test scores. The results suggest that the *4KW* program can successfully teach target words and improve vocabulary scores. The goal of this study was not to test whether the *4KW* instruction is superior to some other instructional design for vocabulary learning, since the computer adaptive aspect of the *4KW* program provides efficiency and individualized instruction that allows for instruction on a much larger corpus of words than is reasonably feasible otherwise.

The *4KW* program includes instructional components that have been repeatedly reported to be essential for effective student learning of vocabulary. Among these are repeated exposure to words (Stahl & Fairbanks, 1986; Nagy, 1988; Graves, 2006), examples of target words used in context (Stahl & Fairbanks, 1986; Biemiller, 2004), provision of definitions in simple language (Stahl & Nagy, 2006), and a focus on words that students are likely to encounter frequently (Nagy, 1988; McKeown & Beck, 2004). Instruction is multimedia in nature, with pictorial cues accompanied by auditory definitions (Mayer & Moreno, 1998), and read-alongs provide opportunities to listen and re-listen to fluent use of words in context (Hulstijn, 2003). Although the *4KW* program implemented in Study 1 was a simplified version, it contained these key elements to effective instruction. Additional game-like activities designed to reinforce the relationship between words and semantic representations and to increase speed of lexical access were incorporated into the *4KW* program in Study 2. These activities increase the number of encounters with target words and focus on words missed at the lesson pretest, capitalizing on the computer adaptive design of the program.

It was encouraging that all three grade levels in Study 1 showed vocabulary gain scores higher for subjects in the treatment condition compared to control groups, although the increase was not statistically significant in 4^{th} grade. Similarly, although effect sizes were large for both boys and girls, the overall gain score improvements for girls in the treatment group did not reach significance. Determining whether a gender difference exists with relation to student learning using the *4KW* program requires a larger sample size, and is addressed in Study 2. In addition, although students involved in this Study 1 included ELL learners, the number of ELL learners was insufficient for statistical analysis. Overall, however, the large improvements in mean post-test scores and effect sizes associated with assignment to the experimental condition provide reason for optimism.

A number of caveats limit the inferences that might be drawn from Study 1. We have little knowledge of the activities in which control group subjects were engaged during treatment sessions. The summer school setting is academically less rigorous in general. It is likely that control students were receiving little direct vocabulary instruction, particularly given the extremely limited amount of explicit vocabulary instruction that is currently delivered in U.S. schools in general (McCutchen, Green, Abbott & Sanders, 2009). Second, vocabulary instruction was not uniformly administered. Duration of sessions varied from 10-30 minutes depending on the teacher whose students were involved on any given day. In addition, how learning the words in this program might generalize to standardized vocabulary tests that include words not explicitly taught in *The First 4000 Words* is unknown, but is addressed in Study 2. We would expect significantly smaller effect sizes when vocabulary assessments include a smaller percentage of words directly targeted by instruction.

The vocabulary instruction program utilized in Study 1 represented only a part of the fully developed *First 4000 Words* program, but still led to significant effects on vocabulary knowledge. Study 2 tested additional possibilities suggested by Study 1, as it included a larger number of subjects, a more complete *4KW* software program to allow for greater individualization of instruction, longer experimental timeframe, and use of standardized vocabulary measures.

Study2

The purpose of Study 2 was to determine whether the *4KW* computerized vocabulary program, shown to be effective in teaching targeted vocabulary words in Study 1, could be scaled up to have an effect on student performance on standardized measures of vocabulary. We viewed this as a logical step toward the larger goal of producing effects on reading comprehension. Although the results of Study 2 indicate that third and fourth grade students who participated in the *4KW* program preformed significantly better on the Gates-MacGinitie-3 vocabulary posttest than students in the control group, effects were weak and significant effects favoring the treatment group were not found for the Seward post-test measures of reading or listening vocabulary, or for first and second grade students on the Gates-MacGinite-2.

The discrepancy between the strong, positive results of Study 1 and the insignificant effects on posttest scores in Study 2 were not completely unexpected, and led to the word-level analysis of student performance detailed in the Results and discussed below. Scaling up the *4KW* intervention and still detecting effects using standardized measures of vocabulary knowledge was challenging in that the number of words a student needed to learn in order to have an expected 1 point improvement from pre-test to post-test increased proportionally as the size of the curriculum increased. The 100-fold increase in curriculum size required that a student learn a far greater number of words in a relatively short amount of time in order to produce a detectable effect of treatment within the time frame of the study. Whereas 40% of target words taught in Study 1 appeared on the posttest, only about 1% of the words potentially taught in Study 2 would appear on any of the standardized or Seward vocabulary posttests. That is, with a large increase in the corpus of target words a vocabulary intervention attempts to teach there is a corresponding decrease in the percentage of target words that will appear on an outcome measure, assuming the number of items on the outcome measure remains constant. This reduction in posttest sensitivity to learning suggests that an intervention must be time-intensive and conducted with high treatment fidelity to produce effects on standardized vocabulary tests when the curriculum domain is large.

