LSHSS

Research Article

The Impact of Dose and Dose Frequency on Word Learning by Kindergarten Children With Developmental Language Disorder During Interactive Book Reading

Holly L. Storkel,[a] Rouzana Komesidou,[a] Mollee J. Pezold,[a] Adrienne R. Pitt,[a] Kandace K. Fleming,[b] and Rebecca Swinburne Romine[b]

Purpose: The goal was to determine whether interactive book reading outcomes for children with developmental language disorder (DLD) were affected by manipulation of dose (i.e., the number of exposures to the target word during a book reading session) and dose frequency (i.e., the number of repeated book reading sessions) and whether pretreatment factors predicted treatment response variation.
Method: Thirty-four kindergarten children with DLD (aged 5;0–6;2 [years;months]) were taught 1 set of words using the Dose 6 and Dose Frequency 6 format from a prior study (Storkel, Voelmle, et al., 2017) and taught a different set of words using an alternative format, either Dose 4 × Dose Frequency 9 or Dose 9 × Dose Frequency 4, determined through random assignment. Word learning was tracked for each treatment via a definition task prior to, during, and after treatment.
Results: Results showed that children with DLD learned a significant number of words during treatment regardless of the dose and dose frequency format but that significant forgetting of newly learned words occurred in all formats once treatment was withdrawn. Individual differences in word learning were related to Clinical Evaluation of Language Fundamentals Core Language and Understanding Spoken Paragraphs scores.
Conclusion: When administered at an adequate intensity, variation in the dose and dose frequency of interactive book reading does not appear to influence word learning by children with DLD. Although interactive book reading continues to show promise as an effective word learning intervention for children with DLD, further development is needed to enhance the effectiveness of this treatment approach.
Supplemental Material: https://doi.org/10.23641/asha.9745181

Children with developmental language disorder (DLD), specifically those children formerly labeled as having specific language impairment (SLI), need two to three times as many exposures as their peers to support short-term word learning (i.e., training and immediate posttest; Gray, 2003; Rice, Oetting, Marquis, Bode, & Pae, 1994). For example, in the study by Rice and colleagues, children with DLD (labeled as SLI in the study) made comparable word learning comprehension gains immediately posttraining as age-matched typically developing children when given 10 exposures to eight new words, but not when given three exposures to eight new words. Despite the impact of vocabulary deficits, there are few effective treatments (Cirrin & Gillam, 2008). Our goal is to develop interactive book reading as an effective word learning treatment for kindergarten children with DLD. During interactive book reading, an adult reads a book to a child and breaks from the text to teach new words. Clinical trials and meta-analyses show strong, replicable evidence that interactive book reading has moderate to large effects on word learning by typically developing children (Marulis & Neuman, 2010; Mol, Bus, & de Jong, 2009; Mol, Bus, de Jong, & Smeets, 2008), but this approach requires development for children with DLD.

[a]Department of Speech-Language-Hearing: Sciences & Disorders, University
of Kansas, Lawrence
[b]Life Span Institute, University of Kansas, Lawrence
Correspondence to Holly L. Storkel: hstorkel@ku.edu
Rouzana Komesidou is now at the Department of Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA.

Editor-in-Chief: Suzanne Adlof
Received November 11, 2018
Revision received March 11, 2019
Accepted April 14, 2019
https://doi.org/10.1044/2019_LSHSS-VOIA-18-0131
Publisher Note: This article is part of the Forum: Vocabulary Across the School Grades.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

In an earlier preliminary clinical trial (Storkel, Voelmle, et al., 2017), we sought to determine the adequate intensity of interactive book reading for kindergarten children with DLD. Using an escalation design (Hunsberger, Rubinstein, Dancey, & Korn, 2005), we randomized 27 kindergarten children with DLD to one of four intensities of interactive book reading: 12, 24, 36, or 48 exposures to new words. Defining and naming the 30 treated words and a comparable set of 30 untreated words was monitored. Results showed that naming and defining accuracy improved as the intensity increased from 12 to 24 to 36 exposures but then plateaued with no further improvement in accuracy as intensity increased from 36 to 48 exposures. Therefore, we concluded that 36 exposures was the adequate intensity of interactive book reading for children with DLD. Notably, 36 exposures was three times the intensity that had previously been shown to be effective for interactive book reading with typically developing children (Justice, Meier, & Walpole, 2005). In terms of the specific word learning outcomes, children with DLD defined five words correctly at the immediate posttest (Storkel, Voelmle, et al., 2017), which compared favorably to the prior study of typically developing children, who defined six words correctly at the immediate posttest (Justice et al., 2005).

