Title

Systematic CHAAOS: Teaching Vocabulary in English/Language Arts Special Education Classes in Middle School.

Journal

Journal of learning disabilities, 54(3)

ISSN

0022-2194

Authors

O'Connor, Rollanda E
Sanchez, Victoria M
Jones, Brian T
et al.

Publication Date

2021-05-01

DOI

10.1177/0022219420922839

Peer reviewed

Systematic CHAAOS: Teaching Vocabulary in English/Language Arts Special Education Classes in Middle School

Rollanda E. O'Connor, Ph.D. (rollanda.oconnor@ucr.edu),

Victoria M. Sanchez, Ph.D. (victoria.sanchez@ucr.edu), Brian T. Jones, MA (bjone011@ucr.edu), Luisana Suchlit (lsuch001@ucr.edu), Valencia Youkhana (vyoul001@ucr.edu)

University of California at Riverside

Graduate School of Education, 1207 Sproul Hall, Riverside, CA 92521

Kristen D. Beach, Ph.D. (kristen.beach@uncc.edu)

University of North Carolina, Charlotte

Department of Special Education and Child Development

9201 University City Blvd., Charlotte, NC 28223

Keith Widaman (Keith.widaman@ucr.edu)

University of California at Riverside

This research was supported in part by Grant R324A160019 from the U.S. Department of Education, Institute of Education Sciences, to the University of California at Riverside. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Contact Author:

Rollanda E. O'Connor

GSOE, 1207 Sproul Hall

Riverside, CA 92521

Rollanda.oconnor@ucr.edu

951-827-6052

There are Supplemental materials with this manuscript.

Abstract

In this multi-year study, we taught English/Language Arts teachers of students with Learning Disabilities in middle school to incorporate 15 minutes of daily vocabulary activities with students in their intact special education English/Language Arts classes. During Year 1, teachers taught 48 words to their sixth grade students, who learned and retained the words significantly better than students in business-as-usual control classes. In the current study, we report the second year results, as the sixth grade students entered seventh grade. Students (n = 42) in treatment classes again learned 48 new vocabulary words significantly better than similar students in business-as-usual (BAU, n = 21) special education classes. In seventh grade, students also outperformed BAU students on maintenance of these age appropriate words ($p < .001$) and on a standardized measure of vocabulary ($p = .04$).

Keywords: Learning Disabilities, vocabulary, middle school instruction, English Language Learners, intervention, CHAAOS

Many students in middle school fail to read at proficient levels (National Assessment of Educational Progress, 2016), and most students with learning disabilities (LD) read even more poorly (Wei, Blackorby, & Schiller, 2011). Moreover, many students with LD are also English Language Learners (Rodriguez & Rodriguez, 2017), which compounds difficulties reading in English (August, Carlo, Dressler, & Snow, 2005; Lesaux & Harris, 2017). Although these students often master the basic reading skills of decoding words, they tend to read slower than typical learners (Jenkins, Fuchs, & Van den Broek, 2003) and comprehend less of what they read (Lesaux & Harris, 2017; Swanson & Deshler, 2003), which inhibits the amount of reading they do in and out of school, their motivation to read (Wigfield, Eccles, & Rodriguez, 1998), and the opportunity to develop the vocabulary in English that supports comprehension of expository text (Mancilla-Martinez & Lesaux, 2011).

Knowledge of vocabulary is important primarily due to its contribution to reading comprehension (Holahan et al., 2018), and this contribution increases as students progress from elementary into middle school (LaRusso et al., 2016; Mancilla-Martinez & Lesaux, 2011). In his lexical quality hypothesis (LQH) of reading comprehension, Perfetti (2007) places word knowledge (i.e., constituent binding among orthographic and phonological aspects of word recognition with the speed of accessing meanings and usage of words) at the center of reading comprehension. The connection among these aspects of word knowledge forms the lexical quality of the word's

representation in memory. Lexical quality influences comprehension by affecting the ease with which words are recognized and understood in the process of comprehending text.

In keeping with the LQH, Rosenthal and Ehri (2008) suggested that bonding printed word forms to pronunciations and meanings in the lexicon forms an amalgam that makes vocabulary learning more memorable. If students have difficulty recognizing words in print or accessing their meanings, insufficient cognitive processing capacity remains for the integrative tasks of comprehension.

Kintsch (2012) also emphasizes the linkage among aspects of word knowledge in his construction-integration theory of comprehension. In Kintsch's view, understanding meanings of words begins with the bottom-up process of word recognition, which incorporates Rosenthal and Ehri's (2008) amalgam of word features and Perfetti's (2007) binding of word features described in the LQH. Comprehension of text uses this bottom-up word knowledge, including meanings of individual words, interactively with top-down knowledge of context. Kintsch proposed that to form a coherent mental representation of text requires interactions among multiple aspects of language processing in addition to word reading and vocabulary. Thus, none of these authors suggests knowledge of word meanings is sufficient to enable comprehension. However, all stress the importance of word knowledge in the process of reading comprehension.

Particularly problematic is learning the meanings and usage of

academic vocabulary—the types of words that occur commonly in textbooks but rarely in everyday speech. Adolescents with LD and many English Language Learners (ELL) participate readily in conversational language that promotes social interaction, but are less inclined to learn academic vocabulary (Cummins, 2007; Swanson & Deshler, 2003), which requires extensive exposure in relevant contexts (Ebbers & Denton, 2008; Lesaux, Harris, & Sloane, 2012). Deep understanding of words' meanings accumulates across different contexts, which provide the nuances that lead to decontextualized representation of meaning—one a reader can use to make sense of subsequent contexts containing the word (Cromley & Azevedo, 2007; Sternberg, 1987).

As part of a multi-year study, we developed a vocabulary intervention for students who have LD and other disabilities in middle school, and also for students who are ELL with LD: Creating Habits that Accelerate Academic Vocabulary of Students (CHAAOS). We implemented it with sixth grade special education teachers and students in its first year (O'Connor et al., 2019), and with seventh grade teachers and students in this current study to determine the effects of a second year of intervention on proximal and distal measures. In the section that follows, we describe the vocabulary research that led to the design of the CHAAOS intervention.

Vocabulary Intervention

Research on intervention to improve vocabulary has a decades-long history, and has been conducted primarily with students in general education

environments, most often in elementary schools (e.g., Carlo et al., 2004; Loftus & Coyne, 2013; McKeown, Beck, Omanson, & Pople, 1985). Reviews of this research (e.g., Elleman, Lindo, Morphy, & Compton, 2009; Stahl & Fairbanks, 1986) have found positive effects consistently for taught words and sometimes for comprehension of passages containing the taught words, but rarely on standardized measures. From these reviews, recommendations for vocabulary instruction have emerged, including pairing definitions of words with contexts that demonstrate how to use them, and providing multiple opportunities for students to practice and apply new words.

In the 1980s and 1990s, several researchers tested intervention components specifically for students who have LD. In a review of these studies, Jitendra, Edwards, Sacks, and Jacobson (2004) reported positive effects for taught words and again emphasized the importance of providing sufficient practice for learning meanings and applications for new words. Especially relevant for the current study, Bos and Anders (1990) taught vocabulary to 61 students with LD in middle school under various conditions. Students who received highly interactive instruction, in which words were discussed and applied in teacher-directed and peer-to-peer interactions, learned the meanings of words better than those who rehearsed definitions. They also retained meanings better four weeks later. Studies in Jitendra et al.'s review incorporated instructional features found to be effective consistently for students with LD, including teaching small sets of new information, careful sequencing to avoid confusion, corrective feedback, daily

and cumulative practice and review, and active participation (Bulgren, Marquis, Deshler, Lenz, & Schumaker, 2013; Grossen, Caros, & Carnine, 2002; Swanson & Deshler, 2003). Nevertheless, the instruction across studies was short in duration (i.e., one to fifteen sessions), and so generalized vocabulary improvement was not expected.

The notion of practice and application of words in context has become salient through a series of studies and practical advice for teachers by Isabel Beck, Margaret McKeown, and their colleagues (e.g., Beck, McKeown, & Kucan, 2002; McKeown, Beck, & Blake, 2009; McKeown, Beck, Omanson, & Pople, 1985; McKeown, Crosson, Moore, & Beck, 2018), who have made popular the term Tier 2 vocabulary to represent academic words. Although neither designed for students with disabilities nor for ELL, their procedures have been tested for many years in general education environments. As an example, McKeown et al. (2009) developed *Robust Academic Vocabulary Encounters* (RAVE) and shifted their earlier, researcher-delivered instruction (McKeown et al., 1985) into scripted routines that can be used by classroom teachers. Instruction is based on several expository text examples that include the taught words so that students have opportunities to integrate word meanings with a range of appropriate contexts. On posttests, students who received RAVE instruction outscored students in the control group significantly. However, results on a far transfer measure of reading comprehension were not significant, and maintenance of words was not assessed.

