Computer Assisted Language Learning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ncal20

The effects of individualized, online vocabulary instruction on picture vocabulary scores: an efficacy study

Charles N. Fehr [a], Mark L. Davison [a], Michael F. Graves [a], Gregory C. Sales [b], Ben Seipel [a] & Sarah Sekhran-Sharma [b]

[a] Department of Educational Psychology, University of Minnesota, Minneapolis, USA

[b] Seward Inc., Minneapolis, USA

Available online: 28 Jun 2011

To cite this article: Charles N. Fehr, Mark L. Davison, Michael F. Graves, Gregory C. Sales, Ben Seipel & Sarah Sekhran-Sharma (2011): The effects of individualized, online vocabulary instruction on picture vocabulary scores: an efficacy study, Computer Assisted Language Learning, DOI:10.1080/09588221.2011.586640

To link to this article: http://dx.doi.org/10.1080/09588221.2011.586640

Computer Assisted Language Learning
2011, 1–16, i*First* article

Routledge
Taylor & Francis Group

The effects of individualized, online vocabulary instruction on picture vocabulary scores: an efficacy study

Charles N. Fehr[a]*, Mark L. Davison[a], Michael F. Graves[a], Gregory C. Sales[b], Ben Seipel[a] and Sarah Sekhran-Sharma[b]

[a]*Department of Educational Psychology, University of Minnesota, Minneapolis, USA;* [b]*Seward Inc., Minneapolis, USA*

Vocabulary knowledge is of fundamental importance to reading comprehension, and many students lack the vocabulary knowledge necessary to facilitate learning to read. A study was conducted to determine the effects of an individualized, online vocabulary program on picture vocabulary test scores. Elementary summer school students ($N = 43$), entering grades 2–4, who scored poorly on a vocabulary pretest were randomly assigned to treatment or control conditions. Students in the treatment condition received computer-delivered vocabulary instruction on a stratified sample of 100 words selected from 4000 of the most common words in written English. Posttest scores on a picture vocabulary test showed that students in the treatment condition outperformed control students by more than one standard deviation. The computer-adaptive, individualized instruction provided by this vocabulary program addresses a need for efficiency in remediation of vocabulary deficits. Further study is planned to determine whether improved vocabulary performance mediated by this computer assisted language learning (CALL) program might transfer to broader measures of vocabulary knowledge or reading comprehension.

Keywords: vocabulary; reading comprehension; online instruction; randomized field trial

Introduction

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. 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 US students are increasingly outperformed by students in other developed countries on reading measures, the research and development (RAND) Reading Study Group (2002) also argued that vocabulary instruction should be a focus from pre-school through the elementary years.

*Corresponding author. Email: fehrx017@umn.edu

ISSN 0958-8221 print/ISSN 1744-3210 online

DOI: 10.1080/09588221.2011.586640
http://www.informaworld.com

Moreover, the RAND group identifies vocabulary knowledge as a "critically important determinant" (p. 43) of reading comprehension. In the United States, 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 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).

This report summarizes the results of a study that implemented a precursor of a computer assisted language learning (CALL) vocabulary instruction program, *The First 4000 Words* (*4KW*). Created by Greg Sales and Michael Graves, this software program delivers vocabulary instruction on 4000 words most commonly found in written English (Zeno, Ivens, Millard, & Duvvuri, 1995) and does so using a research-based instructional design. The program utilizes computer-adaptive technologies to allow for individualized, remedial vocabulary instruction for students in grades 1–4.

Groups at risk for vocabulary deficits

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 both striking and cause for serious concern. In their landmark study, Hart and Risley (1995, 2003) conducted longitudinal research with 42 families categorized on SES as professional, working-class, or welfare. By monthly monitoring of parent–child interactions for over two and a half years, beginning with children at age seven months, they 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 and Risley estimated a 6000 word gap by the age of six. Subsequently, Farkas 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 10-year period, tracking children from age three to age 13. Their results confirmed the finding of large vocabulary gaps between high and low SES children on measures of oral vocabulary. These findings are consistent with the views of Becker (1977) who suggested that insufficient vocabulary knowledge is the primary barrier to achievement for lower SES students. Becker argued that systematic vocabulary instruction is required to address the achievement gap.