Achieving Adequate Intensity

The adequate intensity of 36 exposures can be achieved in a variety of ways. Intensity for interactive book reading can be decomposed into *dose,* that is, the number of exposures to the target word during a book reading session, and *dose frequency,* that is, the number of repeated book reading sessions (Storkel, Voelmle, et al., 2017; Warren, Fey, & Yoder, 2007). In the prior preliminary clinical trial (Storkel, Voelmle, et al., 2017), 36 exposures was achieved by providing six exposures to the target word during a book reading session (dose = 6) and repeating each book reading session six times before the end of treatment (dose frequency = 6). This combination of dose and dose frequency supported word learning by children with DLD, but it is unknown whether a different combination of dose and dose frequency would have promoted better (or worse) word learning by children with DLD. To our knowledge, no prior studies have compared different methods of achieving the same treatment intensity during interactive book reading by children with DLD. Thus, it is unknown whether manipulation of dose and dose frequency, while holding treatment intensity constant, influences learning by children with DLD. This study seeks to test this proposition. Past experimental research on massed versus distributed practice and theories of learning suggest that it is possible that dose and dose frequency during treatment could influence learning (Storkel, 2015).

Massed Versus Distributed Practice

Studies of massed versus distributed practice for short-term learning suggest that dose and dose frequency may influence learning when extreme values are tested. In massed practice, all exposure occurs in a single session. Thus, for massed practice, the dose is equivalent to the total number of exposures and the dose frequency is 1. In contrast, in distributed practice, the same number of exposures occurs, but the exposures are dispersed across multiple sessions or days. Consequently, in distributed practice, the dose frequency is always greater than 1. To illustrate, in one verb learning study of children with DLD (aged 4;7–6;4 [years;months]), children received 12 total exposures to four nonsense verbs (Riches, Tomasello, & Conti-Ramsden, 2005). In the massed version of that intensity, 12 exposures were achieved through a dose of 12 and a dose frequency of 1 (i.e., 12 exposures in a single session). In the distributed version of that intensity, 12 exposures were achieved through a dose of 3 and a dose frequency of 4 (i.e., three exposures per session repeated across four sessions). In these conditions, children with DLD learned to produce, on average, approximately one word in the distributed condition versus less than one word in the massed condition and learned to comprehend, on average, approximately two words in the distributed condition versus one word in the massed condition. In general, results converge on the finding that distributed practice is more effective than massed practice for many types of learning (see Dempster, 1996, and Underwood, 1961, for reviews). Likewise, for word learning, distributed practice leads to better learning than massed practice for both typically developing children (Childers & Tomasello, 2002; McGregor, Sheng, & Ball, 2007; Schwartz & Terrell, 1983) and children with DLD (Riches et al., 2005). This leads to the prediction that greater learning might be achieved by maximizing the dose frequency during interactive book reading rather than by maximizing the dose (but see Meyers-Denman & Plante, 2016). However, it is important to note that the massed versus distributed practice evidence does not map directly onto what is typically done in clinical treatment. That is, there are few, if any, treatments that involve a single session (i.e., the massed practice condition). In other words, the extreme condition of dose frequency of 1 rarely occurs in clinical treatment. It is unclear whether less extreme contrasts (e.g., dose frequency of 3 vs. 5) would support the conclusion that a higher dose frequency is associated with better learning than a lower dose frequency. In fact, one study of conversational recast treatment suggests that this empirical finding may not translate to clinical practice (Meyers-Denman & Plante, 2016). We seek to further examine the role of dose frequency in clinical treatment via an interactive book reading treatment.

Theories of Learning

Two components of learning, encoding and memory consolidation, are useful in thinking about the impact of dose and dose frequency on learning. Encoding requires the learner to extract the novel word form and meaning from the input and hold this information in working memory while storing an initial representation in the hippocampus (Brown, Weighall, Henderson, & Gareth Gaskell, 2012; Davis, Di Betta, Macdonald, & Gaskell, 2009; Davis &

Gaskell, 2009; Henderson, Weighall, Brown, & Gareth Gaskell, 2012; Henderson, Weighall, & Gaskell, 2013; McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly & Rudy, 2000; Storkel, Komesidou, Fleming, & Romine, 2017). Encoding is an area of deficit in word learning by children with DLD (Gray, 2003; McGregor, Licandro, et al., 2013; Rice et al., 1994; Riches et al., 2005). In contrast, *memory consolidation* involves creating an initial representation in the relevant language areas in the cortex during sleep. The new memory is then integrated with similar memories (Brown et al., 2012; Davis et al., 2009; Davis & Gaskell, 2009; Henderson et al., 2012, 2013; McClelland et al., 1995; Norman & O'Reilly, 2003; O'Reilly & Rudy, 2000), potentially strengthening the new memory (Gaskell & Dumay, 2003; Storkel, 2015) or weakening the new memory due to interference between new and old memories (Storkel, 2015; Tamminen & Gaskell, 2013; Vlach & Sandhofer, 2012). Memory consolidation also is a concern in word learning by children with DLD, but to a lesser degree than encoding (Adi-Japha & Abu-Asba, 2014; McGregor, Licandro, et al., 2013; Oetting, 1999; Rice et al., 1994; Riches et al., 2005; Storkel, Komesidou, et al., 2017).

