

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Vocabulary Maintenance of Seventh Grade Students with Learning Disabilities

Permalink

<https://escholarship.org/uc/item/346293bz>

Author

Orozco, Guadalupe

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Vocabulary Maintenance of Seventh Grade Students with Learning Disabilities

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Guadalupe Orozco

June 2019

Dissertation Committee:

Dr. Rollanda O'Connor, Chairperson

Dr. Austin Johnson

Dr. Asha Jitendra

Copyright by
Guadalupe Orozco
2019

The Dissertation of Guadalupe Orozco is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Vocabulary Maintenance of Seventh Grade Students with Learning Disabilities

by

Guadalupe Orozco

Doctor of Philosophy, Graduate Program Education

University of California, Riverside, June 2019

Dr. Rollanda O'Connor, Chairperson

Students with Learning Disabilities (LDs) have less extensive vocabularies than their typically developing peers (Simmons & Kameenui, 1990). This is a critical problem, as the importance of vocabulary knowledge in relation to reading comprehension is widely documented in the literature (Pearson & Gallagher, 1983; Stahl, 1983) and has been shown to enhance students' comprehension of content area texts (Bos & Anders, 1990; O'Connor, Beach, Sanchez, Bocian, & Flynn, 2015). The reading demands placed on students' increase as they enter middle school; however, vocabulary acquisition alone is not sufficient for garnering long-term reading comprehension outcomes. The literature demonstrates a need for students with disabilities to retain vocabulary knowledge in order to make gains on reading comprehension outcomes and accommodate the increased reading load experienced when transitioning to secondary school. However, vocabulary retention research is lacking for students with disabilities in the secondary grades. The

vocabulary retention research that exists investigates vocabulary maintenance of elementary aged students. Moreover, the literature has failed to demonstrate the differences in vocabulary retention over time for English Learners (EL) with LDs. The purpose of this paper is to describe middle school English only and EL students' vocabulary maintenance of taught vocabulary words across five-time points for students with LDs participating in a research-based vocabulary intervention. Results indicate that students in the treatment group were able to maintain vocabulary knowledge over time. Results will be discussed further in terms of implications, limitations, and future research.

Table of Contents

	Page
Introduction.....	1
Method.....	28
Results.....	36
Discussion.....	43
References.....	50
Tables 1-6.....	55
Figures 1-2.....	61

LIST OF TABLES

	Page
1. Demographic Information and Baseline Measure Scores.....	55
2. Treatment vs. BAU Control Independent Samples t-test.....	56
3. Descriptive Statistics for Treatment and BAU control Group.....	57
4. Model Fit Indicators.....	58
5. Estimates of Parameters for Two Latent Difference Score Models	59
6. BAU Control Pretest vs. Spring Y2 Maintenance Paired-Samples t-test	60

LIST OF FIGURES

Page

1. Figure 1: Latent Difference Score Model for Vocabulary at Five Times of Measurement..... 61
2. Figure 2: CREVT-3 and Maintenance Spring-Y2 Correlation for Treatment and Control Groups..... 62

Vocabulary Maintenance of Seventh Grade Students with Learning Disabilities

Vocabulary knowledge is one of the best predictors of educational achievement (Kurdek & Sinclair, 2001). Students with disabilities consistently demonstrate lower levels of academic achievement than their non-disabled peers. Bulgren, Kansas, and Deshler (2013) discussed the difficulties students with Learning Disabilities (LDs) face developing higher order reasoning skills, thereby hindering their ability to attain adequate academic achievement across content areas. Wei, Blackorby, and Schiller (2011) investigated the reading growth trajectories in a nationally representative sample of students with disabilities and found that reading growth trajectories differed across all 11 federal disability categories. Specifically, students with speech or visual impairments performed highest on reading achievement measures while students with multiple disabilities or intellectual disabilities performed lowest. In addition, findings indicated that students with LDs improved at a faster rate than students with speech or hearing impairments or autism. Moreover, Lackaye and Margalit (2006) compared the academic achievement of students with and without LDs and found that students with LDs achieved lower grades across content areas as compared to students without LDs.

Vocabulary acquisition, in particular, has long been considered an essential educational component with which low-performing students struggle. For instance, Hwang, Lawrence, Mo, and Snow (2015) investigated the effects of a vocabulary intervention on 6th, 7th, and 8th grade students with varying levels of English language proficiency (ELP). Findings indicated that language minority students at varying levels of ELP differ in their reading profiles from Native English Speakers (NES). Moreover, a

substantial degree of variation in reading performance of students with varying levels of ELP was demonstrated. Findings suggest that students' level of ELP impacts the extent to which students benefit from research-based vocabulary instruction. In addition, research has clearly demonstrated that students with LDs have less extensive vocabularies than their typically developing peers (Simmons & Kameenui, 1990). Vocabulary acquisition is facilitated with the use of independent word learning strategies and wide reading; however, students with LDs rarely engage in these activities sufficiently to expand their vocabulary knowledge base (Baker, Simmons, & Kameenui, 1995). This is a critical problem, as the importance of vocabulary knowledge in relation to reading comprehension is widely documented in the literature (Pearson & Gallagher, 1983; Stahl, 1983) and has been shown to enhance students' comprehension of content area texts (Bos & Anders, 1990; O'Connor, Beach, Sanchez, Bocian, & Flynn, 2015). In addition, Carver (1994) demonstrated that competent readers are able to withstand a small proportion of unknown words while reading without disruption to comprehension; however, if the proportion of unknown words is too high, reading comprehension falters. The reading demands placed on students' increase as they enter middle school, thus effective vocabulary instruction which ensures a thorough understanding of taught vocabulary words and increases reading comprehension is necessary for students with and without disabilities in the secondary grades.

Fortunately, the literature base on effective vocabulary instruction is extensive. For instance, the literature emphasizes the need for vocabulary instruction to be both direct (e.g., teaching specific words, suffixes, and prefixes) and indirect (e.g., exposing

students to new words and having students engage in wide and varied reading; Sedita, 2005). Phillips, Foote, and Harper (2008) identified the following effective evidence-based strategies for improving vocabulary instruction across content areas: thoughtful selection of vocabulary words, incorporation of graphic organizers, using logic and prediction of word meaning, incorporating synonyms and antonyms to vocabulary instruction, and incorporating word classification in vocabulary instruction. In addition, Beck and McKeown (2007) suggested that effective vocabulary instruction requires extensive and rich vocabulary instruction that provides various opportunities for students to be exposed to words, and for students to use these words in conversation and in their writing. In their meta-analysis investigating the effects of vocabulary instruction for students with LDs, Jitendra, Edwards, Sacks, and Jacobson (2004) found that vocabulary instruction for students with LDs that incorporated direct instruction and other evidence-based strategies, was generally effective and led to increased vocabulary acquisition gains. The literature demonstrates a thorough understanding of effective vocabulary instruction methods for both students with and without disabilities; however, long term research in vocabulary instruction for students with LDs is sparse. In addition, vocabulary acquisition alone is not sufficient for garnering long-term reading comprehension outcomes. Rather, students must retain the vocabulary knowledge acquired through intervention and instruction over time in order to make gains in long-term reading comprehension outcomes.

Long Term Memory

One potential mechanism for a lack of vocabulary maintenance is difficulties with Long-term memory (LTM). LTM refers to the stage of the 1968 model of memory proposed by Richard Atkinson and Richard Shiffrin where information is stored for an extended period of time. The Atkinson and Shiffrin (1968) model of memory proposed that memory is processed in a linear manner such that information from short term memory (STM) is transferred to LTM only if that information is rehearsed. Theoretically, the capacity of LTM is unlimited, the main constraint being accessibility of memories stored in LTM rather than the availability of LTM. Students with LDs compared with non-disabled peers have been found not to apply rehearsal strategies in an effort to maintain good performance on reading tasks (Torgesen, 1980). The lack of rehearsal strategies used may limit students with LD's ability to transfer information from STM to LTM, thus limiting accessibility to long-term semantic memory. In terms of vocabulary acquisition, prior vocabulary knowledge is important given that relevant prior vocabulary knowledge can facilitate the learning of new words (Perfetti & Hart, 2002). Perfetti and Hart (2002) indicated that vocabulary acquisition in the long term is affected by the effortlessness with which children learn words and retrieve words from LTM.

Vocabulary maintenance requires the retrieval of vocabulary knowledge from LTM. Given the tendency for students with LDs not to use rehearsal strategies and the predisposition for students with LDs to have limited vocabulary knowledge, the vocabulary maintenance of students with LDs may differ as compared to non-disabled

peers. Next, I describe vocabulary maintenance in studies with typical learners, and studies which include students at-risk for reading failure and/or students with LDs.

Vocabulary Maintenance

Vocabulary maintenance research is an extension of vocabulary intervention research in that vocabulary maintenance is studied using delayed posttests to measure retention of learned words. Several research groups have investigated vocabulary maintenance as an outcome in vocabulary intervention studies. Biemiller and Boote (2006) conducted a vocabulary intervention study in which general education kindergarten, first grade, and second grade students participated in whole-class vocabulary lessons provided by the classroom teacher. The vocabulary intervention consisted of reading books to students and providing explanations for the target words. Students were administered a general vocabulary assessment to create matched cohorts of children and a vocabulary assessment specific to the target vocabulary words, which was used to assess student vocabulary word knowledge at posttests. Students were tested 2 weeks after the intervention period and again 6 weeks after the intervention period on target vocabulary word knowledge. Results indicated that students in kindergarten through second grade demonstrated a 6% gain from the 2-weeks-immediate post-test to the 6-weeks-delayed post-test on measures of target words vocabulary knowledge. Although instruction did not continue during the 4-week period between the immediate post-test and the delayed post-test students continued to gain vocabulary knowledge. The authors' attribute this continued gain in vocabulary knowledge to a possible increase in word consciousness when encountering the target words used in this study in a different

context. Nash and Snowling (2006) also focused on evaluating the effectiveness of providing young children ages 7 to 8 years with vocabulary instruction. Twenty-four children participated in the study, half of which participated in a definition-based vocabulary intervention and half participated in a context vocabulary intervention that taught students a strategy for deriving meanings from written context. At posttest, students in both groups improved equivalently on measures of vocabulary knowledge. However, at the 3-month delayed posttest, students in the context vocabulary intervention demonstrated significantly greater vocabulary knowledge of target words.