Another important and rapidly growing group in the United States that can be identified as at-risk in English vocabulary development is comprised of those students learning English as a second language (L2). Approximately one in nine K-12 students in the United States is L2, and this number is predicted to grow to nearly one in three by 2015 (Francis & Vaughn, 2009). Children of families wherein English is not the primary language spoken in the home may lag behind their native English-speaking (L1) peers in English vocabulary development. This puts these students at high risk for failure in educational settings. For example, L2 students' lack of adequate vocabulary may contribute to below grade-level reading comprehension achievement and lower standardized test scores, thus making an erroneous diagnosis of reading disability or placement in special education more likely (August, Carlo, Dressler, & Snow, 2005). While nearly a third of native English 8th grade students are proficient-level on national achievement reading tests, only 4% of L2 students meet this threshold (Lee, Grigg, & Donahue, 2007). Lack of academic vocabulary knowledge, which impedes comprehension of content area text, is a primary obstacle for L2 students' academic success (Vaughn et al., 2009). Compounding the problem for many L2 students is that many are also lower SES students, and therefore face challenges in that regard as well. L2 learners encompass a tremendous variety of students. A majority of L2 students in the United States come from homes or backgrounds with Hispanic base languages, and significant variability exists in children's L1 literacy and vocabulary knowledge within that population. These differences in native language proficiency are likely to affect English vocabulary and other literacy-related achievement (Knight, 1994; Li, 2010). L2 students in the United States 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.

Students' vocabulary deficits, which likely stem from limited exposure to spoken English due to a disadvantaged background or a non-English-speaking home, place these students at high risk for failure in educational settings and beyond. The gap these students face relative to their peers develops over a period of years during early childhood (Hirsch, 2001). It is, therefore, likely that remediation of vocabulary gaps will require sustained and intensive efforts. Our current and future research targets schools in low SES areas with high proportions of L2 students.

Can vocabulary instruction aid reading comprehension?

Reading comprehension requires overlapping and concurrent engagement in a number of component sub-skills (e.g. both low-level processes, such as decoding, and high-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. L2 learners face an additional challenge in this regard; new vocabulary words are likely to be translated into L1 before or during semantic retrieval. This is an additional cognitive step that must be performed efficiently if the meaning of a word is to be accessed from print in an automatic fashion (Hulstijn, Van Gelderen, & Schoonen, 2009). L2 learners are, therefore, especially likely to benefit from instructional tasks that require and reinforce connecting new vocabulary words with their meanings in a rapid manner, increasing the speed of lexical access toward automaticity.

Despite the logical connection between vocabulary knowledge and reading comprehension, empirical research demonstrating a causal link between vocabulary instruction and improved reading comprehension is limited. This may be due to the large number of words that must be learned before improvement in comprehension can be detected. Most studies addressing the effects of vocabulary instruction on reading comprehension have taught a relatively small number of words. Changes in vocabulary knowledge that might occur during typical experimental time frames are unlikely to be detected by standardized tests of vocabulary (National Reading Panel, 2000). However, in a meta-analysis of studies relating vocabulary instruction to reading comprehension, Stahl and Fairbanks (1986) argue that the data support a causal role for vocabulary instruction in improving reading comprehension. They found that vocabulary instruction has a strong effect on comprehension of reading passages that include taught words, and a smaller effect on comprehension of passages that do not include explicitly taught words. A more recent meta-analysis using more sophisticated analyses found smaller effects on reading comprehension in general, but suggested that struggling readers were the most likely to benefit from vocabulary instruction (Elleman, Lindo, Morphy, & Compton, 2009).

Other research findings also suggest a causal relationship between vocabulary and reading comprehension. In a series of randomized trials with 4th, 5th, and 6th grade students, Kameenui, Carnine and Freschi (1982) found that replacing easier vocabulary terms with more difficult ones decreased passage comprehension significantly, as measured by both inferential questioning and passage recall. That is, comprehension is facilitated when word meanings are better known by the reader. This study also showed that vocabulary instruction on the difficult words in a text led to better comprehension scores. These results are consistent with a view positing the importance of vocabulary knowledge for reading comprehension. In a similar study conducted with 1st–5th grade students, Roser and Juel (1981) found that students receiving direct vocabulary instruction on words that appeared in subsequent story reading performed significantly better on a follow-up comprehension assessment. The largest improvements came from those students in the lowest reading ability groups.