A review of McGregor, Licandro, et al. (2013) illustrates the interplay between encoding and memory consolidation during word learning. In McGregor, Licandro, et al., adults with a history of DLD and adults with normal developmental history received 48 exposures to 16 novel word–novel object pairings in a single session. Immediate testing at the end of training, which taps encoding, showed group differences in learning of both the word form, the meaning, and the association between the two. Adults were then tested after a 12-hr, 24-hr, and 1-week delay from the end of training. This delayed testing taps memory consolidation. Results were expressed as a difference from the immediate posttest (i.e., how many more or fewer words were responded to correctly at the delay than at the immediate test). Results showed similar change from immediate to delayed testing for adults with DLD and adults with normal development for meaning and the association between form and meaning, indicating similar memory consolidation. However, adults with DLD did show poorer recall of the word form than adults with normal development, suggesting that memory consolidation may be more vulnerable for word forms. This pattern of results led the authors to conclude that encoding was a greater barrier to word learning than memory consolidation for adults with a history of DLD.

During treatment, encoding occurs during the session while information is being presented to the child. Memory consolidation occurs after the session has ended. When multiple sessions are provided, as in distributed learning, encoding and memory consolidation are interleaved. Training on different days allows for re-encoding and reconsolidation. Presumably, distributed practice leads to better performance than massed practice because distributed practice provides an opportunity for consolidation as well as re-encoding and reconsolidation, whereas massed practice does not provide this same opportunity. In terms of dose and dose frequency, dose has the potential to influence encoding. If a child has difficulty with encoding, it is possible that more exposures will be needed during a session (i.e., higher dose) to overcome the child's encoding difficulties. Given that encoding is thought to be the major limiting factor for word learning by children with DLD, a higher dose may lead to better learning than a lower dose.

In contrast, dose frequency determines the number of opportunities to revisit the information and re-encode and reconsolidate that information. This could be important for children with either encoding or consolidation difficulties. If a child has difficulty with encoding, as assumed for children with DLD, but has relatively good memory consolidation, the opportunity to consolidate and re-encode provides an opportunity for intact memory consolidation to support encoding. That is, the child may have only encoded minimal aspects of the word during the first session, but intact memory consolidation may allow the child to retain this partial information in memory. During a second session, the child can now build from this initial representation by encoding a different aspect of the word. In this way, the representation of the word in memory builds over time due to the support of intact memory consolidation. In contrast, if a child has difficulty with memory consolidation, information about a word that is learned in one session is not retained to the next session, but the next session offers an opportunity to re-encode that information. As with poor encoding, the opportunity to interleave stronger and weaker skills, in this case stronger encoding and weaker memory consolidation, facilitates word learning. Taken together, a higher dose frequency may lead to better learning than a lower dose frequency for either encoding or memory consolidation difficulties.

In the current study, children with DLD were treated on two different sets of words sequentially. All children received 36 exposures to one word set using the Dose 6 and Dose Frequency 6 format that was used in the prior preliminary clinical trial (Storkel, Voelmle, et al., 2017). Children also received 36 exposures to a different word set using one of two alternative dose and dose frequency formats: (a) Dose 4 × Dose Frequency 9 and (b) Dose 9 × Dose Frequency 4. The order of the two treatments was counterbalanced across children. The goal of this research was to determine whether different combinations of dose and dose frequency lead to better (or worse) word learning by children with DLD.

Variation in Treatment Response

Another key finding from our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017) was that children with DLD varied in the number of words learned through interactive book reading. Specifically, the number of words learned ranged from 0 to 14 (of 30 taught words). This variability is in keeping with findings from other studies of interactive book reading more generally and experimental word learning studies of children with DLD. However, across these studies, the characteristics that are associated with better versus worse outcomes have varied.

For example, some interactive book reading studies show that children with poorer vocabulary benefit more than children with better vocabulary (Elley, 1989; Justice et al., 2005), whereas other studies report the opposite (Blewitt, Rump, Shealy, & Cook, 2009; Penno, Wilkinson, & Moore, 2002; Robbins & Ehri, 1994). Turning to children with DLD, some studies show that nonword repetition predicts short-term word learning (Alt & Plante, 2006; Jackson, Leitao, & Claessen, 2016), whereas others do not (Gray, 2006; Hansson, Forsberg, Löfqvist, Mäki-Torkko, & Sahlén, 2004).