Although Study 2 was designed to run for an extended period of time (15 weeks), several factors in this study impeded students' progress. The intervention schedule predicted approximately 45 days when students would use the program, but students averaged only 17-18 days of use. It also appears that the average time spent during sessions when students did use the program was less than the 30 minutes called for in the intervention schedule, although determining exact session durations was not possible. These issues precluded treatment students from being exposed to a sufficient number of words to affect overall posttest scores.

Significant technical difficulties encountered in Schools 1 and 2 contributed to the lack of fidelity both directly and indirectly. During the course of the intervention these schools experienced a significant time period for which they had limited or no computer access. School 1 experienced a computer virus that led to a complete district-wide shutdown of all computers for three weeks of the intervention. School 2 was equipped with insufficient internet bandwidth for the first half of the intervention, which resulted in frequent disruptions to the *4KW* program and required tedious repetitions of logins. Teacher and student frustration with technical issues certainly and understandably caused some teachers to reduce the time they committed to the intervention schedule. As a result of these issues, students completed considerably less than half the lessons that would have been expected had the intervention schedule been followed with high fidelity. In addition, it is likely that insufficient effort was given to preparing teachers for their roles in the experiment. Although teachers were initially informed of and encouraged to utilize *4KW* managerial software features, which provide user-friendly mechanisms for tracking student performance and adjusting placement in the lesson hierarchy accordingly, teachers did not use this feature until explicitly requested to do so during the study. Since some students were initially placed into a level that was lower than their actual ability, as shown by series of lessons wherein students demonstrated knowledge of all or most of the words on lesson pretests, many students spent time going through lessons below their ability level, decreasing the amount of time they had to learn new words. Timely use of the managerial software features might have alleviated this problem.

Word-level analyses. Despite the logistical challenges and limited duration of Study 2 referenced above, HGLM and Chi-square word-level analyses showed that students benefited from exposure to vocabulary items while using the *4KW* program. The analysis of student performance at the word-level was motivated by researchers' awareness of the difficulty of detecting growth in vocabulary knowledge using standardized tests of vocabulary (National Reading Panel, 2000), and was made possible by features of the *4KW* software program that closely monitor and record students' progression through the lesson hierarchy. At the word-level both HGLM and Chi-square analyses demonstrated that, despite the lack of statistically significant effects on overall posttest vocabulary scores, treatment students who missed words on the paper-and-pencil pretests were approximately 1.7 times as likely to get those items correct on the posttest if they had encountered those words during the *4KW* intervention. This shows that students gained knowledge of word meanings from use of the *4KW* program and that this learning was retained over a period of weeks. This finding provides support for the hypothesis that the lack of overall improvement by students in the treatment condition on standardized vocabulary measures was a result of insufficient exposure to *4KW* instruction during the study. Further support for this hypothesis comes from data showing the very inconsistent fidelity with which the intervention was implemented by teachers during the study, described above.

The HGLM approach to modeling students' posttest performance on vocabulary items to which they had or had not been exposed during the intervention allowed for analysis of questions arising from Study 1 and the role of a variety of student demographic variables. Study 1 results had suggested differing effects of *4KW* instruction by gender and grade, although the small sample sizes in Study 1 made those conclusions tentative. The word-level analysis in Study 2 found no difference in effects of *4KW* instruction by gender, supporting the supposition that small sample size was important in the differences seen in Study 1. Similarly, Study 1 results suggested that effects of program use differed by grade, with 4th grade students showing insignificant effects from *4KW* instruction. Study 2 showed that the positive effects of word exposure during the intervention were consistent across grades.

Students in the treatment condition performed no better than their control group counterparts on posttest items to which they were not exposed during the intervention. This suggests that positive effects on vocabulary learning from enhanced "word consciousness" (Anderson & Nagy, 1992) as a result of explicit vocabulary instruction were absent or not at detectable levels during this experiment. "Word consciousness" has been posited as a potential benefit of explicit vocabulary instruction. Despite a lack of empirical support in the research literature, proponents theorize that explicit focus on vocabulary learning might lead students to be more cognizant of new words they encounter in their environment, leading to enhanced vocabulary growth. Although the results of Study 2 do not support this supposition, it is possible that better integration of the *4KW* instruction with other literacy-related classroom activities and more overt encouragement from teachers in this regard might have such effects.