In the second year of using this approach (McKeown et al., 2018), students continued to receive RAVE instruction as seventh graders, which allowed the researchers to explore potential long-term effects of their approach. Again, students who received RAVE instruction learned more of the taught words than students in the control group. Moreover, a trend began to emerge suggesting improvement of the RAVE group on a standardized measure of reading comprehension. Students who were ELL were not included in their analyses.

In their seminal study of teaching academic words to students who were ELL, Carlo et al. (2004) used Beck et al.'s (2002) approach by centering instruction on meaningful contexts, cumulative review of taught words, and discussion in groups for 15 weeks of instruction that introduced 12 to 14 new words per week. They extended the work of Beck's research team by teaching mixed groups of students who were ELL or native English speakers (NES) in fifth grade. Their study did not address students with LD; however, their results confirmed the effectiveness of teaching academic words in mixed language groups of students. Although students who were NES achieved higher outcomes overall, students who were ELL in treatment classes outperformed ELL in the control condition significantly.

Two recent groups of studies designed to teach vocabulary in general education classes are especially germane to the current study. The first, *Word Generation,* developed by Snow and colleagues (Lawrence et al., 2015; Snow, Lawrence, & White, 2009), addressed usage of 120 academic words across

subject areas, including English, mathematics, social studies, and science. Like McKeown et al. (2009, 2018), *Word Generation* was implemented by classroom teachers, who attempted to stimulate considerable discussion of words across contexts. School demographics included students who were ELL; however, students receiving special education services were not mentioned. The quality of discussion in *Word Generation* classes exceeded that in control classes, and Lawrence et al. found that classes that spent more time in discussion of academic words had higher outcomes than those with less discussion. Nevertheless, improvement in knowledge of taught words, though significant, was small and no improvement was found on transfer to a standardized measure of vocabulary, a finding that echoed results in studies with younger students (e.g., Loftus & Coyne, 2013).

The second group of studies featured ALIAS (Lesaux et al., 2010; Lesaux, Kieffer, Kelley, & Harris, 2014), which is an ambitious vocabulary approach that was implemented for 45 minutes four times per week during the English/Language Arts block. Notably, Lesaux and colleagues were determined to address the needs of ELL in general education English classes. They designed ALIAS to incorporate extensive discussion as words were introduced and rehearsed, which may be especially beneficial for students who are ELL, who may be less familiar with the words and contexts than NES (August et al., 2005; Carlo et al., 2004). Treatment effects were found for the words taught in ALIAS classes, but not on standardized measures of reading comprehension. A larger study of ALIAS (Lesaux et al., 2014) generated

similar results, with larger effects for students who were ELL and others with lower-than-average vocabularies than for students with higher vocabularies, a finding echoed by Townsend and Collins (2009) in their intervention with middle school ELL.

Common among these studies is the use of discussion to encourage students to process words deeply and use them across contexts and with each other. Indeed, in the studies above and also in research syntheses (e.g., Elleman et al., 2009; Jitendra et al., 2004; Stahl & Fairbanks, 1986), discussion is emphasized as a key instructional condition for improving academic vocabulary. Conversation that stimulates usage of words may be as important as learning definitions because conversing with taught words helps students to understand the pragmatics of particular words and variations of meaning (Applebee, Langer, Nystrand, & Gamoran, 2003; Nagy, Townsend, Lesaux, & Schmit, 2012). Conversation in the process of word learning may be especially important for students with disabilities and students who are ELL, where conversation in the home in English may be less elaborated and less overall than in higher-resourced homes (Hart & Risley, 1992) and vocabulary is less likely to be supported through wide reading of age-appropriate text (Cunningham & Stanovich, 1997), due to poor reading ability in English.

The Current Study

In this multi-year study, we developed a vocabulary intervention based on instructional features found to be effective in earlier vocabulary studies

and also for improving learning of students with LD (Bos & Anders, 1990; Grossen et al., 2002; Jitendra et al., 2004; Pany & Jenkins, 1978; Swanson & Deshler, 2003). CHAAOS differs from approaches we described earlier in being planned specifically for students with LD and other disabilities and for students with disabilities who are ELL, including smaller sets of instructional targets (academic words) than in general education studies and frequent practice with cumulative review. CHAAOS is similar to the general education studies in selection of relevant and grade-appropriate academic words, discussion of familiar and novel examples, and frequent opportunities for students to respond to instructional prompts with their teacher and with peers (Lesaux et al., 2010, 2014; McKeown et al., 1985, 2018; Snow et al., 2009). These instructional features have also been recommended for students who are ELL (August et al., 2005; Carlo et al., 2004; Hall et al., 2017); however, vocabulary studies that focus on students with LD and students with disabilities who may also be ELL are rare. Our study is unique in providing two years of intervention in middle school to students who have LD, many of whom are also ELL.

In Year 1 of this multi-year study, students in CHAAOS classes were taught 48 academic words by their special education teachers. Students in CHAAOS learned and maintained significantly more grade-appropriate academic words than students in BAU classes; however, students in CHAAOS classes did not outperform the BAU on standardized measures of vocabulary or reading (O'Connor et al., 2019). In the work we report here (Year 2), we

follow the original sixth-grade students into seventh grade to add an additional year of vocabulary instruction and determine its effects. To have long-term effects for vocabulary intervention, we reasoned that two conditions should be met. First, we hypothesized that students should demonstrate learning of the taught academic words. Second, students should retain the meanings of these words over time. If students retain what they learn, and many new words are both learned and maintained, they might over time demonstrate improvement in generalized vocabulary.

Our main purpose was to compare our approach to teaching seventh-grade academic vocabulary to learning of academic words in business-as-usual (BAU) special education English/Language Arts classes, which acted as our control condition. Our primary research question was: What are the effects on gains in taught vocabulary of students in CHAAOS special education classes compared to students in BAU classes? We were also interested in maintenance of learned vocabulary, and how gains in vocabulary, if found, might generalize to standardized measures of vocabulary and comprehension.

Method

In this paper, we analyze data from the second year of a multi-year development study of vocabulary intervention for students with LD and other disabilities. We implemented CHAAOS with sixth grade students with disabilities in Year 1, and continued CHAAOS instruction with these students during their seventh grade year, using new academic words in seventh

grade. Special education teachers taught 48 academic words per year, divided into three 4-week sets.

Setting, Teachers, and Students

Setting and assignment to condition. California has the highest percentage of ELL of any state in the United States, at 21% overall and over 30% of the school population in southern California, where this study took place. One school district with approximately 20,000 students hosted this study. The three participating middle schools ranged from 989 to 1280 students each. Free and reduced lunch percentages were 67%, 92%, and 97%, respectively. Across schools, 13% of students received special education services and 27 to 51% received English Language Development services.

For purposes of our study, schools were assigned randomly to condition in Year 1, and each school had one sixth grade special education ELA class and teacher. Teachers and administrators were informed that if they were assigned randomly to the BAU condition, they had the option of being trained to use CHAAOS materials in the following year. A district administrator drew one of three blank envelopes with school names inside to select the BAU school; thus, School 1 was assigned to BAU and Schools 2 and 3 were assigned to CHAAOS.

Teachers. In Year 2, four 7th grade teachers participated in our study (3 female, 1 male). These teachers were credentialed in special education and had taught students with disabilities for 2 to 29 years. Three teachers

were new to the project for the 7th grade school year and taught 7th grade special education English/Language Arts (ELA) classes at their school site. Teachers were assigned to CHAAOS or BAU conditions based on their school's placement to condition in Year 1.

In Year 2 (the current study), Teacher A taught one 7th grade BAU class at School 1, and also a 6th grade CHAAOS class, which was not part of the current study. Teacher B taught one CHAAOS class at School 2; and Teachers C and D each taught one CHAAOS class at School 3. Teacher D had also taught one 6th grade CHAAOS class the previous year.