In our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017), three pretreatment test scores were significantly correlated with the number of words learned by the end of treatment: (a) the Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2; Wagner, Torgesen, Rashotte, & Pearson, 2013) Phonological Awareness composite score, (b) the CTOPP-2 Nonword Repetition standard score, and (c) the Diagnostic Evaluation of Language Variation (DELV; Seymour, Roeper, de Villers, & de Villers, 2005) Semantic adjusted standard score, which takes into account the education level of the child's primary caregiver. That is, children with DLD and better phonological awareness, better nonword repetition, and/or better semantic skills tended to learn more new words during interactive book reading than children with DLD and poorer phonological awareness, poorer nonword repetition, and/or poorer semantic skills. It was thought that these measures were related to word learning outcomes because they indexed the ability to both extract and learn the phonological form of the word and the meaning of the word during encoding. Other measures that captured global or specific language skills (i.e., the Clinical Evaluation of Language Fundamentals–Fourth Edition [CELF-4] Core Language and individual subtest scores), articulation abilities (i.e., the Goldman-Fristoe Test of Articulation–Second Edition), or more global phonological memory (i.e., CTOPP-2 Phonological Memory score) were not related to treatment outcomes.

Although identifying reliable predictors of word learning has been problematic, it is important to continue to examine potential predictors of treatment outcomes so that we can better understand the factors that may be associated with better or worse word learning by children with DLD. This is the foundation of precision medicine, specifically "prevention and treatment strategies that take individual variability into account" (Collins & Varmus, 2015, p. 793). Thus, a secondary goal was to further examine the pretreatment factors associated with the number of words learned by children with DLD during interactive book reading. We, once again, explore a variety of potential predictors, including those that were significant in the prior study (i.e., phonological awareness, nonword repetition, and semantics) and those that were not significant in the prior study but that had the potential to be related to a child's ability to learn during interactive book reading. That is, a child's general language abilities (as measured by the CELF-4 Core Language score) and their ability to understand verbally presented stories (as measured by the CELF-4 Understanding Spoken Paragraphs score) could impact their success in encoding new words during interactive book reading.

Method

Participants

Thirty-four kindergarten children with DLD (M_{age} = 5;6, *SD* = 0;4, range: 5;0–6;2) were recruited through language screenings (91%) and referral from speech-language pathologists (9%). Participating children attended 13 different elementary schools with one to five children in each school (*M* = 2.8, *SD* = 1.5). Within school, there were one to three different classrooms/teachers, for a total of 24 different classrooms/teachers, and one to three children in each class (*M* = 1.2, *SD* = 0.6). A more detailed view of nesting of children in classrooms within schools is shown in Supplemental Material S1. Ninety-one percent of children were seen at their elementary school for most of their sessions. Session location was not specifically tracked, but some final treatment or posttreatment testing sessions did occur after the end of the school year for some children. These final sessions were conducted at a local library or at the child's home. Six percent of children were seen at a local library for all sessions. Three percent of children (i.e., one child) were seen at school initially and then at home for the remainder of the sessions. All children were seen one-on-one in a quiet area of the facility relatively free from distraction.

Sixty-two percent of participants were boys, and 38% were girls. Table 1 shows the demographic characteristics for participants. As shown in Table 1, 24% of parents reported having concerns about their child's speech and language development. However, 59% of children were receiving or had received speech/language services. In addition, 35% of parents reported that individuals in the immediate or extended family had received speech/language services. Most commonly, the individual who had received services was a sibling (58%) or a parent (33%). Likewise, 29% of parents reported that children were receiving or had received special services for other issues outside speech/language concerns. Description of services varied widely with 42% addressing attention, cognition, or behavior; 25% addressing motor issues; 17% addressing academics or learning issues; and 17% addressing social/emotional issues. Preliminary analyses showed no significant difference in performance between children receiving or not receiving other services. Although these special services were reported, none of the children was reported to have any medical or behavioral diagnoses (e.g., attention-deficit/hyperactivity disorder, epilepsy). These two reports seem potentially contradictory, but a medical diagnosis is not a prerequisite to receiving services in the schools, and children with DLD are known to have weaknesses or comorbid disorders in a variety of areas, including coordination, attention, and social interaction (e.g., Bishop, 2004). This co-occurrence of deficits is part of the rationale for

Table 1. Participant demographic characteristics.