Beck, Perfetti and McKeown (1982) provided 4th grade students with long-term vocabulary instruction and compared them to a matched control group, using a variety of vocabulary and reading-related measures. The instructed group received extensive vocabulary instruction on 105 target words over a five-month period. After controlling for prior achievement on standardized tests of vocabulary and reading comprehension, the instructed group not only out-performed the control group on custom vocabulary measures, lexical decision tasks, and sentence verification tasks

involving the target words but also scored significantly higher on standardized tests of vocabulary and reading comprehension. These results suggest that vocabulary instruction can result in faster and more accurate semantic processing of words and that the effects of instruction may transfer to both words not explicitly taught and to general reading comprehension (Beck et al., 1982).

In a randomized study with college-aged L2 learners, Tozcu and Coady (2004) found that subjects assigned to a treatment of three hours per week spent on computer-delivered instruction of 2000 high-frequency words performed significantly better on subsequent tests of both vocabulary and reading comprehension compared to subjects in a control condition who read texts and completed comprehension tasks.

These studies, taken together, indicate the strong relationship between vocabulary instruction and reading comprehension. However, methodological issues have precluded any single study from demonstrating a strong causal relationship. A long-term study, teaching larger numbers of vocabulary items within an adequately powered randomized research design will be necessary to clarify whether a true causal relationship exists.

A CALL approach to explicit vocabulary instruction

Vocabulary instruction has been a favorite target of computer-based instruction since computer use expanded in education in the 1980s (Ma & Kelly, 2006). Typical methods have been games and cloze-like designs developed to aid L2 or low-level learners, but quality and effectiveness of many of these packages have been low due to a lack of theoretical or pedagogical knowledge on the part of program designers (Ma & Kelly, 2006). Many studies of CALL approaches to vocabulary instruction in recent years have focused on the use of computerized glosses, online dictionaries, and multimedia software to facilitate incidental L2 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 L1 and L2 learners. A recent meta-analysis also demonstrated that computer glosses are highly effective in aiding both reading comprehension and vocabulary development for L2 learners (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 CALL 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 students learning English as L2 and found that vocabulary learning and sentence comprehension were enhanced by use of pictorial and auditory contextual cues in combination with decontextualized word training. This is consistent with established research showing that explicit word learning is optimized by a combination of contextualized and decontextualized target word instructions (Beck, McKeown, & Kucan, 2008; Graves, 2006).

A potentially promising avenue for CALL 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 instructions 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 when teaching vocabulary to L1 and L2 learners: how to individualize instruction for a group of students whose vocabulary needs are not equivalent.

The CALL program tested in this study is a prototype of Seward, Inc.'s *The First 4000 Words*, a software program designed to teach high-frequency words to students whose vocabulary knowledge is insufficient to support reading and understanding basic texts. The instructional components of this program are congruent with research and theoretical perspectives from a wide range of scholars in the area of vocabulary and language learning. That is, students and to the target vocabulary items repeatedly (Stahl & Fairbanks, 1986), and to the target items both in reading contexts (Biemiller, 2004) and using simple definitions (Beck et al., 1982), and have the opportunity to engage in read-along activities when using the program (Graves, 2006; Hulstijn, 2003). Students using this program work individually at computer stations wearing headsets. Instruction takes place in an animated environment, "the Vocabitat". Words are presented in 10-word lessons, beginning with a pretest of the 10 target words. During the lesson, students have the opportunity to pronounce words in isolation and receive feedback via voice-recognition technology. The opportunity to pronounce words with feedback is likely to be especially valuable to L2 learners, who may not have well-developed knowledge of L2 graphemic/phonemic rules (Sun & Dong, 2004). User-friendly definitions constructed with high-frequency words are provided for words missed on the pretest and are accompanied by pictorial cues representing word meanings. Students hear the target words used in context in a story three times. These simple stories have been constructed using high-frequency words, including target words from previous lessons, such that stories are likely to contain few unfamiliar words beyond the target vocabulary items. This construction is likely to aid in story comprehension (Hulstijn, 2003). During the story readings, the text of the story is visible, and each sentence is underlined as it is read to the student. Target words missed on the pretest are highlighted in the text, and students can click on any sentence to hear it repeated. The ability to listen and re-listen to sentences provides students, particularly L2 learners, the opportunity to parse L2 words from fluent speech, which aids in both explicit and implicit word-learning, and allows for growth in phonological, morphological, lexical, and syntactic knowledge (Hulstijn, 2003). The *4KW* program represents a theoretically driven CALL approach to explicit vocabulary instruction for young L1 and L2 learners who need knowledge of high-frequency words to support learning to read.