Students. The 64 participating students were seventh graders (72% male) from the four special education (SpEd) ELA classes within our 3 schools that were participating in CHAAOS research for the 2nd year of the project. All received SpEd services, with designations of Specific Learning Disability (73%), Other Health Impairment (6%), Autism Spectrum Disorder (11%), or Speech/Language Impairment (5%). Thus each participating class served one or two students who were eligible for special education under categories other than LD. Seventy percent of students were ELL with their primary language indicated as Spanish, except for two students whose primary language was Vietnamese. Their reading comprehension measured with the Woodcock-Johnson Tests of Achievement-III (Woodcock, McGrew, & Mather, 2001) averaged over two standard deviations below the test mean. Demographic information for students in the BAU and CHAAOS conditions are provided in Table 1.

Participants Across Years. In Year 1 of the study, 3 sixth grade ELA teachers taught 52 sixth grade students. In Year 2 (i.e., the current study), students advanced from sixth to seventh grade with a new set of teachers, except for Teacher D, who taught both sixth grade CHAAOS in Year 1 and seventh grade CHAAOS in Year 2. Of the original 52 sixth grade students, 15 moved out of the participating schools prior to this Year 2 study. We recruited new participants who moved into Schools 2 and 3 and attended special education ELA in 7th grade to participate in the study because their teachers used the CHAAOS materials.

Teacher Training and Fidelity of Intervention

CHAAOS teachers received 1.5 hours of initial training in August of 2017 to explain the rationale supporting the instructional approach, to demonstrate the first week of lessons, and to provide a copy of the observation tool we would use to collect fidelity data in their classes, as well as to identify areas for coaching and capture strengths and surprises during instruction. This tool (Supplemental Appendix A) documented eight aspects of providing CHAAOS words and definitions, context, and activities, and four aspects of student support. Thereafter, training was individual at each school site.

For the two new 7th grade CHAAOS teachers, researchers modeled all four days of the first week of instruction with their intact classes. The returning teacher asked us to model just the first day to establish expectations and introduce us to the students. As researchers modeled this

instruction, teachers made notes on the observation form to document elements of instruction, questions about the lesson or structure, and fidelity to the guidelines for introducing and practicing words and their definitions, student input on definitions and usage of words, teacher modeling, immediate and corrective feedback, scaffolding, monitoring student performance, and student engagement. In addition to completing the observation forms that noted critical aspects of CHAAOS implementation, teachers asked questions about details of implementation, such as how many student turns they should provide for the questions in the PowerPoint slides, and questions about management of teacher and student materials. Following each observation of researchers teaching their class, we met with teachers to discuss their findings and questions, either immediately after class or later in the day. In this way, training was ongoing throughout the first week of implementation.

From the 2nd through 12th week, instruction was delivered by the special education teachers. When problems were noted during an observation (e.g., students seated far from the teacher seemed less engaged than closer students, or too few students offered responses to questions), researchers offered to teach a portion of the lesson the following day to model a possible approach to improving presentation of CHAAOS (e.g., “I wonder whether moving closer to Manuel and Andrea might perk them up. Could I try tomorrow?”). Teachers accepted these offers and because most of our demonstrations improved the lesson flow, incorporated most of these

suggestions in future lessons. Researchers observed every day of instruction during the teachers' first week of lesson delivery, and thereafter at least weekly in each cycle (i.e., a minimum of 12 observations).

We observed the BAU teacher once per cycle (three times) to document the activities used and vocabulary instruction in ELA and collect data on the instructional features that could be common across conditions. These features included minutes spent on vocabulary, modeling and demonstrating, guided practice, independent practice, scaffolding student responses, corrective feedback, pacing, and student engagement.

CHAAOS Content and Instructional Procedures

Vocabulary words. We selected words for utility in middle school, beginning Coxhead's *Academic Word List* (2000), which contains academic words that occur frequently in textbooks across content areas and was also used as source for word selection by Lesaux et al. (2010) and Lawrence et al. (2015). After eliminating words used in sixth grade CHAAOS classes, we developed a core of 75 words and cross-referenced them with Biemiller's *Words Worth Teaching* (2010), which lists academic words based on when students typically acquire them and how difficult they may be to teach. Next, we tracked the list with words from academic grade level lists developed for the *Common Core State Standards* (CCSSI, 2017). Words that appeared on only one of these three lists were dropped from consideration.

The cross referenced lists were reviewed by the research team, 7th grade ELA teachers in the participating schools in both conditions, and the

ELA district administrative team. Among the 48 words that were selected for CHAAOS instruction, the majority were on all three lists. These words were divided into three sets, taking into consideration words that could be easily confused orthographically or in their meanings. These 16-word sets were taught in three cycles across the school year (See Supplemental Appendix B for the list of words taught during each cycle of instruction).

Key lesson components. We constructed lessons based on research on adolescent learning (Kamil et al., 2008), effective instruction for students with LD (Jitendra et al., 2004; Swanson & Deshler, 2003), and effective intervention for students who are ELL (Carlo et al., 2004; Hall et al., 2017), which includes active and interactive learning, cumulative introduction and maintenance, and ongoing formative assessment. Instructional design included essential components of explicit instruction as outlined by Hughes, Morris, Therrien, and Benson (2017). Specifically, instruction built understanding cumulatively across a series of lessons for each word set, teachers modeled usage of words across contexts and in writing, supports and prompts were faded as students gained skill with each word set, and students received ample opportunities for oral and written practice with feedback.

To ease teacher use of lessons and students' learning, we followed a predictable routine for introducing and contextualizing words across the 3 cycles of instruction, which each lasted four weeks. With breaks between cycles, delivery of CHAAOS spanned September through February. Each

week introduced 4 new words, totaling 16 words for the cycle. This means that four new words were introduced on the Monday with the following three days during the week to practice the four words of the week and review previously taught words. This pattern continued the following week with a new set of 4 words. This weekly predictable 15-minute routine is shown in Supplemental Appendix C and outlined below with an additional set of words:

Monday: Introduction. Introduce the four new words of the week (e.g., *adjust, generate, dispose, superb*) and their synonyms or definitions in adolescent-friendly contexts, using appropriate graphics to demonstrate usage. For example, the teacher introduced the word *adjust* by helping students to read the word and telling students that *adjust means make it work better*. The teacher showed additional synonyms for *adjust,* such as *fix* and *correct*. Next, the teacher showed a picture of a person adjusting guitar strings and reviewed the example for the word *"She adjusted the guitar strings to make it sound better."* Teachers prompted students to think about how the vocabulary word applies in the context (e.g., *What did she need to make work better, or adjust? Why?*). Each of the 4 words was introduced in this manner and teachers and students discussed the meanings of words across the provided contexts and how the words were used in each example. Throughout the discussion in class, students or teachers often provided additional examples of words and new contexts. In the last minutes, students wrote the words and

meanings in their vocabulary notebooks with time to review the words as a whole class or in partners.

Tuesday: *Deep Practice.* Teachers started the routine with a brief review of the four new words (e.g., *adjust, generate, dispose, superb)* that could be completed independently, in partners or small groups, whole class, or a combination of those methods. For example, teachers could review the words and definitions all together with choral responses and then provide students 2 minutes to practice with a partner. Teachers followed up by asking questions like, *"Tell me which word means make work better? What does dispose mean? What word means excellent? What does generate mean?"* The 5 minute review was followed by in depth practice on two of the four words (e.g, *adjust* and *dispose*). Students generated sentences with picture prompts and sentence stems, reinforcing reading words in isolation and context. For example, students were provided with two images: a person adjusting a bike seat and a person adjusting a rearview mirror. The teacher poses the question, "What was *adjusted*? Why?" and students are provided with a frame to help phrase their answer, "He *adjusted* the _____ so he could _____." Students completed a similar activity for the other word of the day. Following individual word practice, student groups practiced additional activities, such as generating a sentence for the two words coupled with a discussion of how to depict the words' meaning through that sentence, reading 3-to-4 sentence

stories and determining which words fit appropriately given the contexts, or reviewing scenarios where the students would discuss and use the vocabulary words throughout. For the *adjust* and *dispose* word study day, the final activity involved reading 3-to-4 short sentences about a *superb* Hollywood Lamborghini where a driver had to *adjust* his parking and *dispose* of his moldy sandwich. Review words in the story from previous weeks included *isolate* and *extend*. During all of these activities, teachers used whole group discussion of confusions over meanings, justifications of usage, and extensions of meanings.