Coyne, McCoach, and Kapp (2007) conducted two studies aimed at identifying vocabulary knowledge gains for Kindergarten students. Study one consisted of a vocabulary intervention with 31 Kindergarten students in which an extended instruction condition was compared to an incidental exposure condition. In the extended instruction condition students were provided with simple definitions within the context of the story for three of the six target words. In the incidental exposure condition no definitions for target words were provided. At posttest, administered one to five days after conclusion of the intervention, the researchers found that students learned words taught using extended instruction to a greater extent than words taught through incidental exposure across all measures of vocabulary knowledge (i.e., expressive, receptive, and context measures). Delayed posttest data were collected eight weeks after posttest administrations, and findings indicated that expressive vocabulary knowledge decreased from posttest to delayed posttest; however, on receptive vocabulary knowledge measures and context measures no statistically significant differences between posttest and delayed posttest

measures were present. Coyne et al. (2007) also conducted a vocabulary intervention in which they compared extended instruction to embedded instruction of target words during storybook readings with 34 kindergarten students. The extended instruction was identical to the extended instruction condition in study one. In the embedded condition, simple definitions within the context of the story for all six target words were provided. Posttest data were collected shortly after the intervention ended while delayed posttest data were collected six weeks after posttest data collection. Results indicated that students learned words taught using extended instruction to a greater extent than words taught through embedded instruction across all three measures of vocabulary knowledge. In addition, across all three measures of vocabulary knowledge no statistically significant differences between posttest and delayed posttest measures were present.

In an extension of that study, Coyne, McCoach, Loftus, Zipoli, and Kapp (2009) conducted a vocabulary intervention with 42 kindergarten students with three instructional conditions (i.e., 2 intervention conditions and 1 control condition). The control condition consisted of teaching students' vocabulary words in a storybook reading session through incidental exposure. One intervention condition consisted of teaching students' vocabulary words in a storybook reading session through embedded instruction, which consisted of providing students with simple definitions for target vocabulary words. The second intervention condition consisted of teaching students' vocabulary words in a storybook reading session through extended instruction, which consisted of introducing the target words during the storybook reading. In addition, students participated in activities following the storybook reading session allowing them

to experience the target words in different contexts. Intervention sessions focused on nine target words with three target words taught in each condition. Posttest data were collected one to five days after the end of the intervention and delayed posttest data were collected eight weeks after posttest data were collected. Results on the posttest and the delayed posttest measures indicated that for expressive and receptive definition measures and measures of context knowledge the mean score of words learned through extended instruction was higher than words learned through embedded instruction. The mean score of words learned through incidental exposure was lower than both extended and embedded instruction at both time points.

Loftus, Coyne, McCoach, Zipoli, and Pullen (2010) investigated the effectiveness of a kindergarten vocabulary intervention designed to supplement classroom vocabulary instruction immediately after the vocabulary intervention and 7 weeks later to assess maintenance of word knowledge. The vocabulary intervention incorporated the following evidence-based components of effective supplemental interventions: (a) small groups of three to four students, (b) 30 minutes of additional instructional time, and (c) explicit instruction, multiple opportunities for individual responses, and corrective feedback. Forty-three kindergarten students participated in the study, 20 of whom were determined to be at risk for language and literacy difficulties based on Peabody Picture Vocabulary Test-III (PPVT-III) scores below the 30th percentile. Students not at-risk and students at-risk participated in evidence-based classroom vocabulary instruction; however, at-risk students also participated in an additional supplemental vocabulary intervention while not at-risk students did not. Word knowledge of at-risk and not at-risk students was compared

using measures which assessed word recognition, picture vocabulary, context questions, and expressive definitions. Results on the initial posttest indicated that when all students received the same type and amount of classroom instruction, students who were not identified as at-risk obtained higher mean scores across all four vocabulary knowledge measures. However, when at-risk students were provided with additional supplemental vocabulary instruction, at-risk students obtained vocabulary knowledge scores that were comparable to scores obtained by typical students, not at-risk for language or literacy difficulties. Results from the 7 weeks delayed posttest were similar to that of the results obtained from the initial posttest.

Similarly, Loftus and Coyne (2013) conducted two studies aimed at identifying the effectiveness of a multi-tiered vocabulary intervention for Kindergarten students. Study one consisted of a class wide vocabulary intervention with 124 Kindergarten students (80 students received the treatment and 44 students served as the business as usual (BAU) control group). The class wide intervention included a story book reading activity and a post reading activity. Students were provided with two lessons per week for 18 weeks and focused on three target words per week. In the story book reading activity, students were introduced to the target word, provided with a student-friendly definition, shown a picture that illustrates the target word's meaning, and prompted to pronounce the target word. During the post-reading activities, the target words were reviewed and students were provided with examples and non-examples of the target word's meaning. Results indicated that the treatment group significantly outperformed the control group on measures of vocabulary knowledge. However, a delayed posttest to ascertain the

maintenance of these target words over time was not conducted. The second study examined the effectiveness of a Tier 2 vocabulary intervention that supplemented a Tier 1 vocabulary intervention with 43 kindergarten students. Students' at-risk status was determined based on PPVT-III scores below the 30th percentile. Twenty-three students were considered not at-risk and participated in the Tier 1 whole class story book reading vocabulary intervention (similar to that of study one) two times a week for 2 weeks targeting four words per week. Twenty students were considered at-risk and participated in the Tier 1 vocabulary instruction and in addition participated in a small group (consisting of three to four students) supplemental intervention for an additional 30 minutes two times a week. The supplemental instruction focused on only half of the four target words per week and consisted of a review of the target word's meanings, a repeat of the class wide activity, and two oral language activities. Results indicated that at-risk students learned the target words better when receiving additional Tier 2 instruction on those words as opposed to receiving only Tier 1 instruction. A 7-week delayed posttest revealed that students were able to maintain target word knowledge gained over the course of the intervention.

Marulis and Neuman (2010) conducted a meta-analysis on the effects of vocabulary intervention on young children's word learning to identify the effect of vocabulary interventions on the word learning of pre-kindergarten and kindergarten children. The meta-analysis included 67 studies with a total of 5,929 children included in the sample 60% of which were pre-kindergarten students. Of the studies included in the meta-analysis, 70% were published in a peer-reviewed journal. The results indicated that

vocabulary interventions had a large and significant effect on the word learning of pre-kindergarten ($g = 0.85$, $CI = 0.68, 1.01$ $p < 0.0001$) and kindergarten students ($g = 0.94$, $CI = 0.73, 1.14$, $p < 0.0001$). In addition, analyses conducted to determine word learning maintenance at delayed posttest (defined as measures administered 2-180 days after the end of the intervention) indicated that pre-kindergarten and kindergarten students maintained word knowledge of words learned through intervention over time.

Pullen, Tuckwiller, Konold, Maynard, and Coyne (2010) used the PPVT-4 receptive language assessment to identify students at risk of reading failure (> 39 th percentile = not at risk and ≤ 39 th percentile = at risk). A total of 224 first grade students participated in the study and were grouped into three categories: (a) not at risk for reading failure, (b) at-risk for reading failure treatment condition, and (c) at-risk for reading failure control group condition. Students' at-risk for reading failure were randomly assigned to either the treatment condition or the control group condition. Students in all three conditions received Tier 1 storybook reading instruction in their general education classroom twice a week for 30 minutes. Students in the at-risk treatment condition received a supplemental Tier 2 vocabulary intervention in small groups of two to five students for 20 minutes twice a week. Researcher-developed measures were used to assess students' vocabulary acquisition of taught words. Pullen et al. (2010) found that at the initial posttest (administered at the end of the 2-week intervention), students at-risk for reading failure who received the supplemental Tier 2 vocabulary instruction (treatment group) achieved significantly higher posttest scores on measures of vocabulary acquisition when compared to both the not-at-risk group and the at-risk group who did

not receive the supplemental intervention. Importantly for my study, the delayed posttest, administered 4 weeks after the intervention concluded, revealed that the at-risk students in both conditions achieved lower scores on measures of vocabulary acquisition when compared to students who were not at risk for reading failure. That is, at-risk students lost acquired vocabulary knowledge over time, whereas typical learners did not.

Duff et al. (2008) conducted a 9-week reading intervention with 12 eight-year-old students with severe and persistent reading difficulties who had not responded to a prior reading intervention conducted by Hatcher et al. (2006). The 9-week reading intervention incorporated reading, phonological, and vocabulary training. Over the course of 9 weeks, students participated in daily one-on-one instruction for two, 15-minute sessions.

Instruction was provided by a trained teaching assistant. The first session included 10 minutes of reading and 5 minutes of narrative writing. The second session included a 3 minute review of target vocabulary words learned in session one earlier that day, a 5 minute phonological awareness training session, a 3 minute session on teaching sight words through multi-sensory activities, a 3 minute reading session, and a 1 minute review of the day's target vocabulary words. Findings indicated that, at posttest, students made significant gains on measures of word reading, letter-sound knowledge, phoneme awareness, grammar, expressive language, and knowledge of taught vocabulary. At a 6 month follow up, gains made in reading, phoneme awareness, and vocabulary were maintained. Thus, the results of the Duff et al. (2008) study indicated that eight-year-old students, primarily in third or fourth grade, who consistently struggle in their reading development can acquire and maintain reading skill gains achieved through adequate

intervention support. However, as with the majority of reading intervention research, the focus was on elementary-aged students.