As described in Chapters 1 and 2, ELL and low SES status are known to be key factors that put students at risk for inadequate English vocabulary knowledge and attendant academic difficulties. The HGLM analyses suggest that all student groups identified in Study 2 benefitted on posttest items from exposure to those items during *4KW* instruction, but ELL students who were exposed to words during the intervention were only as likely to respond correctly to those words on the posttests as non-ELL students who were not exposed. It is likely that ELL students have English language skills that are underdeveloped relative to their peers.

Overall, students exposed to words they had missed on the paper and pencil pretests were 1.7 times as likely to respond correctly to those items on the posttest overall than were students who had not been exposed (holding vocabulary ability constant). AIC and BIC model-fitting results, which balance the reward of increases in variance accounted for with the cost of incorporation of extraneous variables, suggest that the best-fitting model fitted in the HGLM analysis conducted here included ELL status as the only Level 2 predictor from the aforementioned demographic variables collected.

Conclusions. The First 4000 Words program represents a thoughtful and well-designed attempt to combine computer-adaptive technologies with theoretically driven, pedagogically sound vocabulary instruction towards the goal of addressing an important educational need. Instructional tasks are research-based, and reflect well-established principles in vocabulary learning. The animation is bright, attractive, and students are typically highly engaged when using the program. Experimental results from Study 1 and Study 2 demonstrate that students gain knowledge of word meanings when using the *4KW* program.

There are weaknesses in the *4KW* program, however, that might be addressed in future iterations of the program. On the lesson pretests, only three alternative representations are available for students to choose from after hearing the target word (see Figure 4), and there is no option for students who don’t know the word meaning except to guess. This means that students who don’t know a word meaning will guess correctly 33% of the time, and subsequently receive no explicit instruction on that word. The fact that students don’t receive further, explicit instruction on a third of the words they encounter and don’t know is problematic. Future modifications might include increasing the number of representations students have to choose from to decrease this percentage, or adding an option that allows students to indicate that they don’t know the meaning of a word. In addition, the repeated reading/listening to lesson stories may be too passive, as suggested by observational data showing most off-task behavior occurred during these lesson components. Repeated reading and shared reading, which the story listening tasks are designed to model, are powerful strategies shown in the research literature to have beneficial effects on literacy. However, in traditional settings they are typically augmented with teacher/student questioning or other interactions. Adding some tasks that require active student participation during the story listening phase itself (rather than after the story), such as a cloze task or multiple-choice questioning, might increase active student engagement.

Lastly, although a great deal of effort has been put into generating the images that serve as visual representations of word meanings in the *4KW* program (including words that do not lend themselves easily to visual representation), ongoing efforts should be made to augment these representations. There is concern that students may identify new words strictly with the corresponding image, which is seen repeatedly during lessons, rather than inferring meaning from the image. This concern is supported by personal observation that some lesson posttest items can be answered successfully in the absence of hearing the word, but rather by visual recognition of a recently familiarized image. Expanding the number of images available to represent each vocabulary item would be an extensive undertaking, but would address this concern and might lead to even better understanding of word meanings as students integrate multiple representations of them.

Future Studies. Study 1 demonstrated the effectiveness of *4KW* in improving students' vocabulary knowledge for words that they were directly taught in the program (Fehr et al., in press). The significant difference found in the current study between the treatment and the control group on the Gates-MacGinitie-3 and, more importantly, the results of the word-level analyses conducted here provide support for the program's potential to be scaled up and have an effect on student's scores on standardized tests of vocabulary. However, the fact that this effect was not found for the other three post-test measures suggests that a more sustained intervention is likely required to produce robust intent-to-treat effects. If teachers actively monitor student performance, adjust student placement accordingly, and work to integrate *4KW* word-learning into other classroom activities such as reading and writing, it is likely that benefits to student use of the program can be enhanced.

The ultimate goal of *4KW* is to produce effects on reading comprehension, thus future research should include measures of reading comprehension as outcome variables. Whether use of the *4KW* vocabulary program over an extended period of time can lead to positive effects on measures of reading comprehension remains to be determined and is a focus of planned research. Future research will include measures to assess reading comprehension, but neither Study 1 nor Study 2 attempted to measure students' reading comprehension before or after the intervention.

In conclusion, although not all of the effects of the *4KW* intervention reported here are robust, there is reason for optimism. The challenge of detecting treatment effects when such an intervention is scaled up will likely require a more sustained treatment period than was implemented here. Given the logistical difficulties that undermined treatment fidelity during this study, it is not surprising that the results on posttest scores showed little or no difference between treatment and control students.