Wednesday: *Deep Practice.* Wednesday is similar to Tuesday's practice except the remaining two words are the focus of the day. Students briefly reviewed all four words (e.g., *adjust, generate, dispose, superb)* followed by more exploration of the remaining two words of the week (e.g, *generate and superb*). Students constructed sentences with picture prompts and sentence stems as they did for Tuesday's practice. For example, students were provided with two images - a lamp switched on and people huddled around a fire. Teachers posed the question "What is *generated*?" and students are provided with a sentence frame to help phrase their answer, "The _____ *generates* _____." Students completed a similar activity for the other word of the day, *superb.* Peer groups practiced in the same types of activities as Tuesday except with the new focus words of the day. For example, one of the scenarios that students reviewed involved

generating money to support a 7th grade Fall Bash dance and students voted in support of the most *superb* idea *generated* by the groups.

Thursday: Review. Each Thursday began with a brief review of the four words of the week, plus any previous words from earlier weeks of the cycle to allow for cumulative practice. Students used flashcards to review the words independently, with partners, or whole-class, as teachers monitored their practice. After review, students worked through small-group tasks in which students justified why or how particular words made sense in given contexts. Some of the activities used during Thursday review included cloze sentences, crosswords, Jeopardy, Bingo, and an Instagram-type activity in which students captioned an image using their vocabulary words. For example, the students were presented with 8 sentences and asked to determine which vocabulary words made sense *(e.g., We used a candle to _____ light when the power was out; That was the most ____ ride I have ever gone on)*. During whole class review of responses, students provided their answer and justified why that word was selected; then the class would agree or disagree with the response.

This 4-day pattern represents one week of instruction, which was repeated 4 times in order to represent one complete cycle of instruction. Our students completed 3 cycles of instruction to learn the 48 novel words. The 3 cycles of teacher presentation PowerPoint slides and student materials are available on the first author's website

(https://gsoe.education.ucr.edu/CHAAOS/index.php) and Cycle 1 is available as Supplemental Appendix C.

CHAAOS was designed to take about 15 minutes of the 55 minutes designated for ELA in this district. During the non-CHAAOS instructional time, two CHAAOS teachers (Teachers C and D) implemented *Corrective Reading, Decoding B* (Engelmann et al., 1988) using whole class instruction, as described in the BAU instruction that follows. Teacher B used a variety of packaged materials that comprised locating and repairing errors in sentences and writing tasks during the non-CHAAOS time, along with *Newsela* (see below) on Fridays.

Business as Usual (BAU) Instruction

We observed ELA instruction in the special education BAU class once per instructional cycle (i.e., 3 times). The special education BAU class included students who had been in the BAU condition the previous year. Three other BAU students from Year 1 were placed in two additional classes for ELA, and given the small number of students (i.e., one and two students, respectively) these classes were not observed.

The teacher in the BAU used two reading programs in his class. Four days per week, he implemented *Corrective Reading, Decoding B* (Engelmann et al., 1988) using whole class instruction. Lessons begin with decoding new words in the day's reading, as well as review of previously taught words, which takes three to four minutes. Many lessons include one or two new vocabulary words found in the day's reading. The majority of the class time

is spent reading aloud the day's story individually or chorally, which has controlled text to practice decoding patterns introduced sequentially. Teachers interrupt student reading at prescribed times to ask scripted comprehension questions and discuss answers when varied responses are appropriate.

On Fridays the teacher used *Newsela* or *Achieve 3000*, which are web-based reading programs with adjustable reading levels implemented with students through Chrome Books. These materials include new vocabulary as incorporated in the reading content, but they do not specifically address academic vocabulary.

Thus, while most classes used *Corrective Reading* as their base program, BAU classes used this material for about 55 minutes daily, whereas CHAAOS classes used it for about 37 minutes daily (i.e., 15 minutes of CHAAOS, 37 minutes of Corrective Reading, and about 3 minutes of transition) during the 12 weeks of CHAAOS implementation. Observations of teaching behaviors are reported in the results section.

Measures

We documented fidelity of implementation with an Observation Tool that teachers used when researchers modeled instruction in their classes and researchers used when observing teacher implementation. We assessed students' proximal receptive understanding of word meanings through multiple choice measures and their near transfer of target words to untaught contexts with cloze passages and comprehension questions following

paragraphs containing taught words. We included norm-referenced measures of vocabulary and reading comprehension as far transfer measures, primarily to describe our sample. Initial scoring was conducted by graduate student researchers soon after test administration. We determined reliability by double scoring 25% of the tests without the raters knowing whether the tests were collected as pre- or posttests, and whether students were in CHAAOS or BAU conditions. Scoring reliability was 97.9%.

Observation Tool

The Observation Tool (Supplemental Appendix A) documented eight aspects of teaching vocabulary words and definitions, context, and activities, and four aspects of student support (e.g., scaffolding, feedback, pacing, and motivation). Each aspect of instruction was followed by a narrow column to rate the feature from 1 to 3. Inter-rater reliability was established between two raters at 92% percent agreement on eight observations. We also left sufficient space for observers to record specific instances of teacher and student language exchanges that occurred during the observation as qualitative data to describe instruction and interactions.

Proximal Measures

For each of the three 16-word cycles, we developed a direct measure of vocabulary learning, which was a 16-item multiple choice test of the taught words. The four choices included one correct answer, one morphologically related incorrect choice, one orthographically similar incorrect choice, and one unrelated incorrect choice. For example, the

choices for *accumulate* were: (a) pile up [correct], (b) dust and dirt [morphologically related incorrect choice], (c) vacuum [orthographically similar incorrect choice], and (d) place to buy things [unrelated incorrect choice]. Test-retest reliability ranged from .88 to .96 across cycles. Correlation coefficients between the multiple choice measure and standardized measures of vocabulary ranged from .34-.52.

Near Transfer Measures

We also developed near transfer measures of vocabulary usage in contexts that were not taught, which included cloze passages and reading comprehension of passages containing vocabulary words from that cycle. These measures were administered before and after each instructional cycle. Test-retest reliability ranged from .88 to .96 across cycles. Estimates of validity were obtained for the usage measures with standardized tests of vocabulary (r = .46-.55) and passage comprehension (r = .30-.47).

Maintenance of Taught Vocabulary Words

The three cycles of instruction spanned September through February, and we were interested in whether treated students maintained knowledge of the 7th grade words they had learned at the end of the school year. To test maintenance of vocabulary learning, we developed an 18-word vocabulary measure that drew six words from each of the three cycles. We selected the six words on which students scored lowest at pretest and highest on the immediate posttest, thus indicating words on which students showed the strongest growth in word knowledge for that cycle. We used this measure at

the end of the school year.

Far Transfer Measures

Standardized measures of reading and vocabulary were administered at the beginning and end of Grade 6, which was the first year of the study, and the end of Grade 7, which is the focus of the current study, to describe our sample and serve as far transfer measures. Furthermore, we were interested in whether any change occurred on standardized measures over the course of 2 years of intervention.

The Comprehensive Receptive and Expressive Vocabulary Test, 3rd ed. (CREVT; Wallace & Hammill, 2013) was administered individually and includes two subtests. Relational vocabulary requires students to point to a picture that represents what the examiner says. Expressive vocabulary is measured by asking students to define a word. Definitions were scored as correct (1) or incorrect (0). As required in the manual, examiners queried all incorrect or vague responses one time to allow the students a second opportunity to pass the item. Internal consistency across subtests was high, with coefficient alphas of .85-.96. The general vocabulary score, which combines results from the receptive and expressive scores, is reported here. None of the words on this measure were taught in CHAAOS lessons.

The reading comprehension portion of the *Woodcock-Johnson Tests of Achievement III* (WJ-III; Woodcock et al., 2001) was administered individually to students to describe reading ability and confirm that participants had severe reading difficulties. Across subtests, reliabilities ranged from .81-.94.

Results

Data Analysis

First, we evaluated teachers' fidelity of implementation. Next, we addressed potential group differences through MANOVA on standardized measures. Then we conducted a series of ANCOVA for each of the three cycles to test receptive vocabulary learning on the multiple choice items and to test student's ability to use vocabulary words in context. We assessed maintenance of learned words with structural equation modeling (pretest, posttest, follow-up). After this, we estimated growth in generalized vocabulary using repeated measures on the CREVT, which is a standardized measure of vocabulary, and growth in comprehension using repeated measures on the Passage Comprehension subtest of the WJ-III to determine whether scores changed over time with two years of participation in CHAAOS.