Characteristic	% of sample
Parent endorsed concerns about child's speech/language development	24
Parent endorsed that child received special services for speech/language development	59
Parent reported child's immediate (e.g., biological parent/sibling) or extended family (e.g., biological grandparents/aunts/uncles/cousins) received speech/language services	35
Parent endorsed that child received special services in other areas	29
Parent endorsed that child had been diagnosed with a medical or behavioral condition (e.g., ADHD, epilepsy, autism)	0
Race and ethnicity	
White: non-Hispanic	79
White: unknown ethnicity	6
Black/African American: non-Hispanic	6
Black/African American: unknown ethnicity	3
Multiple races: Hispanic	3
Multiple races: non-Hispanic	3
Parent marital status	
Married	53
Divorced	26
Single	21
Mother's education	
Graduate degree	12
College graduate	26
Partial college	32
High school graduate	21
Partial high school	9
Father's education	
Graduate degree	3
College graduate	15
Partial college	35
High school graduate	21
Partial high school	0
Not reported	26

Note. ADHD = attention-deficit/hyperactivity disorder.

moving away from the SLI label to the DLD label (e.g., Bishop, Snowling, Thompson, Greenhalgh, & CATALISE Consortium, 2016). Most of the children were White and non-Hispanic (79%), and 53% of the parents were married. Parent education varied with mothers and fathers typically having a college degree, partial college, or a high school diploma. Participant demographic characteristics generally matched the demographics of the recruitment area, which was eastern Kansas.

Table 2 contains a summary of children's percentile scores on all pretreatment test measures. Eligible children were required to (a) be enrolled in or eligible for kindergarten, (b) pass a hearing screening (American Speech-Language-Hearing Association, 1997), (c) score at or above the 16th percentile for nonverbal cognition on the Reynolds Intellectual Assessment Scale (Reynolds & Kamphaus, 2003), (d) have a Core Language Score at or below the 10th percentile on the CELF-4 (Semel et al., 2003), and (e) score at or below the 10th percentile on at least one of two vocabulary measures: the DELV Semantic subtest (Seymour et al., 2005) or the CELF-4 Word Classes subtest. For vocabulary scores, most children (53%) qualified on both the DELV and the CELF, 35% qualified on the DELV only, and 12% qualified on the CELF only. Children also completed supplementary tests to further characterize their abilities, including the CTOPP-2 Elision, Sound Matching, Blending Words, Nonword Repetition, and Memory for Digits subtests (Wagner et al., 2013); the CELF Understanding Spoken Paragraphs subtest; and the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000). The majority showed deficits on all subtests of the CELF and the CTOPP, indicating broad language impairment and deficits in preliteracy skills. Most children (76%) had average articulation skills.

Children were randomly assigned to one of four treatment arms, which varied in the sequencing of the standard Dose 6 and Dose Frequency 6 treatment and the alternative treatment and in which alternative treatment was assigned: (a) Dose 4 × Dose Frequency 9 or (b) Dose 9 × Dose Frequency 4. This is illustrated in Figure 1. Demographic and testing data for children in each of the treatment arms are presented in Supplemental Materials S2 and S3.

Treatment

There were four treatment arms, depicted in Figure 1. As shown in Figure 1, each child received two treatment conditions, Treatment 1 and Treatment 2, focused on

Table 2. Percentile scores for participants on standardized clinical tests.

Test	*M*	*SD*	Range	At or below the 10th percentile (%)
RIAS Nonverbal IQ	55	23	16–95	0
CELF Core Language	2	3	0.1–10	100
Vocabulary: DELV Semantic	6	11	0.1–50	88
Vocabulary: CELF Word Classes	17	21	0.1–75	65
CELF[a] Concepts & Following Directions	6	8	0.1–37	88
CELF[a] Word Structure	5	6	0.1–25	88
CELF[a] Recalling Sentences	6	11	0.1–50	85
CELF[a] Formulating Sentences	6	7	0.1–25	85
CELF Understanding Spoken Paragraphs	8	8	0.1–25	79
CTOPP[b] Nonword Repetition	15	21	1–75	71
CTOPP Phonological Memory	10	15	1–68	71
CTOPP Phonological Awareness	7	6	1–25	76
GFTA	26	23	1–78	24

Note. RIAS = Reynolds Intellectual Assessment Scale (Reynolds & Kamphaus, 2003); CELF = Clinical Evaluation of Language Fundamentals (Semel, Wiig, & Secord, 2003); DELV = Diagnostic Evaluation of Language Variation (Seymour et al., 2005); CTOPP = Comprehensive Test of Phonological Processing (Wagner et al., 2013); GFTA = Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 2000).

[a]Scores on this subtest contributed to the CELF Core Language score. [b]Scores on this subtest contributed to the CTOPP Phonological Memory composite score.

different sets of words, Set 1 and Set 2, respectively. Each treatment provided 36 cumulative exposures to the target words. All children received the Dose 6 × Dose Frequency 6 standard version of the treatment from the prior preliminary clinical trial and one of the new alternative versions: (a) Dose 4 × Dose Frequency 9 or (b) Dose 9 × Dose Frequency 4. In addition, the order of the two treatment conditions (standard Dose 6 × Dose Frequency 6 vs. new alternative version) varied across arms.