Vocabulary knowledge is a strong predictor of reading comprehension, and increasing students' vocabulary knowledge should aid in their reading comprehension. It is the case that many students have vocabulary deficits that are likely to hinder their reading comprehension. No existing study with young students demonstrates strong gains in general reading comprehension measures as a result of vocabulary instruction. It is likely that a large number of words would need to be taught before gains on general reading comprehension measures could be detected. It has been proposed that understanding of 3000–5000 of the highest frequency English word families is required to provide the coverage necessary to comprehend basic English texts (Hirsh & Nation, 1992; Laufer, 1992). Most studies have taught fewer than 100 words over relatively short (i.e. days or weeks) experimental time frames. The current study was conducted to test whether an online, individualized vocabulary instruction program could successfully teach a limited corpus of

vocabulary items (100) to elementary students with low vocabulary knowledge. This is a necessary step toward a larger goal of determining whether a more extensive and sophisticated intervention teaching thousands of words over a period of months or years might result in significant gains in reading comprehension.

Methods

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 females) or control conditions ($N = 21$; 13 females).

Materials

Pretests and posttests

The paper-and-pencil pretests and posttests were identical and consisted of 40 multiple-choice items. Each item consisted of a black-and-white drawing with four printed word choices. The 40 words tested on the pre- and posttests were selected from a stratified random sample of the 4000 words most common in written English from *The Educator's Word Frequency Guide* (Zeno et al., 1995). Testing with similar students has shown that the test is highly reliable (Cronbach's $\alpha = 0.7$–0.9) and scoring correlates highly with the Gates-MacGinitie Vocabulary test.

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 words

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 feedbacks 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 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 are 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

In order to assess program practicality and students' impressions of the computerized vocabulary instruction, a teacher feasibility survey and a student satisfaction survey were developed. The teacher ($N = 8$) 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 (see Appendix 2 for the survey). The student satisfaction survey was informal and administered verbally throughout the program administration. The survey addressed students about impressions and experiences using the computerized vocabulary instruction program. Students' responses were rated on the following scale: 1 = thumbs down, 2 = no opinion, 3 = thumbs up (see Appendices 1 and 2 for teacher and student survey questions).

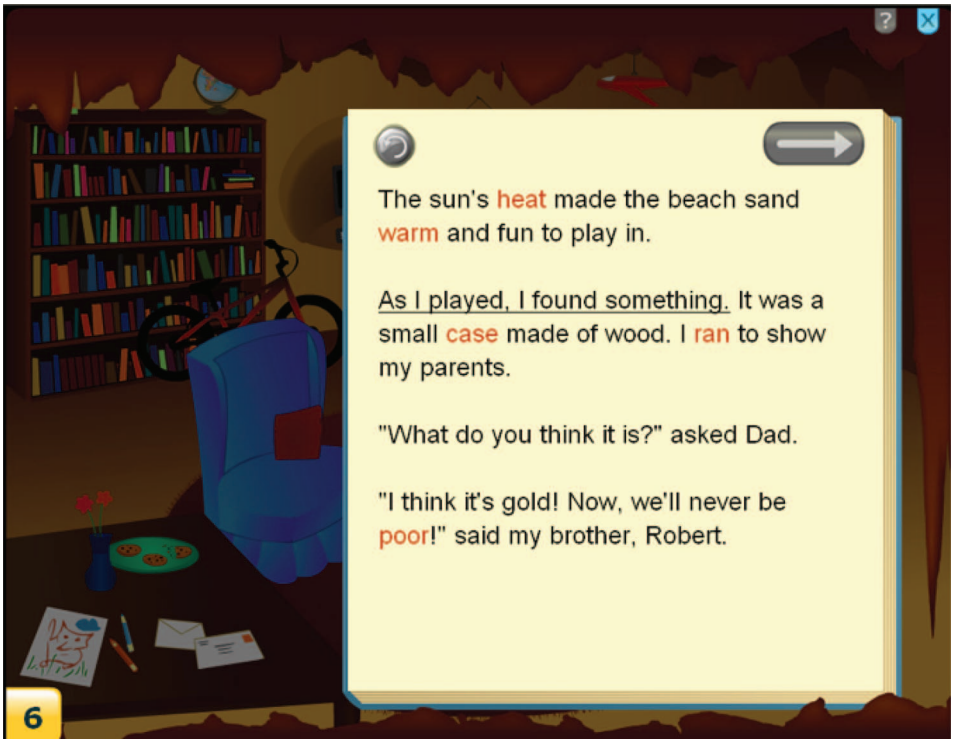

Figure 1. First reading of a story, with highlighted target words and underlining.