Classroom Teachers' Fidelity of Implementation

On average, CHAAOS teachers spent 23 minutes teaching the lessons in Cycle 1. Teachers gradually became more efficient, and Cycle 1 took longer to implement daily than Cycle 3, where teachers reduced the time to an average of 15 minutes daily, as the researchers did during the first week of modeling lessons. Overall, teachers scored 30.83 out of 36 points for fidelity of implementation (86%) across the cycles. Teachers had strong implementation for requiring frequent responses from students, clear modeling of how to use the words in sentences, and providing immediate

corrective feedback to students during the lesson.

Our primary emphasis was on quality of implementation and fine-tuning of lesson components, and so most observations were in treatment classes. Due to time constraints, we observed the BAU teacher formally only three times, once during each of the three cycles. In this class, pacing was brisk, student participation was high, and each student responded to questions and prompts individually and through group responding. The BAU teacher also implemented CHAAOS instruction with a class of sixth grade students, which was also observed though not part of the current study. His pacing, prompting, and scaffolding were similar across CHAAOS and BAU conditions.

Although the focus of the primary curriculum in the BAU, *Corrective Reading,* is on decoding words and building fluency, the teacher addressed vocabulary informally through word morphemes (e.g., vision, visible, invisible) and direct teaching of the meanings of one to three words in each observation (e.g., chunk, offering, clinker, reflection). Although he had received the list of CHAAOS words at the beginning of the school year, he did not have access to the instructional materials for seventh grade and we did not observe him using any of the words instructionally. The average time spent on vocabulary instruction in the BAU class was just under four minutes per observation.

We had not expected teachers to generate additional examples for the CHAAOS words; however, on our observation tool we documented many

instances of teachers proposing tailored examples or extending their students' use of words. As examples, they offered usage of words related to events in the class or school, which may have given students concrete illustrations familiar to particular groups of students. A teacher used an upcoming fieldtrip as context for *consent*. Another showed students their tasks over the course of the week and asked, "Which activity will be your *priority*?" One teacher consistently repeated or rephrased a student's comment to shape correct usage and provide reinforcement, and another repeated a student's response as validation that it was correct. Two of the teachers were exceptionally affirmative in their feedback ("Outstanding! Exactly! Well done!") and scaffolded students' appropriate responses by reasking a question at a lower level to prompt a correct response they then reinforced. We return to data from these observations in the discussion.

Pretest Equivalence

MANOVA on pretests of the WJ-III and CREVT across the teachers' classes did not differ [Wilks' Lambda = .947, $F(5,89) = 0.485$, $p = .818$], suggesting that students in these classes were initially comparable on the standardized reading and vocabulary tests. Moreover, these pretest scores did not differ by treatment or BAU conditions [Wilks' Lambda = .995, $F(2,57) = .084$, $p = .969$] or on the Cycle 1 pretests of vocabulary (multiple choice and usage measures) [Wilks' Lambda = .987, $F(2,57) = .351$, $p = .705$] (See Table 2).

Proximal: Receptive Learning of CHAAOS Words

For the multiple-choice (MC) test in Cycle 1, ANCOVA with posttests covaried by pretests indicated a significant main effect of treatment [F (1, 58) = 61.282, p = <.001, η2 = .57] and a significant effect of the pretest covariate on the posttest [F (1, 58) = 5.07, p = .03, η2 = .10]. Results were similar on the MC scores in Cycles 2 and 3, with main effects for treatment and for pretest (summary of results provided in Table 2). Partial Eta Squared values for treatment were .27 and .21 in Cycles 2 and 3, respectively.

Near Transfer: Usage of Words

On tests of usage of vocabulary, results also significantly favored students in the CHAAOS treatment. ANCOVA of Cycle 1 usage with pretest as covariate indicated significant effects for treatment [F (1, 58) = 6.785, p = .011, η2 = .081], but not for usage pretest [F (1, 58) = 1.219, p = .273]. In Cycles 2 and 3, significant effects for treatment were also found (p < .001 and .005, respectively, η2 = .18 and .093). Note that neither the cloze passages nor the comprehension paragraphs on this measure were used instructionally. These statistics are shown in Table 2.

Maintenance of Taught Vocabulary Words

We used structural equation modeling to analyze the treatment effects on maintenance of CHAAOS vocabulary, with full information maximum likelihood (FIML) estimation due to the presence of small amounts of missing data. To evaluate model fit, we used the chi-square index of model fit and three additional measures of practical fit. Of these latter measures, the root mean square error of approximation (RMSEA) is an index of absolute fit, with

values below .05 indicating close fit to data, and values between .05 and .08 indicating adequate fit. The final two measures were the comparative fit index (CFI) and Tucker-Lewis index (TLI), both of which imply close fit of a model to data if values above .95 are obtained.

We fit an adapted version of a latent growth model, shown in Figure 1, to the data. In this model, the Intercept latent variable had fixed loadings of 1.0 to the pretest, posttest, and maintenance measures. The Slope latent variable had fixed loadings of 0 to the pretest, 1.0 to the posttest, and a freely estimated loading (λ_3) to the maintenance measure. Thus, in this model, the Intercept represented performance at the first time of measurement (pretest), the Slope reflected improved performance to the second measurement (posttest), and the estimated loading (λ_3) for the third measurement (maintenance) provided an estimate of retention of the treatment effect. The closer λ_3 is to 1.0, the more complete the retention of treatment effects.

In the model in Figure 1, the Treatment variable was coded 0 = BAU, 1 = treatment, so α_1 provides an estimate of mean performance of BAU participants at pretest, and β_1 is an estimate of the difference in mean performance of the treatment group compared with the BAU group at pretest. Then, α_2 yields an estimate of the increase in mean performance of BAU participants from pretest to posttest, and β_2 is an estimate of the differential increase in mean performance of the treatment group relative to any increase in mean performance by the BAU group between pretest and

posttest. The parameters θ_{11} through θ_{33} are the measurement residuals for the manifest variable scores from pretest through maintenance, and these estimates were constrained to equality, embodying a homogeneity of variance assumption, to identify the model.

The fit of Model 1 was very good, as shown in Table 3. The statistical index of fit indicated that perfect fit of the model was not rejected, χ^2 (4) = 5.70, p = .22. The RMSEA of .079 was adequate. Finally, the CFI and TLI values easily surpassed values of .95, also indicating close fit of the model to the data. The parameter estimates indicate that the BAU condition scored just over 4.5 words correctly defined at pretest, α_1 = 4.61 (*SE* = 0.71), and that the treatment group exhibited mean performance approximately 1 word correct higher, β_1 = 1.11, although this group difference in mean performance was not significant, p = .17. The BAU condition showed a significant increase in performance of almost 3 more terms correctly defined at posttest, α_3 = 2.74 (*SE* = 1.19), p = .02; however, the treatment group exhibited improved performance of approximately 5.5 additional terms correctly defined relative to the increase by BAU participants, β_2 = 5.49, p < .001. Thus, the treatment led to significant increases in performance relative to that of participants in the BAU condition.

One interesting finding from was that the proportional retention parameter, λ_3, was estimated at 0.96 (*SE* = .05), suggesting that the treatment condition lost only 4 percent of its differential gains from posttest to maintenance, a loss that was non-significant, with CI = [0.87, 1.05]. Thus,

the treatment condition showed substantial differential improvement from pretest to posttest and then retained essentially all of that improved performance on the maintenance test at the end of the year.

Because the proportional retention parameter was estimated so close to 1.0, we formulated Model 2, which involved simply fixing λ_3 at 1.0. Model 2 led to a non-significant change in model fit, $\Delta\chi^2$ (1) = 0.72, and all practical fit indices were improved, with RMSEA = .064 and both CFI and TLI over .980. As seen in Table 3, all other parameter estimates were essentially unchanged. Thus, theoretical implications of Models 1 and 2 are essentially identical, with Model 2 providing a more efficient representation of the treatment effects.