It is unrealistic to suggest that computer-delivered vocabulary instruction can be the sole vehicle for remediation of significant vocabulary deficits or ELL vocabulary learning; the roles of social discourse in the classroom and extensive reading of appropriate level texts will remain indispensable and provide the preponderance of opportunities for implicit word learning. It would seem, however, that efficiencies available through computer-adaptive technology, when combined with theoretically driven and pedagogically sound software programs, may be a very useful tool for introducing students to words and word meanings as part of a comprehensive approach to vocabulary deficit remediation and ELL vocabulary learning. When embedded within a learning environment emphasizing a robust and metacognitive approach to language learning, computer adaptive vocabulary instruction may play an important role in helping at-risk students' literacy achievement (Kojic-Sabo & Lightbown, 1999). The results of the two experiments with students using the *4KW* program documented in this thesis suggest that the program has potential for addressing the need for remediation of students' vocabulary deficits. More extensive interventions will be required to determine if successful remediation of students' vocabulary deficits can lead to better reading outcomes.

References

Abraham, L. (2008). Computer-mediated glosses in second language comprehension and vocabulary learning: A meta-analysis. Computer Assisted Language Learning, 21, 199–226.

Anderson, R. & Nagy, W. (1992) The vocabulary conundrum. *American Educator 16(4*), 14-18, 44-47.

Anderson, R. & Freebody, P. (1981) Vocabulary Knowledge. In J.T. Guthrie (ed) *Comprehension and Teaching: Research Reviews* (p 77-117). Newark, DE: International Reading Association.

August, D., Carlo, M., Dressler, C. & Snow, C. (2005) The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice 20(1)*, 50-57.

Baker, S., Simmons, D. & Kameenui, E. (1995) *Vocabulary acquisition: Synthesis of the research.* National Center to Improve the Tools of Educators, funded by the U.S. Office of Special Education Programs.

Beck, I., Perfetti, C. & McKeown, M. (1982) Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology 74(4)*, 506-521.

Beck, I. & McKeown, M. (2007) Increasing Young Low-Income Children's Oral Vocabulary Repertoires through Rich and Focused Instruction. *The Elementary School Journal* 107(3), 251-273.

Beck, I., McKeown, M., & Kucan, L. (2008). Creating robust vocabulary: Frequently asked questions and extended examples. New York: The Guilford Press.

Becker, C. (1977) Teaching reading and language to the disadvantaged – What we have learned from field research. *Harvard Educational Review 47*, 518-543.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*, 183–200.

Bialystok, E. (2007) Acquisition of Literacy in Bilingual Children: A Framework for Research. *Language Learning 57(1)*, 45-77.

Biemiller, A. (2004). Teaching vocabulary in the primary grades: Vocabulary instruction needed. In J. Bauman & E. Kame'enui (Eds.), *Vocabulary instruction: Research to practice* (pp. 28–40). New York, USA: The Guilford Press.

Bornstein, M., Haynes, O. & Painter, K. (1998) Sources of child vocabulary competence: a multivariate model. *Journal of Child Language 25(2*), 367-394.

Bowey J. (2001) Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, 22(3), 441– 469.

Cain, K. & Oakhill (1999) Inference making ability and its relation to comprehension failure in young children. *Reading and Writing 11*, 489-503.

Calvo, M. (2005) Relative contribution of vocabulary knowledge and working memory to elaborative inferences in reading. *Learning and Individual Differences 15(1),* 53-65.

Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*, 183–198.

Cunningham, A. & Stanovich, K. (1997) Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology 33(6),* 934-945.

Dapretto, M., & Bjork, E. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. *Child Development*, *71(3),* 635–648.

Draper, A. & Moeller, G. (1971) We think with words (therefore, to improve thinking, teach vocabulary). *Phi Delta Kappan, 52(8),* 482-484.

Duncan, G. & Magnuson, K. (2005) Can family economic resources account for racial and ethnic score gaps? *Future of Children, 15(1),* 35-54.

Ehri, L., Rosenthal, J. & Paterson, W, (2007) Spelling of words: A neglected factor of vocabulary learning. *Journal of Literacy Research 39(4),* 389-409.

Elleman, A., Lindo, E., Morphy, P., & Compton, D. (2009) The impact of vocabulary instruction on passage-level comprehension of school age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2(1*), 1-44.

Farkas, G. & Beron, K. (2004) The detailed trajectories of oral vocabulary knowledge: differences by class and race. *Social Science Research, 33(3),* 464-497.