Zeller (2010) investigated the vocabulary retention of 63 third grade students 1 year after students participated in a second-grade vocabulary intervention. The vocabulary intervention was provided 4 days per week for 18 weeks and focused on a total of 108 Tier 2 target words selected from *The Academic Word List* (Coxhead, 2000). Students' retention of 42 of the 108 taught vocabulary words was measured using a decontextualized definition task (requiring students to express their knowledge of the target word) and an expressive labeling task (requiring students to identify the target word using a picture and the definition of the target word). Results indicated that on average students' decontextualized word knowledge decreased by 28% from initial posttest to the 1-year delayed posttest, whereas 64% of decontextualized word knowledge remained stable over time. In addition, on average, students' expressive word knowledge decreased by 32% from initial posttest to delayed posttest, whereas 64% of expressive word knowledge remained stable over time. That is, on average, students were able to retain decontextualized and expressive word knowledge for 64% or 27 words of the 42 words used in the delayed posttest 1 year after participating in a vocabulary intervention.

In addition, McLaughlin et al. (2000) conducted a multi-year intervention designed to improve the vocabulary knowledge and reading comprehension skills of fourth and fifth grade English Learners. The intervention was provided for 12 weeks with students receiving supplemental vocabulary instruction four days a week for 20-40 minutes. After the first 12 weeks of intervention in year two of the study, students

improved on measures of breadth of vocabulary knowledge, but not on measures of depth of vocabulary knowledge or on measures of reading comprehension. However, after a second year of intervention (an additional 12 weeks during students' fifth grade year), intervention students outperformed comparable students on measures of breadth and depth of vocabulary knowledge and on measures of reading comprehension. These findings suggest that adequate supplemental vocabulary instruction carried out over several years can narrow the gap between ELs and non-ELs in vocabulary knowledge and reading comprehension skills. However, the sample population in this study again focused on elementary aged children, and neither this study nor Zeller's (2010) included subjects eligible for special education.

Research focused on the vocabulary acquisition/maintenance of at-risk students has been conducted primarily with young children in elementary school grades. In addition, much of the vocabulary instruction research conducted with older at-risk populations has neglected to identify the long-term gains, if any, of vocabulary instruction with this population of students.

Turning to older subjects, Mastropieri, Scruggs, Levin, and McLoone (1985) investigated the effects of a pictorial mnemonic vocabulary intervention and the effects of direct instruction on the vocabulary acquisition of 32 7th, 8th, and 9th grade students with LDs. The results indicated that students' vocabulary acquisition was significantly greater in the mnemonic condition; however, maintenance of vocabulary knowledge was not evaluated as part of the study. In addition, Seifert and Espin (2012) examined the effects of a text reading intervention, vocabulary learning intervention, and a combined text

reading with vocabulary learning reading intervention on the science text reading of 20 10th grade students with LDs. The text reading intervention included components such as automatic word identification, reading aloud to an adult, adult modeling, error correction, and repeated reading of the text. In the vocabulary learning condition students were explicitly taught the meanings of 10 scientific terms encountered in 10th grade science reading texts. The combined condition contained elements from both the text reading and vocabulary learning conditions. Students also participated in a control condition in which no intervention was provided. All students participated in each of the four conditions over the course of 4 days with one condition delivered on each of 4 days for approximately 30 minutes. Results indicated student performance on measures of reading fluency and vocabulary knowledge was greater in the three intervention conditions as compared to the control condition. However, for measures of reading comprehension there was no difference between student performance in the intervention conditions and student performance in the control conditions. Given that reading comprehension is the most difficult reading skill to improve, these results are not surprising. In fact, these findings indicate that brief reading interventions can have an immediate positive effect on the text reading fluency and vocabulary knowledge of secondary school students with LDs; however, students' retention of vocabulary knowledge gained as part of the intervention is unknown.

Johnson, Gersten, and Carnine (1987) compared two methods of computer-assisted vocabulary instruction for teaching definitions of 50 words to 25 high school students in 9th through 12th grade with LDs over the course of a maximum of eleven 20-

minute sessions. The experimental method consisted of vocabulary sets of no more than seven words presented at a time and included daily reviews of learned words and periodic cumulative reviews of learned words to ensure retention. The comparison method consisted of vocabulary sets of 25 words at a time and did not include cumulative review of learned words. An experimenter-created multiple-choice measure of vocabulary acquisition was used to assess students' acquisition of the 50 taught vocabulary words. The results indicated that students in neither group differed in the average amount of vocabulary words learned or retained at posttest or at maintenance 2 weeks later, though there was a slight decline in maintenance vocabulary acquisition scores. However, there was a significant difference between conditions in the time it took students to reach mastery on the taught vocabulary words. Students in the experimental condition on average reached mastery on all 50 words after 7.6 sessions compared to 9.1 sessions for students in the comparison group. The results of these analyses indicated that 9th through 12th grade high school students with LD can acquire and maintain vocabulary knowledge more efficiently when words are presented in sets of no more than seven words and daily and cumulative review of learned words is provided, as compared to learning words in sets of 25 without cumulative review of learned words. However, maintenance of taught vocabulary words was assessed only 2 weeks after the conclusion of the intervention. Student vocabulary knowledge increases only if students are able to retain vocabulary knowledge of newly taught words. Because vocabulary knowledge plays such a critical role in reading comprehension, identifying the extent to which students retain vocabulary knowledge over time is essential. Studies that have investigated vocabulary retention with

secondary level students have failed to identify the effects of long term vocabulary maintenance on general reading outcomes, focusing primarily on maintenance of vocabulary words at a 2-week follow-up posttest. For instance, Stump, Lovitt, Fister, Kemp, Moore, and Schroeder (1992) conducted two studies investigating the effectiveness of a precision-teaching vocabulary intervention for secondary-level youth grades 7 through 12. For Study 1 a total of 351 students, 115 of whom were special education students, participated in the intervention. The intervention consisted of four phases: (a) introduction and explanation of unit vocabulary through whole-group instruction, (b) discussion of target vocabulary, (c) small-group instruction, and (d) independent seat work with target words. Results indicated that students improved significantly on measures of vocabulary knowledge from baseline to posttest for both general education and special education students. In addition, students retained target vocabulary knowledge at a 2-week delayed posttest. Study 2 replicated the intervention procedures from Study 1. A total of 343 students, 95 of whom were special education students, in grades 7 through 12 participated in the intervention. Results for Study 2 were similar to the results from Study 1, generating the same general findings differing only in degree of the significance of the findings.

Similarly, Xin and Rieth (2001) investigated the effects of video-assisted vocabulary instruction for 76 4th, 5th, and 6th grade students with LDs. Students were randomly assigned to video-assisted vocabulary instruction group or a BAU dictionary and printed materials vocabulary instruction group. Both groups learned word meanings and concepts in special education resource classrooms three times a week for six weeks

with each session lasting 30 minutes. The interactive video program used was a commercial videodisc that focused on the 1989 San Francisco earthquake. The 30 target words used in the study were selected from 4th, 5th, and 6th grade basal reading textbooks and were related to the context depicted in the selected videodiscs. Results indicated that at posttest students in the video instruction group had statistically higher word acquisition scores than those in the BAU control group. Similarly, at the 2-week follow-up posttest, students in the video instruction group had higher word maintenance scores than those in the BAU control group; however, the difference was not statistically significant.

The literature demonstrates a need for students with disabilities to retain meanings of vocabulary words in order to make gains on reading outcomes and accommodate the increased reading load experienced when transitioning to secondary school. In addition, the differential impacts of vocabulary instruction on language minority students at varying levels of English proficiency have also been documented in the literature. However, the literature on vocabulary retention research is lacking for students in secondary grades and for students with disabilities. Moreover, the literature has failed to demonstrate the differences in vocabulary retention over time for English minority students with LDs at various levels of English proficiency. The purpose of this paper is to examine differences between middle school NES and EL students at varying levels of ELP, vocabulary knowledge of taught vocabulary words across five-time points (i.e., pretest, immediate posttest, maintenance Y1 Spring 4 months after the initial posttest, maintenance Y2 Fall 11 months after the initial posttest and maintenance Y2 Spring 17

months after the initial posttest) for students with LDs participating in an evidence-based vocabulary intervention.

Research Questions

- 1) To what extent does word knowledge decrease, increase, or remain stable from pretest, to immediate posttest, maintenance Y1 Spring, maintenance Y2 Fall and maintenance Y2 Spring on measures of word knowledge for students with LDs from 6th grade to 7th grade for control and treatment groups?
- 2a) Are there significant differences in vocabulary knowledge for control and treatment groups on pretest, immediate posttest, maintenance Y1 Spring, maintenance Y2 Fall and maintenance Y2 Spring measures of vocabulary knowledge?
- 2b) Does the control group demonstrate statistically significant growth on vocabulary knowledge measures from Pretest to maintenance Y2 Spring?
- 3) Are there significant differences in vocabulary knowledge for treated English Learners (ELs) as compared to their treated NES peers on immediate posttest, maintenance Y1 Spring, maintenance Y2 Fall and maintenance Y2 Spring measures of vocabulary knowledge?
- 4) Are there significant differences in vocabulary knowledge for treatment group English Learners (ELs) at varying levels of ELP on immediate posttest, maintenance Y1 Spring, maintenance Y2 Fall and maintenance Y2 Spring measures of vocabulary knowledge?

5) Are there differences in generalized vocabulary outcomes for students in the control and treatment groups?

Method

Participants

Thirty-six sixth grade students from three middle schools in southern California participated in the study. The middle schools serve a large population of low income students, with 85% of the student population qualifying for Free/Reduced Lunch. In addition, 60% of students attending these middle schools are classified as ELs. Middle school A was used as a comparison group and did not receive the vocabulary intervention. Middle schools B and C served as the treatment group with the exception of four students in middle school B, who also served a Business as Usual (BAU) control participants. These four students were dispersed across classes and therefore could not be targeted for intervention. Eighteen students served as the BAU control group and eighteen students participated in the treatment. All study related activities occurred in participating students' English Language Arts (ELA) classes.