Far Transfer: Reading and Vocabulary Performance on Norm-Referenced Measures

Scores on CREVT and WJ-III were collected at the beginning and end of Grade 6 and end of Grade 7. Planned comparisons on CREVT scores revealed that there were no differences in student performance at the beginning or end of Grade 6 (F (1,60) = 0.52, p = .48, see Table 2); however, CREVT scores differed at the end of Grade 7 (F (1,60) = 4.44, p = .04, with students in CHAAOS classes scoring significantly higher than those in the BAU classes. Repeated measures ANOVA on WJ-III comprehension scores revealed a similar pattern, with no difference between students in CHAAOS or BAU classes on the pretest or posttest in Grade 6 (F (1,60) = .123 and p = .810, respectively); however, scores at the end of Grade 7 showed a positive,

though not significant trend (F (1,60) *3.34*, p = .07, eta squared = 0.6, see Table 2). Note that the standard scores on the WJ-III Passage Comprehension subtest remained over two standard deviations below the test mean.

Discussion

In this second year of a multi-year study, students receiving special education in 7th grade were taught 48 academic words across three instructional cycles that spanned September through February. They learned significantly more words than students in BAU classes and maintained those words' meanings three months later with no significant drop in scores. Students who participated in both years of CHAAOS instruction (i.e., over half of the students in CHAAOS classes) learned the meaning and usage of up to 96 important academic words. Our premise was that if students could learn key academic words and retain their meanings, their generalized vocabulary might also grow. Indeed, that is what we found, even though the specific words targeted for instruction were not items on the standardized vocabulary measure.

Connection with Prior Research

Several features of CHAAOS instructional design mirror the design of earlier vocabulary studies, including extensive verbal interactions between teachers and students, and among students (e.g., Beck et al., 2002; Bos & Anders, 1990; Carlo et al., 2004; Lesaux et al., 2012). Although Lawrence et al. (2015) found only small effects for Word Generation overall, higher effects

were found in classes where higher proportions of time were spent discussing words. Using words in a range of relevant contexts is another feature common to effective vocabulary interventions.

Carlo et al.'s (2004) study did not include students who had disabilities; however, as in their study, CHAAOS instruction was conducted in mixed classes of students who were ELL or NES, and small group discussions occurred routinely among students whose first language differed. The instructional graphics that used contexts familiar to the students may have enabled students regardless of first language to discuss the words and meanings among themselves. We often heard students who were ELL using some Spanish during small group discussions, but rarely heard instances of students using Spanish cognates for CHAAOS words, as in Carlo et al.'s study. Given that students who were ELL in our study also had disabilities, they may have had difficulty with academic language in Spanish as well as in English.

Although CHAAOS shares similarities with prior studies of vocabulary intervention, it differed in several relevant aspects. First, earlier studies (Carlo et al., 2004; Lawrence et al., 2015; Lesaux et al., 2012; McKeown et al., 2018) introduced ten or more new academic words per week. By contrast, CHAAOS introduced only four new words per week, in keeping with recommendations for students with LD to teach fewer targets with more repetition and cumulative review (Bos & Anders, 1990; Jitendra et al., 2004; Swanson & Deshler, 2003). To our knowledge, CHAAOS is also unique in

being a vocabulary intervention conducted entirely in middle school special education classes that included high proportions of students who are ELL in addition to having disabilities.

Jitendra et al.'s review also found most studies in special education were of very short duration (i.e., 1 to 15 weeks), whereas we provided two years of vocabulary instruction, although only 54% of our sample participated in both years. Moreover, when maintenance has been assessed (e.g., Jitendra et al.), it is usually one-to-four weeks following instruction; however, we assessed maintenance of taught words three months after the close of CHAAOS instruction without a significant drop in knowledge.

As in the vocabulary studies reviewed by Stahl and Fairbanks (1986), Jitendra et al. (2004), and Elleman et al. (2009), we found positive effects on taught words. Some of the studies in these reviews found transfer to comprehension of passages that contained the taught words and we also found positive effects for transfer to untaught contexts in all three instructional cycles. The positive effect on our far transfer measures of vocabulary ((i.e., the CREVT) in seventh grade, but not in sixth grade, is more puzzling because effects on standardized measures of vocabulary are rare.

McKeown et al. (2018) reported a similar finding in general education classes with transfer to a standardized measure after two years of RAVE intervention. However, just over half of students in the CHAAOS intervention classes in Grade 7 received two years of intervention. Nearly half the

students received CHAAOS only in seventh grade, and thus we cannot link far transfer directly to a second year of intervention. Nor can we attribute improvement on the CREVT to improvements in CHAAOS materials or instruction in seventh grade. Although we revised and strengthened the materials based on class observations after each year of implementation, those improvements were made after implementing CHAAOS in the current study, moreover, teachers' fidelity of implementation was similar in Grades 6 and 7.

One possibility is that CHAAOS created habits of exploring words, as the intervention intended, that took time to establish. Students who received CHAAOS for both years may have become gradually more eager to play with meanings of unknown or partially known words, much as typical readers build partial to full representations of meanings as they encounter words in texts and conversations. Returning CHAAOS students in seventh grade may have helped to initiate new students in how to partake in the lively activities that formed the CHAAOS routines. If so, students may have responded more willingly to prompts for more information on the expressive portion of the CREVT. A larger study with a more stable group of participants would be needed to enable analyses of cumulative years of intervention and whether students' enthusiasm and skill with wordplay improved alongside their knowledge of core academic words.

Connection with Theory

The lexical quality hypothesis (Perfetti, 2007) stresses the importance

of linking decoding of words with immediate access to words' meanings. Although explicit decoding instruction was not part of the CHAAOS vocabulary routine, teachers gave decoding prompts for the words, especially on Mondays when the words for the week were introduced. Teachers pointed out affixes the students were likely to know already, or places to chunk long words into decodable pieces. One teacher also picked out the two longest or most difficult to read words for the week (e.g., *apprehensive* and *negotiate*) and modeled how to use word parts the students could recognize to break apart and put the word back together.

Throughout CHAAOS instruction, we helped students to bond the printed word form (orthography) with its pronunciation and meaning through pictured and captioned contexts that displayed the word as well as demonstrated use of the word and formed the backdrop for discussions in which students used the word in relevant contexts. Rosenthal and Ehri (2008) suggested that bonding the printed word form with pronunciation and usage may be necessary for memorable vocabulary learning. These activities that consistently connected printed words with pictured and verbal meanings all may have improved the quality of the word's lexical representation in memory. With eased access to a clear lexical representation, students may be more likely to form a coherent situated representation of a context that incorporates the word, as Kintsch (2012) suggested. Although we do not have a direct test of either theory, students in CHAAOS classes were able to read and understand the meanings of

passages that contained the taught words better than students in BAU classes, which suggests they may have been better able to form a coherent representation of what they had read.

Teaching CHAAOS

Teachers in both conditions were credentialed special education teachers who were experienced in teaching English/Language Arts to students with high-incidence disabilities, as well as to students who are ELL. Teachers in both conditions scored high on the Observation Tool items that document effective features of special education instruction (Swanson & Deshler, 2003), instruction for adolescents (Kamil et al., 2008), and instruction for students who are ELL (August et al., 2005; Carlo et al., 2004), such as scaffolding, feedback, brisk pacing, and student motivation.

For CHAAOS teachers, instructional expectations were outlined clearly in the Observation Tool the teachers used as they observed the researchers teaching their intact classes during the first week of instruction. This scale addressed not just whether aspects of the lessons were implemented, but also the degree of student participation, number of turns given to students, review of previously taught words, and description of scaffolding students' usage of words. As teachers observed the degree of interaction between the researcher and their students during CHAAOS modeling in Week 1, some were surprised by high levels of engagement shown by students with low levels of language and low levels of English, in particular. As teachers took over all instruction, we noted in our observations teachers using techniques

we had used during modeling to elicit participation, such as, “Raise your hand when you know a word that means…” or “thumbs up when you know…” and waited for many hands to go up before calling on students. In CHAAOS classes, teachers allowed students to call out definitions for words, or words to fit definitions without raising hands, which stimulated more student responding than when individuals were called upon, as was common in the BAU class.

Teachers offered multiple turns for students to provide examples of usage; sometimes as many as seven students responded with appropriate examples for each word. A teacher modeled generating a sentence with “I could *vary* my breakfast by alternating between cold and hot foods. How might you *vary* your breakfast?” [teacher calls on five students to give examples] “Now write a sentence with how you *vary* your breakfast from day to day.” Most teachers also encouraged use of multiple forms of the word (e.g., restrict, restricted, restriction). One teacher shared a book with her students on visual *perception*, and asked them to take notes on what they *perceived*. These behaviors may have stimulated the exceptional level of participation in oral usage of words, which may have contributed to firmer lexical representations of words bonded to meanings.