Fernald, A., Perfors, A. & Marchman, V. (2006) Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42(1),* 98-116.

Francis, D. & Vaughn, S. (2009) Effective practices for English language learners in the middle grades: Introduction to the special issue of *Journal of Research on Educational Effectiveness. Journal of Research on Educational Effectiveness 2(4),* 289-296.

Ganger, J. & Brent, M. (2004) Re-examining the vocabulary spurt. *Developmental Psychology, 40(4),* 621-632.

Gathercole S., Willis C., Emslie H. & Baddeley, D. (1992) Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887– 898.

Goldenberg, C. (2006) Improving achievement for English learners: What the research tells us. *Education Week 25(43),* 34-36.

Goldfield, B. & Reznick, S. (1990) Early Lexical Acquisition: Rate, Content, and the Vocabulary Spurt. *Journal of Child Language, 17 (1*), 171-83.

Graves, M. & Prenn, M. (1986) Costs and benefits of various methods of teaching vocabulary. *Journal of Reading* 29(7), 596-602.

Graves, M. (2006) *The Vocabulary Book: Learning and Instruction*. New York, USA: Teachers College Press.

Graves, M. (2008) Instruction on individual words: One size does not fit all. In A.E.Farstrup & S.J.Samuels (Eds). *What Research Has to Say About Vocabulary Instruction*. Newark, DE, International Reading Association.

Hammer, C., Nimmo, D., Cohen, R., Draheim, H. & Johnson, A. (2005) Book reading interactions between African-American and Puerto Rican Head Start children and their mothers*. Journal of Early Childhood Literacy 5(3),* 195-227.

Hart, B. & Risley, T. (1995) *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore, MD; Brookes Publishing.

Hart, B. & Risley, T. (2003) The early catastrophe. The 30 million word gap. *American Educator, 27(1),* 4-9.

Heath, S. (1983) *Ways With Words: Language, Life, and Work in Communities and Classrooms.* Cambridge, UK: Cambridge University Press.

Helman. L. (2008) English words needed: Creating research based vocabulary instruction for English learners. In A.E.Farstrup & S.J.Samuels (Eds). *What Research Has to Say About Vocabulary Instruction*. Newark, DE, International Reading Association.

Hirsch, E. (2001, May 2) The latest dismal NAEP scores. *Education Week, 20.*

Hirsh, D. & Nation, P. (1992) What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8(2),* 689-696.

Hoff, E. (2003) The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development 74(5),* 1368-1378.

Hoff, E. & Tian, C. (2005) Socioeconomic status and cultural influences on language. *Journal of Communication Disorders 38(4),* 271-281.

Hulstijn, J. (2003) Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning, 16(5),* 413-425.

Hulstijn, J., Van Gelderen, A. & Schoonen, R. (2009) Automatization in 2nd language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*, 555-582.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology* 27(2),236–48.

Jacobsen, J., Olsen, C., Rice, J., Sweetland, S. & Ralph, J. (2001) *Educational Achievement and Black-White Inequality* (NCES 2001–061). US Department of Education .National Center for Education Statistics. Washington, DC: US.

Jusczyk, P. (1997) Finding and remembering words: Some beginnings by English- learning infants. *Current Directions in Psychological Science, 6(6)*, 170-174.

Kameenui, E., Carnine, D. & Freschi, R. (1982) Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly 17*, 367-388.

Kintsch, W. (1988) The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review 95(2*), 163-182.

Knight, S. (1994) Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, *78*, 285-299.

Kojic-Sabo, I. & Lightbown, P. (1999) Students' approaches to vocabulary learning and their relationship to success. *The Modern Language Journal, 83(2)*, 176-192.

Kuhl, P. & Meltzoff, A. (1996) Infants' vocalizations in response to speech: Vocal imitation and development differences. *Journal of the Acoustical Society of America 100(4),* 2425-2438.

Laberge D. & Samuels, S.J. (1974) Toward a theory of automatic information processing in word reading and reading comprehension. *Cognitive Psychology 6(2),* 293-323.

Landauer, T. & Dumais, S. (1997) A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review 104*, 211-240.

Laufer, B. (1992) How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (ed.s) *Vocabulary and Applied Linguistics*. London: Macmillan.

Lee, V. & Burkam, D. (2002) *Inequality at the starting gate: Social background differences in achievement as children begin school.* Washington D.C., Economic Policy Institute.

Lee, J., Grigg, W. & Donahue, P. (2007). *The Nation's report card: Reading 2007* (NCES 2007-496). Washington, DC: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.