Procedure

All summer school students entering grades 2–4 completed the 40-word, paper-and-pencil 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. Students were able to complete one to three lessons during a typical session. Two or three experimenters monitored students' 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 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 paper-and-pencil posttest. After the posttest was administered, the

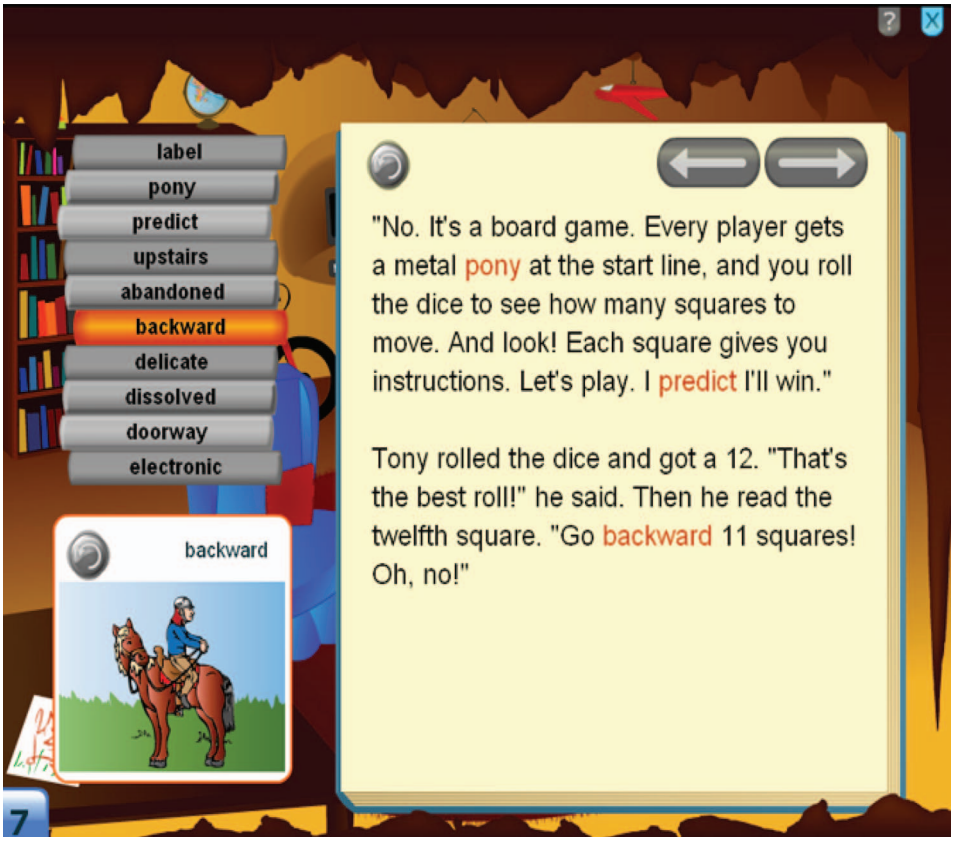

Figure 2. Second reading of a story: student has clicked the word "backward".

experimenters met with the teachers to administer the survey and participate in the brief group interview based on the survey questions.

Results

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 posttest raw scores and standard deviations (SDs) 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 pretest scores exists between those assigned to experimental and control conditions in any of the three grade levels, or overall. This suggests that randomization of subjects was successful in balancing prior vocabulary knowledge between the experimental and control conditions.

Table 2 reports mean gain scores (posttest 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, likely due to insufficient power. The effect sizes for gain scores averaged

Table 1. Mean pretest and posttest scores by group.

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

Note: SD, standard deviation.

Table 2. Mean gain scores and effect sizes by group.

	Control		Experimental		
Group	Gain score	*N*	Gain score	*N*	Effect size[a]
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

Notes: [a]Cohen's *d*; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

more than one SD overall. Teachers ($N = 8$) expressed generally favorable views and felt that the program might integrate well with existing curricula and classroom frameworks. Teachers' ratings of overall perception of the program yielded a mean = 2.9 (SD = 0.4) on a 0–3 scale. Ratings of how the program would fit within existing curricula resulted in a mean = 2.8 (SD = 0.4). 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 expressed that the software would integrate well into existing literacy centers, especially given young students' increasing comfort with autonomous use of computers.