All participants were eligible for special education under the categories of Specific Learning Disability (SLD, $n = 28$), Speech/ Language Impairment (SLI, $n = 2$), Autism ($n = 3$), and Other Health Impairment (OHI, $n = 3$). All participating students were administered the verbal and matrix reasoning portion of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011) to measure general intelligence and the Woodcock-Johnson Tests of Achievement IV (WJ-IV; Schrank, McGrew, & Mather, 2014) to gather information related to reading skill using the Word Identification, Word

Attack, and Passage Comprehension subtests. These measures were used to identify any discernible differences between control and treatment groups. Participants were also administered the Comprehensive Receptive and Expressive Vocabulary Test (CREVT-3; Wallace & Hammill, 2013) to measure oral vocabulary with the receptive and expressive vocabulary subtests. Table 1 provides a summary of the baseline test scores for all participants. Independent t-tests resulted in no significant differences between BAU control and treatment groups on baseline measures (see Table 2).

Business as usual. Of the 18 BAU students, 11 (61%) were males and 7 (39%) were females. The ethnicity of the participants in the BAU group was Hispanic ($n = 18$, 100%). All 18 BAU students had identified disabilities. Fourteen participants were classified under SLD, one participant was classified under OHI, two participants were classified under Autism and one participant was classified under SLI. Five BAU students (28%) were proficient in English while 13 (72%) had limited ELP.

Treatment. Of the 18 treatment students, 17 (94%) were males and 1 (6%) was female. The ethnicity of the participants was predominantly Hispanic ($n = 13$, 72%) followed by Caucasian ($n = 3$, 17%), African American ($n = 1$, 5%) and one student declined to state ethnicity ($n = 1$, 5%). All eighteen treatment students had identified disabilities. Fourteen participants were classified under SLD, two participants were classified under OHI, one participant was classified under SLI and one participant was classified under Autism. Six (33%) treated students were proficient in English while 12 (67%) had limited English proficiency. Treatment and BAU groups were similar in terms of basic demographic information and baseline measure scores (see Tables 1 and 2).

Procedures

Participants in the treatment group received a vocabulary intervention in their 6th grade Special Education Language Arts class. Maintenance of the words taught over the course of the vocabulary intervention is the focus of this research study and was measured across five-time points: (a) pretest, (b) immediate posttest, (c) maintenance Y1 Spring 4 months after the initial posttest, (d) maintenance Y2 Fall 11 months after the initial posttest and (e) maintenance Y2 Spring 17 months after the initial posttest.

Intervention. As part of the Creating Habits Accelerating Academic Language of Students (CHAAOS) research in 6th grade, treated special education students received 12 weeks of vocabulary instruction aimed at broadening usage and practice opportunities of taught vocabulary words. The goals of the CHAAOS intervention included improving vocabulary acquisition and maintenance for students with disabilities as well as teacher implementation and fidelity of vocabulary lessons. CHAAOS includes several distinctive features: (a) teaching definitions students can understand, (b) commencing vocabulary intervention procedures in 6th grade in order to build a set of maintained academic vocabulary words by the end of 8th grade, (c) using a gradual release of responsibility model for assisting teachers who deliver vocabulary instruction, and (d) incorporating explicit and extended practice of vocabulary words into the intervention.

Students participating in the treatment group were taught 48 words over the course of 12 weeks. The 48 words were selected based on Coxhead's academic word list and the Common Core State Standards. Taught vocabulary words were grouped into three sets of 16 words, which were further grouped into sets of 4 words each. Vocabulary

instruction took place for approximately 20 minutes, 4 days a week (i.e., Monday to Thursday) and included a predictable routine that introduced four new words per week. During the first week of instruction, the Monday lesson included an introduction of the four new words using their synonyms, student-friendly definitions, and appropriate graphics demonstrating the usage of target words. The Tuesday lesson included a brief review of the four new words; however, two of the four words were emphasized in instruction. The Wednesday lesson also included a brief review of the four new words; however, the other two of the four words were emphasized in instruction. The Thursday lesson incorporated instruction and usage across all four words. Weeks 2-3 of instruction followed the same routine; however, previously taught words were used in contexts to introduce the four new words of the week and Thursday activities integrated words taught across the 3-week set. Intervention procedures from weeks 4-12 followed the same routine; however, Thursday activities integrated a minimum of eight previously taught words from across weeks.

Treatment integrity. An experimenter created treatment fidelity data sheet was used to collect treatment integrity data for 28 observations in which the treatment teachers were providing the intervention. Treatment integrity data components included the following: instructional components, instructional time, and quality of treatment implementation. Inter-rater reliability was established between two raters on the treatment fidelity data sheet at 92% percent agreement across eight observations. Treatment integrity data indicated that instructional components were present 100% of the time for all treatment teachers. Instructional time varied and ranged from 13-27

minutes per session. However, the goal of providing instruction for at least 15 minutes per session was met on most occasions with sessions averaging 23 minutes. The quality of implementation also varied but was generally high for all treatment teachers. Average quality of implementation for both teachers was 2.5 out of 3 (O'Connor et al., 2018 in press).

Assessments. An experimenter-developed vocabulary knowledge measure was used in this research study to measure students' retention of taught words. Students' receptive word knowledge was measured using a multiple-choice vocabulary assessment, which specifically targets the 18 of the 48 vocabulary words on which students made the most gains during the first year of intervention. Adequate reliability estimates for the experimenter-developed vocabulary knowledge measure were obtained with correlation coefficients ranging from .70-.72.

A standardized general vocabulary measure, CREVT-3, was used to measure students' general vocabulary growth over the course of one year. None of the words included in the CHAOS study are on the CREVT-3. The CREVT-3 is a frequently used norm-referenced standardized measure of receptive and expressive oral vocabulary. The CREVT-3 is an individually administered, comprehensive instrument for assessing general vocabulary knowledge for individuals ranging in age from 5-0 to 89-11. Criterion validity for the CREVT-3 was addressed by correlating the CREVT-3 with other measures of vocabulary knowledge including: The Expressive One-Word Picture Vocabulary Test-2000, the Receptive One-Word Picture Vocabulary Test-2000, The WORD Test-Second Edition, the Peabody Picture Vocabulary Test-Revised, the

Expressive One-Word Picture Vocabulary Test-Revised, and the Wechsler Intelligence Scale for Children-Third Edition. Average coefficients from the aforementioned correlations ranged from .72 to .87. Reliability evidence for the CREVT-3 was provided using coefficient alpha to estimate internal consistency. The CREVT-3 has excellent internal consistency as evidenced by coefficient alphas ranging from .85-.96. Test-retest reliability evidence for the CREVT-3 was also excellent with average correlation coefficients ranging from .84-.91.

Data Analysis

The first research question focuses on whether word knowledge on taught vocabulary words decreases, increases, or remains stable over time. To answer this question, descriptive statistics including means, standard deviations, and effect sizes for treatment and control groups were derived using the experimenter-developed multiple-choice vocabulary knowledge measures across five-time points (i.e., pretest, immediate posttest, maintenance-Spring Year 1, maintenance- Fall Year 2, and maintenance- Spring Year 2). Hedges g is a measure of effect size that indicates how much one group differs from another. Both Cohen's d and Hedges g produce upward biased results with small sample sizes; however, Hedges g with the correction factor tends to outperform Cohen's d with small sample sizes (Hedges, 1981). The following equations were used to calculate the aforementioned descriptive statistics:

Equation 1 $\bar{x} = \frac{\sum x}{n}$

where \bar{x} = *mean of* all values in the data set

$\sum x$ = *sum of all data values*

n = *number of data items in sample*

Equation 2 $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$ where $s = \text{standard deviation of sample}$

$\bar{x} = \text{mean of the } x_i$

$x_i = \text{each of the values of the data set}$

$n = \text{number of data items in sample}$

Equation 3 $g = \frac{\bar{x}_1 - \bar{x}_2}{Sp} \times \frac{N-3}{N-2.25} \times \sqrt{\frac{N-2}{N}}$ where $\bar{x}_1 = \text{the mean of sample 1}$

$\bar{x}_2 = \text{the mean of sample 2}$

$Sp = \text{the pooled standard deviation}$

Formula for pooled standard deviation $Sp = \frac{\sqrt{(n_1-1)s_1^2 + (n_2-1)s_2^2}}{(n_1-1) + (n_2-1)}$

Question 2a focuses on differences in vocabulary knowledge rates for control and treatment groups over time (i.e., pre-test, post-test, maintenance- Spring Year 1, Maintenance- Fall Year 2, and maintenance- Spring Year 2). Using M Plus version 8.1, a latent variable structure model was used to evaluate the relation between vocabulary retention rates and group status.

Question 2b focuses on the vocabulary knowledge growth of the BAU control group from pretest to the maintenance Y2 Spring measure. The BAU control group was expected to grow on measures of vocabulary knowledge over time due to selecting words for intervention based on Coxhead's academic word list and the Common Core State Standards. Using SPSS version 24, a paired samples t-test was used to identify if the BAU control group demonstrated significant growth on measures of vocabulary knowledge over time.

The third research question focuses on differences in vocabulary knowledge for students identified as ELs in the treatment group and students not identified as ELs in the treatment group over time (i.e., pre-test, post-test, maintenance -Spring Year 1, Maintenance- Fall Year 2, and maintenance- Spring Year 2). Using M Plus version 8.1, a latent variable structure model was used to evaluate the relation between vocabulary knowledge and language status.