Levy, M. (1997) Computer Assisted Language Learning. Oxford: Clarendon Press.

Li, J. (2010) Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning*, *23(3)*, 253-275.

Ma, Q. & Kelly, P. (2006) Computer assisted language learning: Design and evaluation. *Computer Assisted Language Learning* 19(1), 15-45.

Mayer, R.E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312–320.

McBride–Chang, C., Wagner, R., Muse, A., Chow, B. & Shu, H. (2005). The role of morphological awareness in children's vocabulary acquisition in English. *Applied Psycholinguistics,* 26 , 415-435.

McCutchen, D., Green, L., Abbott, R. & Sanders, E. (2009). Further evidence for teacher knowledge: supporting struggling readers in grades three through five. *Reading and Writing 22(4)*, 401-423.

McKeown, M., Beck, I., Omanson, R. and Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication'. *Journal of Literacy Research,* 15(1), 3-18.

McKeown, M. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly* 20(4), 482–496.

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and fequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly* 20(5), 522–535.

McKeown, M. & Beck, I. (2004). Direct and Rich Vocabulary Instruction. In J. Bauman & E. Kame'enui (Eds.) *Vocabulary Instruction: Research to Practice* (pp. 13-27). New York, USA: The Guilford Press.

Meijer, R. & Nering, M. (1999) Computer-adaptive testing: Overview and introduction. *Applied Psychological Measurement* 23(3), 187-194.

Metsala J. (1999) Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91, 3– 19.

Nagy, W., Herman, P. & Anderson, R. (1985) Learning words from context. *Reading Research Quarterly 20*, 233-253.

Nagy, W. (1988) *Teaching Vocabulary to Improve Reading Comprehension*. Urbana, IL: National Council of Teachers of English.

Nagy, W. (2005) Why vocabulary instruction needs to be long-term and comprehensive. In EH Hiebert & ML Kamil (Eds), *Teaching and Learning Vocabulary Bringing Research to Practice (p27-44)*. Mahwah, NJ Erlbaum.

National Center for Education Statistics (2005) Reading Assessment. http: nces.ed.gov/nationsreportcard/itemmaps/?year=2005&grade=4&subject=reading.

National Reading Panel (2000). *Report of the National Reading Panel: Teaching children to read.* Bethesda, MD: National Institute of Child Health and Human Development.

Pavio, A. (1991) Dual Coding Theory: retrospect and current status. *Canadian Journal of Psychology 45*, 255-287.

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

Reitsma, P. (1983) Printed word learning in beginning readers. *Journal of Experimental Child Psychology 36,* 321-338.

Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental Psychology, 28,* 406–413.

Richman, A., Miller, P. & leVine, R. (1992) Cultural and educational variations in maternal responsiveness. *Developmental Psychology 28,* 614-621.

Roser, N. & Juel, C. (1981) Effects of vocabulary instruction on reading comprehension. Paper presented at the *Annual Meeting of the National Reading Conference*, Dallas, TX Dec 2-5, 1981.

Rowe, M. (2008) Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language 35,* 185-205.

Rumelhart, D. (1980) Schemata: The building blocks of cognition. In RJ Spiro, BC Bruce & WE Brewers (Ed.s) *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ, Lawrence Erlbaum Associates.

Sadoski, M. (2005) A dual coding view of vocabulary learning. *Reading & Writing Quarterly 21(3),* 221-238.

Saffran, J., Aslin, R. & Newport, E. (1996) Statistical learning by 8-month-old infants. *Science 274,* 1926-1928.

Scarr, S., & Weinberg, R. A. (1978). The influence of “family background” on intellectual attainment. *American Sociological Review, 43,* 674–692.

Scott, J., Nagy, W. & Flinspach, S. (2008) More than merely words: Redifining vocabulary learning in a culturally and linguistically diverse society. In: A.E.Farstrup & S.J.Samuels (Eds). *What Research Has to Say About Vocabulary Instruction*. Newark, DE, International Reading Association.

Stahl, S. & Fairbanks, M. (1986) The effects of vocabulary instruction:A model-based meta-analysis. *Review of Educational Research 56(1),* 72-110.

Stahl, S. (1983) Differential word knowledge and reading comprehension. *Journal of Reading Behavior 15(4),* 33-50.

Stahl, S. & Nagy, W. (2006) Teaching Word Meanings. Mahwah, NJ. Lawrence Erlbaum Associates.

Stanovich, K. (1986) Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly 21*, 360-407.

Sun, Y. & Dong, Q. (2004) An experiment on supporting children’s English vocabulary learning in multimedia contexts. *Computer Assisted Language Learning, 17(2),* 131-147.