Students' affective responses were also very favorable. 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).

Discussion

The purpose of this study was to test whether students' use of this CALL 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 this CALL 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 allow for instruction on a much larger corpus of words than is reasonably feasible otherwise.

The *4KW* program includes instructional components that research has shown to be essential for effective student learning of vocabulary in both L1 and L2. Among these are repeated exposure to words (Graves, 2006; Nagy, 1988; Stahl & Fairbanks, 1986), examples of target words used in context (Biemiller, 2004; Stahl & Fairbanks, 1986), provision of definitions in simple language (Stahl & Nagy, 2006), and a focus on words that students are likely to encounter frequently (McKeown & Beck, 2004; Nagy, 1988). Instruction is multimedia in nature, with pictorial cues accompanied by auditory definitions (Mayer & Moreno, 1998), and read-alongs providing opportunities to listen and re-listen to fluent use of words in context (Hulstijn, 2003). Although the *4KW* program implemented in this study 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 are being incorporated into future studies. These activities will increase the number of encounters with target words and, as with the program used here, will focus on words missed at the lesson pretest, capitalizing on the computer-adaptive design of the program.

It is encouraging that all three grade levels studied here showed vocabulary gain scores higher for subjects in the treatment condition compared to control groups. Although the increase was not statistically significant in 4th grade, this appears to be due to lack of statistical power stemming from small sample size. 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. The trends apparent in the results, however, suggest that this is due to the small size of the sample in this study. Determining whether a gender difference exists with relation to student learning using this CALL technology will require a larger sample size, and is a focus of planned research. In addition, although students involved in this study included L2 learners, the number of L2 learners was insufficient for statistical analysis. Our future research plans target schools with high proportions of L2 learners. In general, the large improvements in mean posttest 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 this study. First, 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 US 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. Students received the program instruction outside of their classrooms, and teachers were only minimally involved. It is likely that the effectiveness of this CALL instruction would be enhanced when teachers actively motivate and monitor students' progress and integrate this remedial vocabulary learning into other literacy-related activities. It is also the case that the students who might benefit most from the *4KW* program are a subset of those included in this study. It is unlikely that one program would be appropriate for all students with low vocabulary knowledge in grades 1–4. Future

research using significantly larger sample sizes will allow for identification of specific students' profiles most likely to benefit from this type of instruction. The results do not identify the specific program components that lead to positive outcomes. For example, if a student successfully learns a word (as defined by missing the item on paper-and-pencil pretest but correctly responding on posttest), is it the exposure to the word's definition or hearing it in the story that mattered? If the student requested sentences containing the word to be repeated, is that important? The literature suggests that the combination of these instructional components leads to effective vocabulary learning. While the software used in this study did not record all students' online activities, the *4KW* software program does have this capability, and future research will allow for factor analysis of specific student's actions that lead to word learning.

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 under current investigation. We would expect significantly smaller effect sizes when vocabulary assessments include a smaller percentage of words directly targeted by instruction. It is important to note that learning the meaning of a word is not a one-step process. Even highly literate adults are likely to be continually adding nuances to their understanding of many words. Complicating the task for young students is the fact that the most common English words are almost always polysemous, and new meanings are occasionally assigned to common words (i.e. the word "text" is now a common verb). The *4KW* program is designed as a remedial introduction to words that students will commonly encounter in typical English texts, and provides a basic meaning of a word representing the word's most common usage. It is likely that having such a basic word understanding would facilitate more sophisticated understanding of word meaning as students subsequently encounter words in reading, listening, and conversation. Whether implementation of the *4KW* vocabulary program in its entirety, which includes more interactive, individualized instruction and tasks designed to promote automaticity in lexical access, leads 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 this study did not attempt to measure students' reading comprehension before or after the intervention.

Conclusion

It is unrealistic to suggest that computer-delivered vocabulary instruction can be the sole vehicle for remediation of significant vocabulary deficits or L2 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. However, the efficiencies available through computer-adaptive technology, combined with a theoretically driven instructional software program, as described in this study, 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 L2 vocabulary learning. When embedded within a learning environment emphasizing a robust and metacognitive approach to language learning, computer-adaptive CALL vocabulary instruction may play an important role in helping at-risk students' literacy achievement (Kojic-Sabo & Lightbown, 1999). The vocabulary instruction program utilized in this study

represented only a part of the fully developed *First 4000 Words* program, but still led to significant effects on vocabulary knowledge. Our ongoing research, including a larger number of subjects, a more complete software program to allow for greater individualization of instruction, longer experimental time frame, and use of standardized vocabulary and reading comprehension measures will test possibilities suggested by this study.