Treatment Materials

Materials consisted of 10 commercially available books with colorful illustrations and Tier 2 vocabulary words (i.e., less frequent vocabulary words but common in academic settings). The books and words were taken from the study of Justice et al. (2005) and were used in our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017). Six target words were selected from each book, for a total of 60 words. Target words consisted of 16 nouns, 25 verbs, and 19 adjectives. The 10 books were divided into two sets of five books, Set A and Set B, with each set containing 30 target words. The two sets of books were randomly assigned to the two treatment conditions the child received. The words assigned to Treatment 1 are referred to as *Set 1 words,* and the words assigned to Treatment 2 are referred to as *Set 2 words.*

Dose × Dose Frequency

As shown in Table 3, this study involved three treatment versions within intensity 36: (a) 4 × 9 consisted of four exposures to a target word during book reading (i.e., dose = 4) and nine repeated readings of each book (i.e., dose frequency = 9), (b) 6 × 6 consisted of six exposures to a target word during book reading (i.e., dose = 6) and six repeated readings of each book (i.e., dose frequency = 6), and (c) 9 × 4 consisted of nine exposures to a target word during book reading (i.e., dose = 9) and four repeated readings of each book (i.e., dose frequency = 4). Table 3 also shows how dose and dose frequency affected time, a less precise measure of treatment intensity. Specifically, as dose increased, the length of the book reading session in minutes increased slightly from a low of 13 min (Dose 4) to a high of 16 min (Dose 9). In complement, as dose frequency increased, the number of sessions increased from a low of 10 sessions (Dose Frequency 4) to a high of 23 sessions (Dose Frequency 9). Typically, children received two treatment sessions per week. Therefore, there also was a corresponding increase in the number of weeks of treatment from a low of 5 weeks (Dose Frequency 4) to a high of 12 weeks (Dose Frequency 9).

Treatment Form

Treatment form is defined as the "typical activity within which teaching episodes are delivered" (Warren et al., 2007, p. 71). A summary of our prior stimuli is presented in Storkel, Voelmle, et al. (2017) and in our online archive (Storkel, 2016). We made slight adjustments to the treatment scripts based on our experience during the prior preliminary clinical trial and based on the needs of the current research design. The scripts used for the current study are included as Supplemental Material S4. In a treatment session, children participated in activities for two books. The two books were treated sequentially. For all treatment conditions, the six target words in a given book were taught through (a) a pre–book reading activity that previewed the six target words using colorful pictures that were not from the storybook and a lowercase orthographic label in 40-point Calibri font under the picture; (b) reading of the storybook with the target words highlighted by a box, potentially with departures from the text to teach the target words (depending on treatment condition); and (c) a post–book reading

Figure 1. Overall research design including pretreatment definition testing of Set 1 and Set 2 words (white box); first treatment focusing exclusively on Set 1 words, definition testing of Set 1 and Set 2 words immediately and 2 weeks post–Treatment 1 (black boxes); and second treatment focusing exclusively on Set 2 words, definition testing of Set 1 and Set 2 words immediately and 2 weeks post–Treatment 2 (black boxes). During treatment, only the treated words were tested via naming and definition tasks. Testing during treatment is indicated by an asterisk next to the number of exposures when testing occurred.