Tozcu, A. & Coady, J. (2004) Successful learning of frequent vocabulary through CALL also benefits reading comprehension and speed. *Computer Assisted Language Learning* 17(5), 473-495.

Tsao, F., Liu, H., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, *75*, 1067–1084.

Vaughn, S., Martinez, L., Linan-Thompson, S., Reutebuch, C., Carlson, C. & Francis, D. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English Language Learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness, 2(4)*, 297-324.

Verhoeven, L. (1990) Acquisition of reading in a second language. *Reading Research Quarterly 25(2)*, 90-114.

Walker, D., Greenwood, C., Hart, B. & Carta, J. (1994) Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, *65(2)*, 606-621.

Weizman, Z. & Snow, C. (2001) Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology 37(2)*, 265-279.

White, T., Graves, M., & Slater, W. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology,* 82, 281–290.

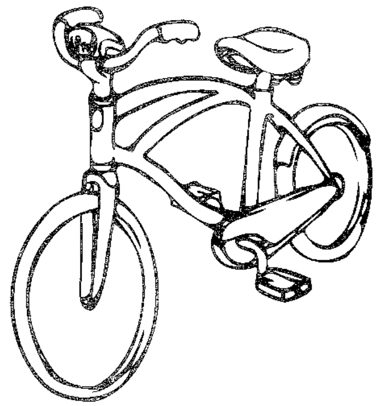
Whitehurst, G., Arnold, D., Epstein, J., Angell, A., Smith, M. & Fischel, J. (1994) A Picture Book Reading Intervention in Day Care and Home for Children From Low-Income Families. *Developmental Psychology* 30(5), 677-689.

Wood, J. (2001) Can software support children's vocabulary development? *Language, Learning and Technology* 5(1), 166-201.

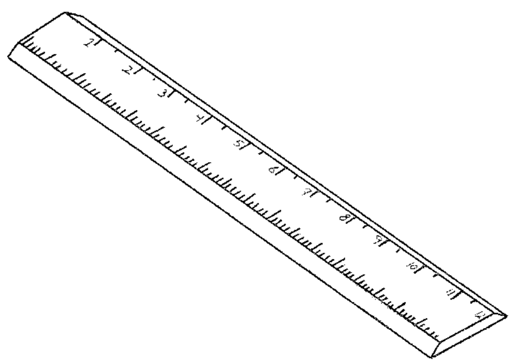
Yoshii, M. (2006) L-1 and L-2 glosses: Their effects on incidental vocabulary learning. *Language, Learning and Technology, 10(3)*, 85-101.

Zeno, S., Ivens, S., Millard, R., Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: Touchstone Applied Science Associates.