The fourth research question focuses on differences in vocabulary knowledge for students identified as ELs at varying levels of ELP over time (i.e., pre-test, post-test, maintenance- Spring Year 1, Maintenance- Fall Year 2, and maintenance- Spring Year 2). Using M Plus version 8.1, a latent variable structure model was used to evaluate the relation between vocabulary knowledge and California English Language Development Test (CELDT) status. CELDT proficiency level scores were used to cluster students into level 1, 2, or 3. The subject pool did not include any students with CELDT proficiency level scores of 4 or 5, thus only clusters of CELDT proficiency levels of 1, 2, and 3 were used in these analyses.

The fifth research question focuses on identifying differences in general vocabulary outcome measures between students in the control and treatment groups as they relate to number of words maintained at the Spring Y2 delayed posttest measure. To answer this question, a correlation between outcome scores on the delayed posttest measure (Spring Y2) and on the generalized vocabulary outcome measure (CREVT-3) was conducted for both control and treatment groups.

Results

Research Question 1

Means, standard deviations, and effect size estimates for each time point broken down by treatment condition are presented in Table 3. At pretest, the BAU control and treatment group did not differ significantly indicating that prior to intervention students in both groups had similar vocabulary knowledge levels across the 18 target words. However, at posttest, the treatment group obtained average scores that were significantly greater than that of the control group. The magnitude of the effect at posttest was large ($g = 1.286$) indicating that the CHAOS vocabulary intervention had a large positive effect on the vocabulary knowledge of taught words of treatment group students. The magnitude of the effect at the 4-month delayed maintenance measure (Maintenance Spring-Y1) could not be determined because 4-month delayed maintenance data were not collected for the BAU group; however, average scores for the treatment group at the initial posttest and at the 4-month delayed maintenance measure remained relatively stable. At the 11-month delayed maintenance measure (Maintenance Fall-Y2) the magnitude of the effect was again large ($g = 1.071$) indicating that the CHAOS vocabulary intervention had a large positive effect on the vocabulary knowledge of treatment group students even after an 11-month period. At the 17-month delayed maintenance measure (Maintenance Spring-Y2) the magnitude of the effect was again large ($g = 0.904$) indicating that the CHAOS vocabulary intervention had a large positive effect on the vocabulary knowledge of treatment group students even after a 17-month period.

Research Question 2a

A latent variable structure model was first fitted to the data. The model was a latent difference score model, shown in Figure 1. In this model, each of the five vocabulary manifest variables from pretest to Spring-Y2 is shown in its own rectangle. Associated with each manifest variable is a latent variable; for example, the latent variable Status Pre is associated with the pretest manifest variable. The Status Pre variable is an error-free latent variable, because measurement error is partialled out through specification of the measurement error term θ_{11} . This form of measurement structure is used for each of the five manifest vocabulary scores, pretest through Spring-Y2, with measurement error variance constrained to equality across times of measurement. Then, latent status at each time after the pretest is represented as a function of latent status at the previous time of measurement plus a latent difference (or change) score. For example, at the second time of measurement, Status Post is a function of Status Pre plus Diff Post (which is the difference score at the posttest). Because difference scores are dependent on scores at the previous time of measurement, autoprotection parameters β_1 through β_4 , are specified to account for this dependence. The mean parameters are α parameters, with α_1 representing the control group mean at pretest, α_2 through α_5 the intercepts of the difference scores Diff Post through Diff S-Y2, and α_6 the mean of the treatment indicator. Treatment effects are represented by β coefficients, with β_5 the treatment group difference from the BAU group at pretest, and β_6 through β_9 the treatment effect on difference scores at the four succeeding times of measurement. Finally, the latent variable variances were represented by ψ parameters, with ψ_{11}

representing individual difference variance in Status Pre, and ψ_{22} through ψ_{55} representing individual difference variance in difference scores at succeeding times of measurement.

All paths in Figure 1 associated with path coefficients of 0 or 1 had coefficients fixed to those values to identify the model. In addition, the measurement residual variances θ_{11} through θ_{55} were constrained to equality to identify the model. All 21 remaining parameters – α_1 through α_6 , β_1 through β_9 , ψ_{11} through ψ_{55} , and θ_{11} through θ_{55} (with θ estimates constrained equal) – were freely estimated.

Model 1 (see Figure 1) was the initial latent difference score model that was specified as shown in Figure 1. As discussed above, this model had 21 parameter estimates. In terms of model fit (see Table 4), the root mean square error of approximation (RMSEA), was 0.000, less than 0.08, and thus considered a good fit (Grimm, Ram, & Estabrook, 2016). The comparative fit index (CFI) and the Tucker-Lewis index (TLI), were 1.000 and 1.039, respectively which indicate good model fit (Grimm, Ram, & Estabrook, 2016). Thus, Model 1 provides very good fit to the data.

Model 2 was identical to Model 1 except that the four autoprotection parameters were fixed at 0. If this constraint did not harm fit, the resulting model would be essentially identical to a latent growth model. As shown in Table 4, the statistical fit of Model 2 was very poor, with $\chi^2(10) = 21.95$, $p < .01$, and the change in model fit was also significant, $\Delta\chi^2(4) = 18.09$, $p < .0001$. Further, all practical fit indices fell in unacceptable ranges, suggesting that the autoprotection parameters needed to be retained

in the model and that a latent difference score model was more appropriate than a latent growth model for these data.

Model 3 is similar to Model 1; however, the effects of treatment on latent difference scores at times 3, 4, and 5 were fixed equal to 0. That is, the model was constrained to allow for treatment effects at time 1 and, importantly, at time 2, but no addition effects on later times of measurement. Ideally, this model would exhibit a trivial difference in performance at time 1, a significant treatment effect on the difference score at time 2; and the zero treatment effects at the remaining three times of measurement would suggest maintenance of the treatment effect at time 2. As shown in Table 4, the fit of Model 3 was very good, with $\chi^2(9) = 4.13$, *ns*, and the change in model fit was not significant, $\Delta\chi^2(3) = 0.27$, *ns*. All model fit indices for this model also indicated good model fit: RMSEA = 0.000, CFI = 1.000, and TLI = 1.060. The non-significant change in model fit supports the contention that no residual treatment effects were apparent at times 3 through 5.

Model 4 was similar to Model 3, but added a constraint of equality of residual variances of the latent difference scores at times 2 through 5. The overall model fit for Model 4 was very good, and the change in fit relative to Model 3 was not significant, $\Delta\chi^2(3) = 0.29$, *ns*. As with Model 3, all model fit indexes for Model 4 indicated good model fit: RMSEA = 0.000, CFI = 1.000, and TLI = 1.070. This model is efficient and has a relatively small number of parameter estimates. Given the excellent fit of this model, Model 4 is the optimal model for these data.

Parameter estimates for Models 1 and 4 are shown in Table 5. Here, I will mention the key parameter estimates in these models. In Model 1, the BAU group had a mean performance of $\alpha_1 = 4.28$ ($SE = 0.62$) at pretest, and the treatment condition scored only slightly and non-significantly higher at pretest, $\beta_5 = 0.28$ ($SE = 0.88$). The key theoretical parameter was the treatment effect at posttest, which was large and significant, $\beta_6 = 5.86$ ($SE = 1.24$). Notably, the treatment effects at times 3 through 5, represented by β_7 through β_9 , were relatively small and non-significant.

The parameter estimates for Model 4 tend to be similar to those for Model 3, but with standard errors that tended to be smaller for most estimates. The key parameter estimates remained largely unchanged, with BAU group having a mean performance of $\alpha_1 = 4.28$ ($SE = 0.63$) at pretest, the treatment condition scoring only slightly and non-significantly higher at pretest, $\beta_5 = 0.28$ ($SE = 0.88$), and the treatment effect at posttest being large and significant, $\beta_6 = 6.00$ ($SE = 1.17$). As shown in Table 5, the treatment effects at times 3 through 5, represented by β_7 through β_9 , were fixed at 0. Further, the equality constraint on the difference score residual variances ψ_{22} through ψ_{55} led to substantial reductions in the SE s for these parameters, indicating more precise estimates of these values. Point estimates for remaining parameters in the model were little changed from values in Model 1, but tended to have improved, with smaller SE s. Thus, Model 4 provides a much more efficient model for the data than does Model 1, given the more tightly constrained sets of parameter estimates. Models 5 and 6 will be discussed in relation to questions 3 and 4.

Research Question 2b

To test the hypothesis that the pretest ($M = 4.28$, $SD = 2.44$) and Spring Y2 Maintenance means ($M = 8.28$, $SD = 3.48$) were equal, a paired samples t-test was performed. Prior to conducting the analysis, the assumption of normal distribution of difference scores was examined. The assumption was considered satisfied, as the skew and kurtosis levels were estimated at $-.657$ and $.358$, respectively, which is less than the maximum allowable values for a t-test (i.e., skew $< |2.0|$ and kurtosis $< |9.0|$; Posten, 1984). It will also be noted that the correlation between the two conditions was estimated at $r = .675$, $p < .05$, suggesting that the dependent samples t-test is appropriate in this case. The null hypothesis of equal pretest and Spring Y2 Maintenance means was rejected, $t(18) = -6.61$, $p < .001$. Thus, the Spring Y2 Maintenance mean was significantly higher than the pretest mean (see Table 6).

Research Question 3

Model 5 was based on the efficient Model 4 specification, but added the language variable (i.e., English as a second language) as a covariate to determine whether language had any effect on vocabulary performance at pretest or on difference scores at times 2 through 5. As shown in Table 4, the fit of Model 5 was very good, with a non-significant chi-square statistic and practical fit indices that were in fully acceptable ranges. None of the effects of the language variable was significant. The effect of language on pretest performance was nonsignificant, $\beta = -.96$ ($SE = 0.95$), $z = -1.00$, and the effects of language on the four difference scores were also non-significant, all z s $< |1.00|$, *ns*.

Thus, language did not have an appreciable effect on vocabulary performance at any of the times of measurement.