Acknowledgement

This work was supported by Grant R305C050059 from the Institute of Education Sciences in the US Department of Education.

Notes on contributors

Charles N. Fehr is a doctoral student in the Psychological Foundations Track of the Department of Educational Psychology at the University of Minnesota. He is also a fellow in the Minnesota Interdisciplinary Training in Education Research (MITER) Program.

Mark L. Davison is the American Guidance Service, Inc. and John P. Yackel Professor of Educational Assessment and Measurement in the Department of Educational Psychology at the University of Minnesota. He is also the co-director of the Minnesota Interdisciplinary Training in Education (MITER) Program and the editor-in-chief of the journal *Applied Psychological Measurement*.

Michael F. Graves is professor emeritus of Literacy Education at the University of Minnesota and a member of the Reading Hall of Fame. His research and writing focus on vocabulary learning and instruction and comprehension instruction. His most recent books include *Teaching vocabulary to English-language learners* (in preparation, with Diane August and Jeannette Mancilla-Martinez), and *Reading better, reading smarter* (in press, with Deborah Appleman).

Gregory C. Sales is a former associate professor of instructional design and technology at the University of Minnesota and CEO of Seward Incorporated, a company he founded in 1990.

Ben Seipel is a doctoral student in the Psychological Foundations Track of the Department of Educational Psychology at the University of Minnesota. He is also a fellow in the Minnesota Interdisciplinary Training in Education Research (MITER) Program.

Sarah Sekhran-Sharma is a communications specialist with over 10 years of technical production experience providing assistance with problem-solving activities associated with product design, development and layout, graphic, video and multimedia development.

References

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

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.

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

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

Becker, W. (1977). Teaching reading and language to the disadvantaged: What we have learned from field research. *Harvard Educational Review*, *47*(4), 518–543.

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.

Chun, D., & Plass, J. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, *80*, 183–198.
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*, 464–497.
Francis, D., & Vaughn, S. (2009). Effective practices for English language learners in the middle grades: Introduction to the special issue. *Journal of Research on Educational Effectiveness*, *2*, 289–296.
Graves, M. (2006). *The vocabulary book: Learning and instruction*. New York: Teachers College Press.
Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.
Hart, B., & Risley, T. (2003). The early catastrophe. The 30 million word gap. *American Educator*, *27*(1), 4–9.
Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*, 689–696.
Hirsch, E. (2001, May 2). The latest dismal NAEP scores. *Education Week*, *20*, 41, 60.
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*, 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.
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.
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*, 176–192.
Laberge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in word reading and reading comprehension. *Cognitive Psychology*, *6*, 293–323.
Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Macmillan.
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, US Department of Education.
Lee, V., & Burkam, D. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.
Li, J. (2010). Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning*, *23*, 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.
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*, 401–423.
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*, 187–194.

Nagy, W. (1988). *Teaching vocabulary to improve reading comprehension*. Urbana, IL: National Council of Teachers of English.
National Reading Panel (2000). *Report of the National Reading Panel: Teaching children to read*. Bethesda, MD: National Institute of Child Health and Human Development.
RAND Reading Study Group (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Education.
Roser, N., & Juel, C. (1981, December). *Effects of vocabulary instruction on reading comprehension*. Paper presented at the annual meeting of the National Reading Conference, Dallas, TX.
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., & Nagy, W. (2006). *Teaching word meanings*. Mahwah, NJ: Lawrence Erlbaum Associates.
Sun, Y., & Dong, Q. (2004). An experiment on supporting children's English vocabulary learning in multimedia contexts. *Computer Assisted Language Learning*, *17*, 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*, 473–495.
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*, 297–324.
Yoshii, M. (2006). L-1 and L-2 glosses: Their effects on incidental vocabulary learning. *Language, Learning and Technology*, *10*, 85–101.
Zeno, S., Ivens, S., Millard, R., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

Appendix 1. Teacher interview questions

1. How well does the vocabulary instruction align with your normal school year curriculum?

Poor	Fair	Good	Excellent	N/A

2. Based on your interaction with the software, how would you rate the program?

Poor	Fair	Good	Excellent	N/A

Appendix 2. 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