Appendix A. Seward Reading Vocabulary Test

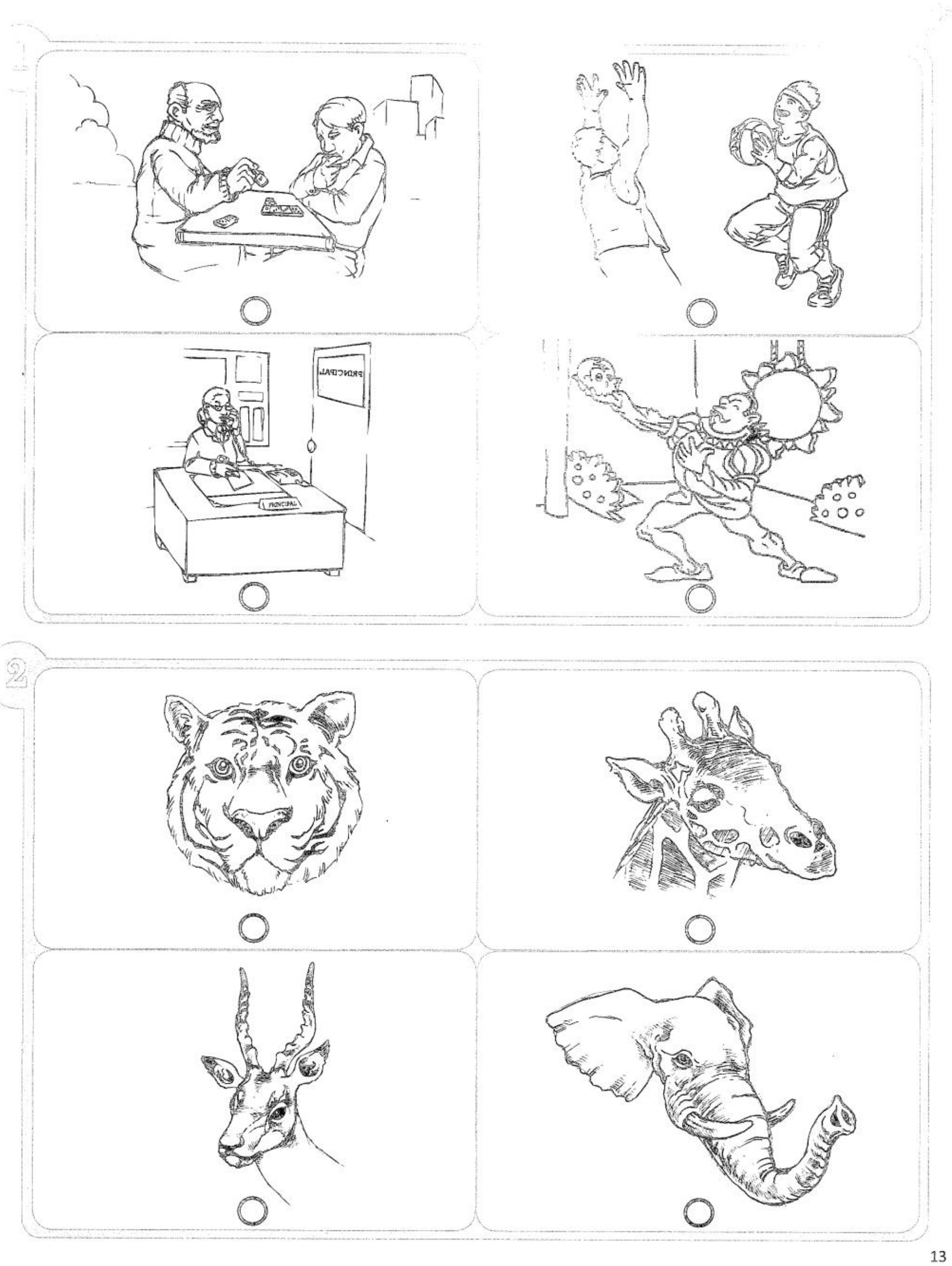


mirror ○
bicycle ○
car ○
highway ○



cooler ○
lawn ○
pencil ○
ruler ○

Appendix B. Seward Listening Vocabulary Test



Appendix C. Study 1 Teacher Survey

The First 4000 Words Teacher Survey

Dear Teachers,

Thank you very much for your participation in this research effort. Your opinions will help guide future research in this area. Please read the following statements and indicate your agreement or disagreement with them. In addition, your insights, critiques, or comments on your experience implementing this program would be greatly appreciated.

1 = Strongly agree   2 = Agree   3 = Neutral   4 = Disagree   5 = Strongly disagree

1. Implementation of the *4KW* program caused significant problems for my classroom management.   1   2   3   4   5

2. When using the *4KW* program, students were very much “on-task”.   1   2   3   4   5

3. I had a lot of difficulty integrating the *4KW* program into my normal curriculum.   1   2   3   4   5

:

4. The students assigned to the *4KW* program really enjoyed using it. 1 2 3 4 5

5. Students needed a lot of supervision when using the *4KW* program. 1 2 3 4 5

6. The *4KW* program would be a much more useful teaching tool in a computer lab or library than in the classroom. 1 2 3 4 5

7. The students using the *4KW* program were more "on-task" than they would typically be during whole-class activities. 1 2 3 4 5

8. The quality of the instruction the *4KW* program is high. 1 2 3 4 5

(over)

9. What grade level do you teach? 1 2 3 4

Please provide any additional thoughts, comments or critiques:

Appendix D. Study 1 Student Survey Questions

1) Was doing these lessons fun? 3 2 1

2) How well did you like the stories? 3 2 1

3)Do you think the software looks interesting? 3 2 1

4) Did this program help you learn new words? 3 2 1

5) Would you like to do more of this? 3 2 1

Appendix E. Study 2 Observation Form

4KW Observation Form

Date: School: Grade: Teacher:

Treatment Students: 1 = engaged 0 = not engaged * = teacher interaction /

A=assessment; R1, R2, R3= respective readings; S=subgame; P=puzzle; M=matching; B= bonusgame

A) 5’___/____ 10’___/____ 15’___/____ 20’___/____

B) 5’___/____ 10’___/____ 15’___/____ 20’___/____

Control Students: Circle all that apply

Independent/Group reading

Storytelling (listening)

Vocabulary tasks

Other literacy tasks (writing, rhyming, etc)

Other (math, games, etc) specify:

Teacher Notes (teacher comments, teacher strategy):