Research Question 4

Model 6 was very similar to Model 5, but substituted the CELDT variable as a covariate to determine whether verbal achievement in English was related to vocabulary performance at pretest or to difference scores at times 2 through 5. As shown in Table 4, the fit of Model 6 was very good, with a non-significant chi-square statistic and practical fit indices that were in fully acceptable ranges. None of the effects of the CELDT variable was significant. The effect of the CELDT variable on pretest performance was nonsignificant, $\beta = 0.28$ ($SE = 0.36$), $z = 0.78$, and the effects of the CELDT variable on the four difference scores were also non-significant, all z s $< |1.60|$, *ns*. Thus, ELP status did not have an appreciable effect on vocabulary performance at any of the times of measurement.

Research Question 5

Differences in generalized vocabulary outcomes for students in the control and treatment groups as they relate to number of words maintained on the Spring Y2 delayed posttest measure can be seen in Figure 2. For both treatment and control groups higher scores on the CREVT-3 standardized vocabulary outcomes were associated with higher scores on the Spring Y2 delayed posttest measure. However, the treatment group outperformed the control group on the Spring Y2 delayed posttest measure and thus the linear trend line for the treatment group has a steeper slope as compared to the control group. In addition, the Pearson correlation coefficient for the control ($r = 0.30$) and

treatment ($r = 0.60$) groups differed with the treatment group outperforming the control group.

Discussion/ Implications

Research Question 1

As expected, the word knowledge across time points differed for BAU control and treatment groups. Generally speaking, the intervention had a strong positive effect at the initial posttest which remained relatively stable across all maintenance time points (i.e., maintenance-Spring Year 1, maintenance- Fall Year 2, and maintenance- Spring Year 2). Specifically, Hedges' g can be interpreted using Cohen's suggested guidelines with 0.2 indicating a small effect, 0.5 indicating a medium effect, and 0.8 indicating a large effect (Cohen, 1977). However, Cohen emphasized caution when using the aforementioned guidelines as a "small" effect in one scenario may not necessarily be a "small" effect in a different scenario. That said, Durlak (2009) suggests referring to prior studies of similar content to identify how your results compare to the larger literature. Thus, per Cohen's guidelines the effects (i.e., the magnitude of the difference between treatment and BAU control groups) were small at pretest and large at posttest, at maintenance-Spring Year 1, at maintenance-Fall Year2, and at maintenance- Spring Year 2. Comparing across the vocabulary maintenance literature, the CHAOS intervention demonstrates larger effects that maintain overtime.

For example, as mentioned earlier, Loftus et al., 2010 investigated the effectiveness of a kindergarten vocabulary intervention designed to supplement classroom vocabulary instruction immediately after the vocabulary intervention and 7

weeks later to assess maintenance of word knowledge. At posttest Cohen's d effect size estimates across measures of word recognition, picture vocabulary, context questions and expressive definitions ranged from 0.08 (small effect) to 0.69 (medium effect). At the 7-week delayed posttest Cohen's d effect size estimates across measures of word recognition, picture vocabulary, context questions and expressive definitions ranged from 0.48 to 0.68. In another study, Pullen et al., 2010 investigated the effectiveness of a first-grade Tier 2 vocabulary intervention. At posttest Cohen's d effect size estimates across receptive, context, and expressive vocabulary measures ranged from 0.37 to 0.64. At the 4-week delayed posttest Cohen's d effect size estimates across receptive, context, and expressive vocabulary measures ranged from 0.20 to 0.38.

However, the aforementioned comparison studies focused on younger age groups and subjects without disabilities making the effect size comparisons difficult to interpret. In a study of 10th grade students with LDs, Seifert and Espin (2012) examined the effects of a text reading intervention, vocabulary learning intervention, and a combined text reading with vocabulary learning reading intervention on reading of science text. At posttest Cohen's d effect size estimates on the vocabulary knowledge measure demonstrated a large positive effect ($d = 1.11$); however, students' retention of vocabulary knowledge gained as part of the intervention is unknown. Thus, both by Cohen's guidelines and by relevant literature comparisons, the CHAOS intervention demonstrate larger positive effects that maintain of time.

Research Questions 2a

The treatment had a large and significant effect at posttest; however, the treatment effect at times 3 through 5 were relatively small and non-significant. That is, the treatment group made initial gains at posttest and maintained these gains over time. By comparison, the control group did not demonstrate substantial gains at posttest and word knowledge remained relatively stable across the five time points for this group. These findings indicate that the CHAAOS vocabulary intervention was effective for improving the vocabulary acquisition and vocabulary maintenance of middle school students with disabilities.

These results have direct implications for middle school classroom teachers. The intervention provided to students as part of this research study was developed using empirical evidence and was implemented by classroom Special Education teachers. In addition, The CHAAOS intervention incorporated grade appropriate words and treatment group students learned and retained these words even though their reading level was several grade levels below these academic words.

Research Question 2b

The BAU control group was expected to grow in vocabulary knowledge of the target words because they were selected based on Coxhead's academic word list and the Common Core State Standards. A statistically significant difference between the BAU control pretest and BAU control Spring Y2 Maintenance vocabulary knowledge measure was found as expected. However, the average vocabulary word knowledge for BAU control group students at the fifth time on measurement (i.e., Spring Y2 Maintenance)

was over 20% less than that of the treatment group. These findings have direct implications for middle school Special Education teachers. By incorporating a 20-minute vocabulary intervention into classroom instruction for just 12 weeks, students make significantly larger gains on measures of vocabulary knowledge than would be expected without intervention.

Research Question 3 and 4

The effects of English as a second language/ ELP have been shown to impact the extent to which students with and without special education needs benefit from research-based vocabulary instruction (Hwang et al., 2015). Models 5 and 6 were based on the efficient Model 4 specification, but added the English as a second language variable as a covariate and added the CELDT score of ELP as a covariate respectively to Model 5 and 6. The results of these analyses indicated that learning English as a second language and proficiency in English did not have a meaningful effect on vocabulary performance at any time of measurement in this study. That is, statistically significant differences in vocabulary maintenance for treated ELs and treated NES groups on pretest, posttest, maintenance Y1 Spring, maintenance Y2 Fall and maintenance Y2 Spring measures were not found on measures of vocabulary knowledge.

Similarly, statistically significant differences in vocabulary maintenance for ELs in the treatment group at varying levels of ELP were not found. These findings indicate that significant differences between BAU control and treatment groups found at post-test and delayed maintenance measures are not affected by classification as an English learner or level of English proficiency. That is, student's English as a second language and ELP

status did not negatively affect student performance at any time of measurement. Thus, the results of this study demonstrate that classroom Special Education teachers can generate strong positive results in the vocabulary learning of groups of students whom are largely limited in their ELP and meet special education classification under SLD using a vocabulary intervention that takes only 20 minutes a day and that these effects are maintained over time. By implementing such interventions, students with disabilities may be able to retain vocabulary knowledge learned over time, which may allow them to make gains on reading comprehension outcomes and accommodate the increased reading load experienced when transitioning to secondary school. This possibility deserves research attention.

Research Question 5

The correlations between the CREVT-3 and the Spring Y2 delayed maintenance measure for the control and treatment group indicate that students who were able to maintain a higher degree of vocabulary knowledge over the course of one school year were also able to obtain higher generalized vocabulary outcome scores. That is, the CHAOS intervention not only impacted students' ability to improve on the researcher developed measures of vocabulary acquisition but also had a positive effect on improving students' generalized vocabulary as measured by standardized measure of vocabulary knowledge. These findings are important given that researcher-developed measures typically yield larger effect sizes than standardized measures (National Reading Panel, 2000; Swanson, Hoskyn, & Lee, 1999). Although several studies have demonstrated small to large effects on standardized vocabulary measures (Duff et al., 2008; Loftus &

Coyne, 2013; Marulis & Neuman, 2010), such findings are not commonly found, especially for students in special education ELA classes.

Limitations

A few limitations of the study need to be noted. One limitation concerns the relatively small sample size of the study. The main limitation with small sample sizes is the interpretation of the results. That is, studies with small sample sizes tend to result in large standard errors resulting in wide confidence intervals and imprecise estimates of the effect. However, sample size in longitudinal research cannot always be controlled by the researcher. In this particular study the sample consisted of a group of students followed over the course of two years and the sample size at the end of Y2 diminished due primarily to students moving to different schools, making it impossible to continue data collection for those students.

A second limitation relates to the decision to measure vocabulary knowledge maintenance for only 18 of the 48 words taught during Y1 of the intervention. However, given the constraints of the testing situation and the large amount of testing students went through in the intervention study, the maintenance assessment needed to be a relatively quick assessment. Thus, we decided to focus only on those words for which students showed the most growth. In addition, for the purpose of the present study, the focus was exclusively on breadth of vocabulary knowledge maintained over time; however, depth of vocabulary knowledge maintained over time is also a phenomenon worth studying with this population of students. Furthermore, a single study like the present one cannot be

definitive concerning the impact of the CHAAOS intervention on vocabulary maintenance over time.

Conclusion

The literature demonstrates a need for students with disabilities to retain vocabulary knowledge in order to make gains on reading comprehension outcomes and accommodate the increased reading load experienced when transitioning to secondary school. The results of these analyses indicate that with the implementation of 12 weeks of 20-minute evidence-based vocabulary intervention 4 days a week, students can make significant gains on vocabulary knowledge and maintain those gains over time. Moreover, findings indicated that the CHAAOS intervention not only positively impacted students' ability to improve on a researcher developed measure of vocabulary acquisition, but also had a positive effect on improving student generalized vocabulary knowledge. Furthermore, based on a thorough review of the literature, though there are several studies that have demonstrated small to large effects on standardized vocabulary measures, such findings are not common with the population of students included in this study.