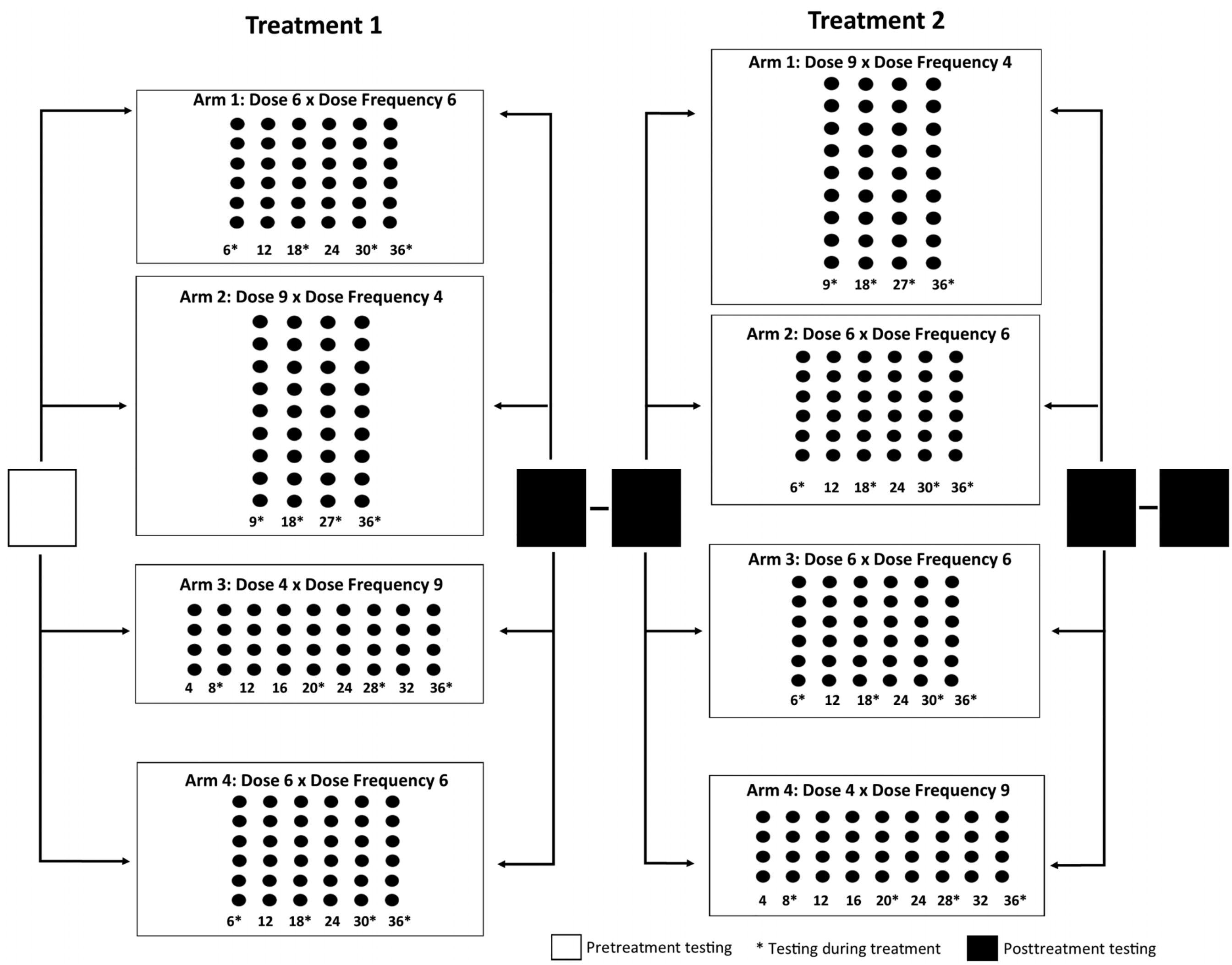


Table 3. Dose, dose frequency, and treatment time for each treatment condition.

Treatment condition	Treatment		Treatment time		
	Dose[a]	Dose frequency[b]	Average session length (min)	No. of sessions	No. of weeks[c]
4 × 9	4	9	13	23	12
6 × 6	6	6	14	15	8
9 × 4	9	4	16	10	5

[a]Dose refers to the number of exposures to a target word during a book reading session. [b]Dose frequency refers to the number of repeated book reading sessions. [c]Typically two sessions occurred per week.

activity that reviewed the six target words with a different set of colorful pictures than the pre–book reading or storybook reading activity and a lowercase orthographic label in 40-point Calibri font under the picture. During these activities, children with DLD were taught the target words via storybooks, definitions, synonyms, and supportive context sentences. An example of the treatment form for each treatment condition is shown in Table 4 for the target word *overjoyed.*

Dose 4 × Dose Frequency 9

In Version 4 × 9, children are exposed to a target word four times during the session (i.e., Dose 4). As shown in Table 4, children hear the target word *overjoyed*: (a) via a definition (Exposure 1) and a synonym (Exposure 2) during the pre–book reading activity, (c) via the text from book (Exposure 3) during the book reading activity, and (c) via a context sentence (Exposure 4) during the post–book reading activity. This set of activities is repeated on nine different occasions (i.e., Dose Frequency 9).

Dose 6 × Dose Frequency 6

In Version 6 × 6, children are exposed to a target word six times during the session (i.e., Dose 6). As shown in Table 4, children hear the target word *overjoyed*: (a) via a definition (Exposure 1) and a synonym (Exposure 2) during the pre–book reading activity, (b) via the text from book (Exposure 3) and a synonym (Exposure 4) during the book reading activity, and (c) via a context sentence (Exposure 5) and a definition (Exposure 6) during the post–book reading activity. This set of activities is repeated on six different occasions (i.e., Dose Frequency 6).

Dose 9 × Dose Frequency 4

In Version 9 × 4, children are exposed to a target word nine times during the session (i.e., Dose 9). As shown in Table 4, children hear the target word *overjoyed*: (a) via a definition (Exposure 1), a synonym (Exposure 2), and a context sentence (Exposure 3) during the pre–book reading activity; (b) via the text from book (Exposure 4), a synonym (Exposure 5), and a definition (Exposure 6) during the book reading activity; and (c) via a context sentence (Exposure 7), a synonym (Exposure 8), and a definition (Exposure 9) during the post–book reading activity. This set of activities is repeated on four different occasions (i.e., Dose Frequency 4).