Limitations

All classes and schools were located in a single district, and findings in these schools might not generalize to students in other locations. The nature of studies that develop and test new curricular approaches often

necessitates small sample sizes, as in the current study, and so our sample was limited to the students who received special education services in three middle schools. The small sample size warrants several cautions in interpretation of our results. Because a majority of students in these classes were ELL, our study included too few students with LD who were NES to test for differences by ELL status. Although small sample sizes can limit the ability to find differences, that was not the case here. The small sample could miss finding real differences between the vocabulary posttest and maintenance tests; however, examination of scores in Table 2 reinforces our conclusion that students retained knowledge of words they learned during CHAAOS instruction.

Although we had hoped to determine effects of a second year of CHAAOS intervention, student mobility in this district prohibited doing so. By the end of the second year of intervention, we had lost 15 students from the first year sample and gained 27 seventh graders who had not participated in the sixth grade intervention or BAU classes. Replication with a larger and more stable student sample might be able to address the cumulative effect of CHAAOS in future studies.

The BAU teacher taught only in the BAU condition in Year 1 when students were in sixth grade, and continued with BAU students in seventh grade, but also taught CHAAOS to sixth grade students in Year 2, who were not part of this study. As could be expected, qualities of scaffolding, feedback, pacing, and motivation were similar across conditions because

these qualities reflected his overall style of teaching. The differences across conditions were related to the amount of time spent on vocabulary activities (i.e., two-to-four minutes in the BAU class compared to 15-to-18 minutes when teaching CHAAOS lessons), and increased reading time in the BAU class because reading supplanted CHAAOS time. We observed no use of CHAAOS activities in his BAU class; however, only three lessons were observed in the BAU condition.

Core components of CHAAOS, including student friendly definitions, multiple contextual representations, multiple opportunities for students to respond, phonetic and orthographic representations of words, and support for using words in speaking and writing were specified in the instructional package. Nevertheless, the details of implementation (e.g., 4 words per week, 48 words per year, twelve weeks of instruction spread over six months of the school year) were kept consistent deliberately. As a result, we cannot determine whether 4 words per week is optimal pacing for students with disabilities in middle school to learn grade-level words. Nor can we determine through the current study whether twelve weeks was needed or more weeks would be useful, or whether growth would continue if the instruction continued another year with another 48 academic words. These limitations could be addressed through further research.

Although we are pleased with results that suggest students with disabilities are highly responsive to instruction geared to teach age-appropriate academic language, we have no evidence that improvements

will contribute meaningfully to academic success in high school, which is often (as in this study) the purpose for trying to improve students' understanding and use of academic language. Longer-term studies are needed to determine whether small improvements in knowledge of academic words impact long-term outcomes.

Conclusions and Implications for Instruction

Lack of opportunity to learn academic vocabulary is egregious particularly because vocabulary assumes an increasingly important role in reading comprehension from Grade 5 through Grade 9 (Holahan et al., 2018). When we introduced these grade-appropriate CHAAOS words to teachers and principals in both conditions, teachers expressed skepticism that their students would be able to learn them, worried because their students read several years below grade level. A common theme among teachers was, "They can't even read these words, let alone understand them." Indeed, many of their students could not read them at pretest; however, we observed that most students read them correctly with practice in the CHAAOS classes. Another common theme among teachers was to suggest we teach easier words, more in keeping with students' current level of vocabulary. However, our intent was to help students learn important academic words that could enable participation in middle school courses and beyond.

Resources are available on how to design vocabulary instruction that focuses on useful academic words and student-friendly contexts (e.g., Beck

et al., 2002); however, it is unusual for teachers to have the time available to develop instructional materials on their own. Teachers are often advised to incorporate opportunities for students to play with words and their meanings and to initiate potential uses for interesting words (Beck et al.; Carlo et al., 2002; McKeown et al., 2018); however, doing so takes time and skill to develop such a range of examples. Furthermore, most students with disabilities and students who are ELL need extensive practice with words in order to retain their meanings over time. We are unaware of any materials that have been developed for special educators in middle school that embody these important features, which led to our development of CHAAOS.

This study offers a test of effects of vocabulary instruction that can be layered over existing curricula in middle school. Moreover, these materials are available at no cost to download from a website that includes teacher instructions, presentation slides, and student materials. Teachers of CHAAOS classes spent more time on vocabulary instruction than did teachers in BAU classes; however, the time spent may be worthwhile because student maintenance measures revealed no significant drop in word knowledge during the months following instruction. Focusing a fraction of the available special education ELA class time on this often neglected area could lead to long-term retention of taught words with the potential to improve students' generalized vocabulary and comprehension.

References

Applebee, A.N., Languer, J.A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*, 685-730. doi: 10.3102/00028312040003685

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice, 20*, 50-57. doi: 10.1111/j.1540-5826.2005.00120.x

Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

Biemiller, A. (2010). *Words Worth Teaching*. Columbus, OH: SRA/McGraw-Hill.

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:10.2307/1510390

Bulgren, J.A., Marquis, J.G., Deshler, D.D., Lenz, B., & Schumaker, J.B. (2013). The use and effectiveness of a question exploration routine in secondary English language arts classrooms. *Learning Disabilities Research & Practice, 28*, 156-169. doi: 10.1111/ldrp.12018

Carlo, M., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D.N.,

Lively, T.J., & White, C.E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*, 188-215. doi: 10.1598/RRQ.39.23

Common Core State Standards Initiative (CCSSI, 2017). *California Common Core State Standards*. Retrieved July 18, 2018, from https://www.cde.ca.gov/be/st/ss/documents/finalelaccssstandards.pdf

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. doi:10.2307/3587951

Cromley, J.G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311-325. doi: 10.1037/0022-0663.99.2.311

Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics, 10*, 221-240.

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

Ebbers, S.M., & Denton, C.A. (2008). A root awakening: Vocabulary instruction for older students with reading difficulties. *Learning Disabilities Research & Practice, 23*, 90-102. doi: 10.1111/j.1540-5826.2008.00267.x

Elleman, A.M., Lindo, E.J., Morphy, P., & Compton, D.L. (2009). The impact of

vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on educational Effectiveness, 2*, 1-44. doi: 10.1080/19345740802539200

Engelmann, S., Johnson, G., Carnine, L., Meyer, L., Becker, W., & Eisele, J. (1988). *Corrective Reading Decoding Strategies*. Chicago: SRA, MacMillan/McGraw-Hill.

Grossen, B., Caros, J., & Carnine, D. (2002). Big ideas plus a little effort produce big results. *Teaching Exceptional Children, 34*, 70-73.

Hall, C., Roberts, G.J., Cho, E., McCulley, L.V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review, 29*, 763-794. doi: 10.1007/s10648-016-9372-4

Hart, B., & Risley, T.R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28*, 1096-1105. doi: 10.1037/0012-1649.28.6.1096

Holahan, J.M., Ferrer, E., Shaywitz, B.A., Rock, D.A., Kirsch, I.S., Yamamoto, K., Michaels, R., Marchione, K.E., & Shaywitz, S.E. (2018). Growth in reading comprehension and verbal ability from grades 1 through 9. *Journal of Psychoeducational Assessment, 36*, 307-321. doi: 10.1177/0734282916680984

Hughes, C.A., Morris, J.R., Therrien, W.J., & Benson, S.K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities*

Research & Practice, 32, 140-148. doi: 10.1111/ldrp.12137

Jenkins, J.R., Fuchs, L.S., & Van den Broek, P. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719-729. doi: 10.1037/0022-0663.95.4.719

Jitendra, A.K., Edwards, L.L., sacks, G., & Jacobsoon, L.A. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children, 70*, 299-323. doi:10.1177/001440290407000303

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., and Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A Prac- tice Guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evalu- ation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc

Kintsch, W. (2012). Psychological models for reading comprehension and their implications for assessment. In J.P. Sabatini, E.R. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pps. 21-38). New York, NY: Rowman & Littlefield.

LaRusso, M.D., Donovan, S., & Snow, S. (2016). Implementation challenges for Tier One and Tier Two school-based programs for early adolescents. *New Directions for Child and Adolescent Development, 154*, 11-30.