References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The psychology of learning and motivation: II; the psychology of learning and motivation: II* (pp. 249-Chapter xi, 249 Pages) Academic Press, Oxford.[http://dx.doi.org/10.1016/S0079-7421\(08\)60422-3](http://dx.doi.org/10.1016/S0079-7421(08)60422-3)
- Baker, S.K., Simmons, D.C., & Kame'enui, E.J. (1995). *Vocabulary acquisition: Synthesis of the research*. (Tech. Rep. No. 13). Eugene: University of Oregon, National Center to Improve the Tools of Educators.
- Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, 107(3), 251-271. <http://dx.doi.org/10.1086/511706>
- Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, 98(1), 44-62. doi:<http://dx.doi.org/10.1037/0022-0663.98.1.44>
- Bos, C. S., & Anders, P. L. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning disabled students. *Learning Disability Quarterly*, 13(1), 31-42. doi:<http://dx.doi.org/10.2307/1510390>
- Bulgren, J. A., Kansas, P., & Deshler, D. D. (2013). Literacy challenges and opportunities for students with learning disabilities in social studies and history. *Learning Disabilities Research & Practice*, 28(1), 17-27. doi:<http://dx.doi.org/10.1111/ldrp.12003>
- Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of reading behavior*, 26(4), 413-437. doi:<http://dx.doi.org/10.1080/10862969409547861>
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed. ed.) Lawrence Erlbaum Associates, Inc, Hillsdale, NJ. Retrieved from <https://search.proquest.com/docview/617353861?accountid=14521>
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238. doi:10.2307/3587951
- Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly*, 30(2), 74-88. doi:<http://dx.doi.org/10.2307/30035543>

- Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, 110(1), 1-18. doi:<http://dx.doi.org/10.1086/598840>
- Duff, F. J., Fieldsend, E., Bowyer-Crane, C., Hulme, C., Smith, G., Gibbs, S., & Snowling, M. J. (2008). Reading with vocabulary intervention: Evaluation of an instruction for children with poor response to reading intervention. *Journal of Research in Reading*, 31(3), 319-336. doi:<http://dx.doi.org/10.1111/j.1467-9817.2008.00376.x>
- Durlak, J. A. (2009). How to select, calculate and interpret effect sizes. *Journal of Pediatric Psychology*, 34(9), 917-928. doi:<http://dx.doi.org/10.1093/jpepsy/jsp004>
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications.
- Hammill, D. D., & Larsen, S. C. (2009). *Test of written language 4 (TOWL-4)*. San Antonio, TX: Pearson Assessments.
- Hatcher, P. J., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., Gibbs, S., & Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: a randomized controlled trial. *Journal of Child Psychology and Psychiatry*, 47(8), 820-827. doi:<http://dx.doi.org/10.1111/j.1469-7610.2005.01559.x>
- Hedges, L. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics*, 6(2), 107-128. doi:10.2307/1164588
- Hwang, J. K., Lawrence, J. F., Mo, E., & Snow, C. E. (2015). Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of English proficiency. *International Journal of Bilingualism*, 19(3), 314-332. doi:<https://doi.org/10.1177/1367006914521698>
- Jitendra, A. K., Edwards, L. L., Sacks, G., & Jacobson, L. A. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children*, 70(3), 299-322. doi:<http://dx.doi.org/10.1177/001440290407000303>
- Johnson, G., Gersten, R., & Carnine, D. (1987). Effects of instructional design variables on vocabulary acquisition of LD students: A study of computer-assisted instruction. *Journal of Learning Disabilities*, 20(4), 206-213. <http://dx.doi.org/10.1177/002221948702000402>
- Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of*

Educational Psychology, 93(3), 451-455. doi:<http://dx.doi.org/10.1037/0022-0663.93.3.451>

Lackaye, T. D., & Margalit, M. (2006). Comparisons of achievement, effort, and self-perceptions among students with learning disabilities and their peers from different achievement groups. *Journal of learning disabilities*, 39(5), 432-446. doi:<http://dx.doi.org/10.1177/00222194060390050501>

Loftus, S. M., & Coyne, M. D. (2013). Vocabulary instruction within a multi-tier approach. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 29(1), 4-19. doi:<http://dx.doi.org/10.1080/10573569.2013.741942>

Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice*, 25(3), 124-136. doi:<http://dx.doi.org/10.1111/j.1540-5826.2010.00310.x>

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, 80(3), 300- 335. doi:<http://dx.doi.org/10.3102/0034654310377087>

Mastropieri, M. A., Scruggs, T. E., Levin, J. R., Gaffney, J., & McLoone, B. (1985). Mnemonic vocabulary instruction for learning disabled students. *Learning Disability Quarterly*, 8(1), 57-63. doi:<http://dx.doi.org/10.2307/1510908>

McLaughlin, B., August, D., Snow, C. E., Carlo, M., Dressler, C., White, C., et al. (2000). Vocabulary improvement and reading in English language learners: An intervention study. Proceedings of a Research Symposium on High Standards in Reading for Students from Diverse Language Groups. Washington, DC: U.S. Department of Education.

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders*, 41(3), 335-354. doi:<http://dx.doi.org/10.1080/13682820600602295>

O'Connor, R. E., Beach, K. D., Sanchez, V. M., Bocian, K. M., & Flynn, L. J. (2015). Building BRIDGES: A design experiment to improve reading and United States history knowledge of poor readers in eighth grade. *Exceptional Children*, 81(4), 399-425. <https://doi.org/10.1177/0014402914563706>

- Pearson, P. D., & Gallagher, M. C. (1983). The instruction of reading comprehension *Contemporary educational psychology*, 8(3), 317-344.
doi:[https://doi.org/10.1016/0361-476X\(83\)90019-X](https://doi.org/10.1016/0361-476X(83)90019-X)
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67-86.
- Phillips, D. C. K., Foote, C. J., & Harper, L. J. (2008). Strategies for effective vocabulary instruction. *Reading improvement*, 45(2), 62.
- Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice*, 25(3), 110-123. doi:<http://dx.doi.org/10.1111/j.1540-5826.2010.00309.x>
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.
- Sedita, J. (2005). Effective vocabulary instruction. *Insights on Learning Disabilities*, 2(1), 33-45.
- Seifert, K., & Espin, C. (2012). Improving reading of science text for secondary students with learning disabilities: Effects of text reading, vocabulary learning, and combined approaches to instruction. *Learning Disability Quarterly*, 35(4), 236-247. doi:<http://dx.doi.org/10.1177/0731948712444275>
- Simmons, D.C., & Kameenui, E.J. (1990). The effect of task alternatives on vocabulary knowledge: A comparison of students with and without learning disabilities. *Journal of Learning Disabilities*, 23, 291-297.
doi:<http://dx.doi.org/10.1177/002221949002300508>
- Stump, C. S., Lovitt, T. C., Fister, S., Kemp, K., Moore, R., & Schroeder, B. (1992). Vocabulary intervention for secondary-level youth. *Learning Disability Quarterly*, 15(3), 207-222. <http://dx.doi.org/10.2307/1510244>
- Stahl, S. (1983). Differential word knowledge and reading comprehension. *Journal of Literacy Research*, 15(4), 33-50.
doi:<http://dx.doi.org/10.1080/10862968309547495>
- Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes* Guilford Press, New York, NY. Retrieved from <https://search.proquest.com/docview/619399491?accountid=14521>

- Torgesen, J. K. (1980). Conceptual and educational implications of the use of efficient task strategies by learning disabled children. *Journal of Learning Disabilities, 13*(7), 364-371. <http://dx.doi.org/10.1177/002221948001300704>
- Wallace G, Hammill DD. (2013). *Comprehensive Receptive and Expressive Vocabulary Test-Third Edition (CREVT-3)*. Austin, TX: Pro-Ed.
- Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II)*. San Antonio, TX: NCS Pearson. DOI: 10.1177/0734282912467756
- Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children, 78*(1), 89-106.
doi:<http://dx.doi.org/10.1177/001440291107800106>
- Xin, J. F., & Rieth, H. (2001). Video-assisted vocabulary instruction for elementary school students with learning disabilities. *Information Technology in Childhood Education Annual, 12*, 87-103. Retrieved from <https://search.proquest.com/docview/619647940?accountid=14521>
- Zeller, A. (2011). *Vocabulary retention of third grade students from Low-Income homes following second grade vocabulary instruction* (Doctoral dissertation, The Ohio State University).

Tables

Table 1

Demographic Information and Baseline Measures Scores

Variable	Statistic	Treatment Group	Business As Usual
Gender			
Males	<i>N</i>	17	11
Females	<i>N</i>	1	7
Ethnicity			
Hispanic	<i>N</i>	13	18
Caucasian	<i>N</i>	3	0
African American	<i>N</i>	1	0
Declined to State	<i>N</i>	1	0
SPED Classification			
SLD	<i>N</i>	14	14
SLI	<i>N</i>	1	1
Autism	<i>N</i>	1	2
OHI	<i>N</i>	2	1
ELP			
Proficient	<i>N</i>	6	5
Limited Proficiency	<i>N</i>	12	13
WASI-II	<i>M (SD)</i>	76.65 (15.99)	79.06 (11.03)
WJ-IV Word ID	<i>M (SD)</i>	74.13 (10.13)	65.00 (13.43)
Word Attack	<i>M (SD)</i>	59.93 (14.93)	71.63 (14.91)
Passage Comprehension	<i>M (SD)</i>	59.00 (06.89)	65.67 (12.68)
CREVT-3	<i>M (SD)</i>	76.62 (09.51)	75.89 (07.81)

Note: SPED = Special Education; SLD = Specific Learning Disability, SLI = Specific Learning Impairment; OHI = Other Health Impairment; ELP = English Language Proficiency; WASI-II = Wechsler Abbreviated Scale of Intelligence-II; SD = Standard deviation; WJ-IV = Woodcock-Johnson Tests of Achievement IV; Word ID = Word Identification; CREVT-3 = Comprehensive Receptive and Expressive Vocabulary Test-3.

Table 2

Treatment vs. BAU Control Independent Samples t-test

	t	df	p-value	Mean Difference	Std. Error Difference	95% Confidence Interval	
						Lower Bound	Upper Bound
WASI	-0.521	33	0.606	-2.409	4.621	-11.810	6.993
WJ- Word ID	-0.934	31	0.357	-4.846	5.187	-15.424	5.733
WJ- Word Attack	0.428	31	0.671	2.232	5.209	-8.392	12.855
WJ- Passage Comp	-1.886	31	0.069	-8.529	4.523	-17.753	0.694
CREVT-3	0.248	32	0.806	0.736	2.972	-5.317	6.789

Note: WASI-II = Wechsler Abbreviated Scale of Intelligence-II; WJ = Woodcock-Johnson Tests of Achievement IV; Word ID = Word Identification; CREVT-3 = Comprehensive Receptive and Expressive Vocabulary Test-3; df = degrees of freedom.

Table 3
Descriptive Statistics for Treatment and BAU Control Group

	Control (BAU)		Treatment		<i>g</i>
	M	SD	M	SD	
Pretest	4.28	2.45	4.56	2.98	0.108
Posttest	6.28	4.17	12.67	5.22	1.286
Maintenance Spring Y1	--	--	12.06	5.63	--
Maintenance Fall Y2	7.17	4.29	11.78	3.89	1.071
Maintenance Spring Y2	8.28	3.48	12.22	4.70	0.904

Note: M = Mean; SD = Standard Deviation; *g* = Hedges *g* effect size with correction factor.

Table 4

Model Fit Indicators

	Chi-squared	df	p-value	RMSEA (90% CI)	CFI	TLI	SRMR
Model 1	3.86	6	0.69	0.000 (0.000, 0.166)	1.000	1.039	0.026
Model 2	21.95	10	0.01	0.182 (0.076, 0.286)	0.912	0.868	0.189
Model 3	4.13	9	0.90	0.000 (0.000, 0.079)	1.000	1.060	0.039
Model 4	4.42	12	0.97	0.000 (0.000, 0.000)	1.000	1.070	0.035
Model 5	4.10	13	0.99	0.000 (0.000, 0.000)	1.000	1.108	0.028
Model 6	3.88	12	0.99	0.000 (0.000, 0.000)	1.000	1.106	0.021

Note: df = degrees of freedom; RMSEA = Root Mean Square Error of Approximation; CI = Confidence Interval; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residuals.

Table 5

Estimates of Parameters for Two Latent Difference Score Models

Parameters	Model 1	Model 4
Means		
α_1	4.28 (0.62)	4.28 (0.63)
α_2	- 1.85 (2.82)	- 1.45 (1.67)
α_3	- 1.89 (1.99)	- 1.58 (1.64)
α_4	4.14 (1.30)	3.99 (1.03)
α_5	1.72 (1.27)	1.80 (1.18)
α_6	0.50 (0.08)	0.50 (0.08)
Autoproportion parameters		
β_1	0.90 (0.63)	0.79 (0.34)
β_2	0.05 (0.14)	0.06 (0.13)
β_3	- 0.35 (0.10)	- 0.35 (0.09)
β_4	- 0.08 (0.15)	- 0.11 (0.12)
Treatment effects		
β_5	0.28 (0.88)	0.28 (0.88)
β_6	5.86 (1.24)	6.00 (1.17)
β_7	0.71 (1.62)	0.00 (-----)
β_8	- 0.17 (1.30)	0.00 (-----)
β_9	- 0.28 (1.13)	0.00 (-----)
Variances		
ψ_{11}	4.57 (2.09)	4.79 (1.73)
ψ_{22}	2.33 (5.84)	3.24 (1.42)
ψ_{33}	2.79 (2.57)	3.24 (1.42)
ψ_{44}	3.47 (1.88)	3.24 (1.42)
ψ_{55}	2.60 (2.60)	3.24 (1.42)
Measurement residuals		
$\theta_{11}, \theta_{22}, \theta_{33}, \theta_{44}, \theta_{55}$	2.43 (1.28)	2.25 (0.80)

Note: Tabled values are parameter estimates, with standard errors in parentheses.

Table 6

BAU Control Pretest vs. Spring Y2 Maintenance Paired-Samples t-test

	t	df	p-value	Mean Diff	Std. Error Diff	95% Confidence Interval	
						Lower Bound	Upper Bound
BAU	-6.61	17	<0.001	-4.000	0.605	-5.276	-2.724

Note: df = degrees of freedom.

Figures

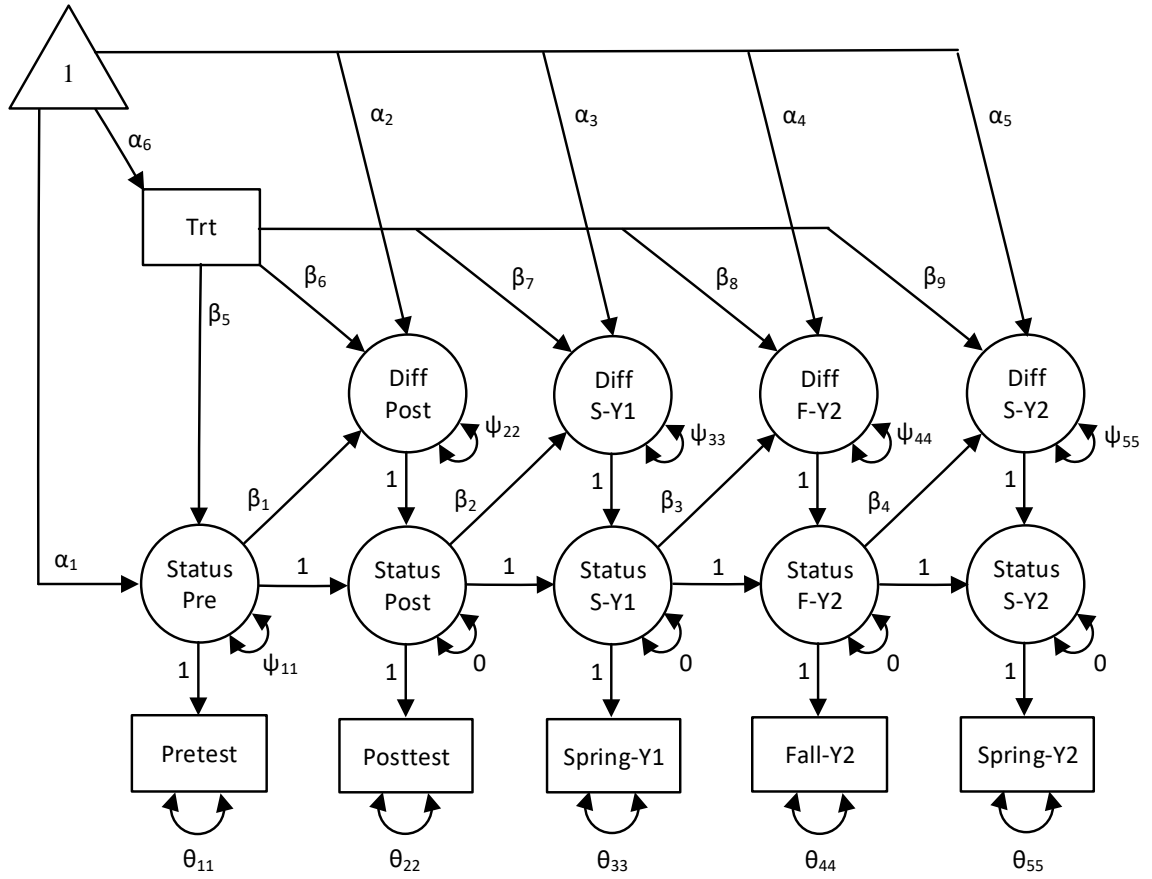


Figure 1. Latent Difference Score Model for Vocabulary at Five Times of Measurement.

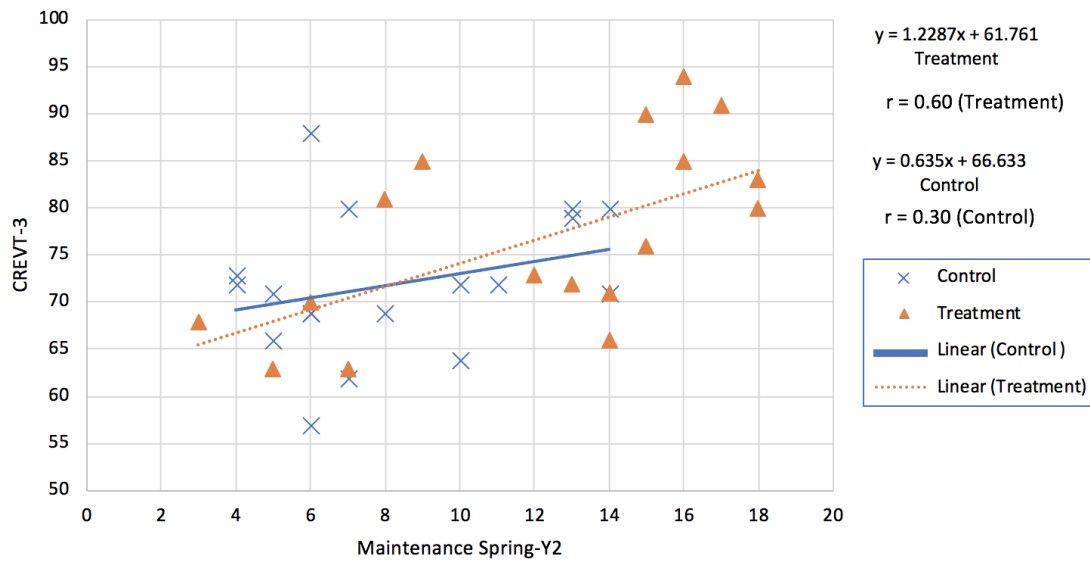


Figure 2. CREVT-3 and Maintenance Spring-Y2 Correlation for Treatment and Control Groups