Lawrence, J.F., Crosson, A.C., Pare-Blagoev, E.J., & Snow, C.E. (2015). Word generation randomized trial: Discussion mediates the impact of

program treatment on academic word learning. *American Educational Research Journal, 52*, 750-786. doi: 10.3102/0002830215579485

Lesaux, N.K. & Harris, J.R. (2017). An investigation of comprehension processes among adolescent English learners with reading difficulties. *Topics in Language Disorders, 37*, 182-203. doi:10.1097/TLD.0000000000000120

Lesaux, N.K., Harris, J.R., & Sloane, P. (2012). Adolescents' motivation in the context of an academic vocabulary intervention in middle school classrooms. *Journal of Adolescent and Adult Literacy, 56*, 231-240. doi: 10.1002/JAAL.00132

Lesaux, N., Kieffer, M.J., Faller, S.E., & Kelley, J.G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*, 196-228. doi.org/10.1598/RRQ.45.2.3

Lesaux, N., Kieffer, M.J., Kelley, J.G., & Harris, J.R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal, 51*, 1159-1194. doi: 10.3102/0002831214532165

Loftus, S.M., & Coyne, M.D. (2013). Vocabulary instruction with a multi-tier approach. *Reading & Writing Quarterly, 29*, 4-19. doi:10.1080/10573569.2013.741942

Mancilla-Martinez, J., & Lesaux, N.K. (2011). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study.

Child Development, 82, 1544-1560. doi: 10.1111/j.1467-8624.2011.01633.x

McKeown, M.G., Beck, I.L., & Blake, R. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44,* 218-253. doi: 10.1598/RRQ.44.3.1

McKeown, M.G., Beck, I.L., Omanson, R.C., & Pople, M.T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20,* 522-535. doi.org/10.2307/747940

McKeown, M.G., Crosson, A.C., Moore, D.W., & Beck, I.L. (2018). Word knowledge and comprehension effects of an academic vocabulary intervention for middle school students. *American Educational Research Journal, 55*, 572-616. doi: 10.3102/0002831217744181

Nagy, W., Townsend, D., Lesaux, N., & Schmidt, N. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47,* 91-108. doi: 0.1002/RRQ.011

O'Connor, R.E., Beach, K.D., Sanchez, V.M., Knight-Teague, K., Orozco, G., & Jones, B. (2019). Teaching academic vocabulary to sixth grade students with disabilities. *Learning Disability Quarterly*, doi.org/10.1177/0731948718821091

Pany, D., & Jenkins, J.R. (1978). Learning word meanings: A comparison of instructional procedures*. Learning Disability Quarterly, 5*, 202-215.

Perfetti, C.A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383. doi: 10.1080/10888330701530730

Rodriguez, A., & Rodriguez, D. (2017). English learners with learning disabilities: What is the current state? *Insights on Learning Disabilities, 14*, 97-112.

Rosenthal, J., & Ehri, L.C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology, 100*, 175-191. doi: 0.1037/0022-0663.100.1.175

Snow, C.E., Lawrence, J.F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2,* 325-344. doi: 10.1080/19345740903167042

Stahl, S.A., & Fairbanks, M.M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72-110. doi:10.2307/1170287

Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M.E. Curtis (Eds.), *The nature of vocabulary acquisition* (pps. 89-105). Hillsdale, NJ: Erlbaum.

Swanson, H.L., & Deshler, D. (2003). Instructing adolescents with learning disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities, 36*, 124-135. doi:10.1177/002221940303600205

Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school

English learners: An intervention study. *Reading & Writing, 22*, 993-1019. doi: 10.1007/s11145-008-9141-y

Wallace, G., & Hammill, D.D. (20130. *The Comprehensive Receptive and Expressive Vocabulary Test-Third Edition*. Austin, TX: Pro-Ed.

Wei, X., Blackorby, J., & Schiller,E. (2011). Growth in reading achievement of students with disabilities, ages 7-17. *Exceptional Children, 78*, 89-106. doi:10.1177/001440291107800106

Wigfield, A., Eccles, J.S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. *Review of Research in Education, 23*, 73-188.

Woodcock, R. W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside.

Table 1

Demographic Data by Treatment Group

	Business As Usual		Treatment		Total	
Variable	(N = 18)		(N = 46)		(N = 64)	
	%	n	%	n	%	n
Gender						
Male	55.6%	10	90%	33	71.7%	43
	44.4%	8	10%	13	28.3%	21
Female						
Special Education Classification						
SLD	77.8%	14	71.7%	33	73.4%	47
SLI	5.6%	1	4.3%	2	4.7%	3
Autism	11.1%	2	10.9%	5	10.9%	7
OHI	5.6%	1	6.5%	3	6.3%	4
	0%	0	6.5%	3	4.7%	3
Missing						
Ethnicity						
	100%	18	76.1%	35	82.8%	53
Hispanic						
White	0%	0	13%	6	9.4%	6
Black	0%	0	4.3%	2	3.1%	2
	0%	0	6.5%	3	4.7%	3
Missing						
Language Preference						
	27.8%	5	26.1%	12	26.6%	17
English						
	72.2%	13	69.9%	32	70.3%	45
Spanish						
	0%	0	4.3%	2	3.1%	2
Missing						

Note: SLD = Specific Learning Disability, SLI = Speech/Language

Impairment, OHI = Other Health Impairment

Table 2

Scores on Measures by Treatment Condition

Measure	Treatment (n = 46)	Business as Usual (n = 18)
CREVT		
Pre Year 2 Intervention	75.62 (9.29)	79.78 (6.74)
Year 2 Spring	78.84 (9.48)*	71.88 (7.71)*
WJ-III Passage Comprehension		
Pre Year 2 Intervention	62.58 (12.16)	62.25(10.13)
Year 2 Spring	69.56 (10.83)	63.36 (11.39)
Cycle 1		
Multiple Choice Pretest	6.85 (2.69)	6.50 (2.48)
Multiple Choice Posttest	14.77 (2.34)*	6.86 (2.00)*
Usage Pretest	3.18 (1.92)	2.78 (1.25)
Usage Posttest	3.73 (2.04)*	3.21 (1.25)*
Cycle 2		
Multiple Choice Pretest	8.36 (3.60)	7.00 (3.00)
Multiple Choice Posttest	14.09 (3.05)*	7.53 (3.36*)
Usage Pretest	4.40 (1.96)	3.60 (2.26)
Usage Posttest	5.72 (1.73)*	3.76 (2.05)*
Cycle 3		
Multiple Choice Pretest	7.45 (3.57)	5.65 (3.20)
Multiple Choice Posttest	13.91 (3.65)*	7.82 (3.43)*
Usage Pretest	3.77 (1.75)	2.65 (1.54)
Usage Posttest	4.34 (2.02)*	3.06 (1.89)*
Maintenance		
Total Pretest	5.66 (3.14)	5.17 (2.53)
Total Posttest	14.96 (3.55)*	6.11 (1.99)*

Note: For students who participated in the Year 1 intervention, standard CREVT scores from Spring of Year 1 were used. For students began their participation in Year 2, standard CREVT scores are from Fall of Year 2, before

intervention was implemented. WJ-III is the Woodcock Johnson Tests of Achievement, Passage Comprehension Subtest. Standard scores are reported and were used in analyses. Significant differences are marked with an asterisk.

Table 3

Estimates of Parameters for Two Latent Growth Models

Parameters	Model 1	Model 2
Means		
α_1	4.61 (0.71)	4.59 (0.70)
α_2	2.74 (1.19)	2.71 (1.17)
α_3	0.75 (0.04)	0.75 (0.04)
Treatment effects		
β_1	1.11 (0.82)	1.13 (0.81)
β_2	5.49 (1.38)	5.35 (1.34)
Retention		
λ_3	0.96 (0.05)	1.00* (----)
Covariances		
ψ_{11}	3.69 (1.70)	3.48 (1.69)
ψ_{22}	15.65 (4.45)	14.71 (4.16)
ψ_{21}	−3.10 (2.21)	−2.82 (2.15)
Measurement residuals		
θ_{11}, θ_{22}, θ_{33}	4.82 (0.85)	4.93 (0.87)
Indices of model fit		
χ^2 (*df*)	5.70 (4)	6.42 (5)
RMSEA [CI]	.079 [.000, .211]	.064 [.000, .188]
CFI, TLI	.983, .974	.985, .982

Note: Tabled values are parameter estimates, with standard errors in parentheses. RMSEA [CI] = root mean square error of approximation, with its confidence interval in brackets. CFI = comparative fit index, TLI = Tucker-Lewis index.

Figure 1. Latent growth model for treatment effects from pretest to maintenance