Treatment Providers

Ten graduate students in speech-language pathology conducted most of the treatment sessions (99.7%). A small number of treatment sessions (0.3%) were conducted by a lab coordinator with an out-of-field degree. In terms of the graduate students, 30% of treatment sessions were conducted by two PhD students who were certified speech-language pathologists, 23% of treatment sessions were conducted by two PhD students who had completed the clinical Master of Arts (MA) requirements but had not yet completed the clinical fellowship year for certification, and 47% of treatment sessions were conducted by six MA speech-language pathology students. Training was the same for all treatment providers and included three steps. First, treatment providers reviewed the scripts and treatment materials (e.g., books, preview and review pictures) and watched prior treatment sessions until they felt comfortable with the scripted protocol. For the initial treatment providers, prior treatment sessions were taken from the previous preliminary clinical trial (Storkel, Voelmle, et al., 2017). Second, treatment providers conducted practice treatment sessions with experienced treatment providers acting as the target children. Finally, treatment providers were observed by an experienced treatment provider during their initial treatment session to ensure that the protocol was followed and that the provider did not have any questions. All providers were successful during the first observation.

Table 4. Example treatment for the target word *overjoyed* for each treatment condition.

Activity [visual image]	Dose 4 × Dose Frequency 9	Dose 6 × Dose Frequency 6	Dose 9 × Dose Frequency 4
Pre–book reading [color picture of a girl playing in a park with the printed word "overjoyed" underneath the picture]	*Overjoyed* means filled with joy; very happy. [definition] *Overjoyed* is like excited. [synonym]	*Overjoyed* means filled with joy; very happy. [definition] *Overjoyed* is like excited. [synonym]	*Overjoyed* means filled with joy; very happy. [definition] *Overjoyed* is like excited. [synonym] The little girl was *overjoyed* to play outside at the park. [context sentence]
Book reading: *Imogene's Antlers* (Small, 1985) [storybook picture of Imogene and her family looking happy and accompanying printed text with the word "overjoyed" marked with a box]	"When she came down to breakfast, the family was *overjoyed* to see her back to normal…" (p. 25) [text from book]	"When she came down to breakfast, the family was *overjoyed* to see her back to normal…" (p. 25) [text from book] *Overjoyed* is like excited. [synonym]	"When she came down to breakfast, the family was *overjoyed* to see her back to normal…" (p. 25) [text from book] *Overjoyed* is like excited. [synonym] *Overjoyed* means filled with joy; very happy. [definition]
Post–book reading [color picture of a boy getting a puppy for his birthday with the printed word "overjoyed" underneath the picture]	He was *overjoyed* to get a puppy for his birthday. [context sentence]	He was *overjoyed* to get a puppy for his birthday. [context sentence] *Overjoyed* means filled with joy; very happy. [definition]	He was *overjoyed* to get a puppy for his birthday. [context sentence] *Overjoyed* is like excited. [synonym] *Overjoyed* means filled with joy; very happy. [definition]

Although we intended for each child to be seen consistently by only one or two treatment providers, scheduling needs often caused inconsistency. In some cases, these were short-term adjustments (e.g., covering for a treatment provider who was sick or out of town). In other cases, these were longer-term adjustments (e.g., changes due to semester schedule change for graduate students). To capture the consistent pattern, the providers who conducted 25% or more of the treatment sessions were tallied and examined to determine the number of treatment providers for each child. Results showed that 44% of children were seen primarily by one treatment provider, 29% were seen primarily by two treatment providers, 15% were seen primarily by three treatment providers, 9% were seen by one treatment provider for Treatment 1 and a different treatment provider for Treatment 2, and 3% were seen by four treatment providers. Supplemental Material S5 provides more details about the distribution of treatment providers across children for Treatments 1 and 2.

Treatment Fidelity

Treatment fidelity was checked for 20% of sessions. Research assistants watched videos of selected sessions and used a checklist to determine that the correct number of exposures to a target word and correct treatment form were delivered. Two scores were derived. The first score, derived by dividing the total number of exposures administered by the intended number of exposures, was 99.90%. The second score, derived by dividing the total number of correct treatment forms administered by the total number of treatment forms, was 99.90%.

Outcome Measure

The primary outcome measure typically was collected and scored by research assistants who were unaware of the participant's assigned treatment condition. Research assistants administered the secondary outcome measure during treatment sessions, but other research assistants who were unaware of treatment assignment scored that measure. Research Electronic Data Capture tools hosted at the University of Kansas Medical Center were used to collect and manage study data (see Harris et al., 2009, for more information).

Primary Outcome Measure: Definition Task

As in our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017), the primary outcome measure was a definition task. We chose a definition task for two reasons: (a) to afford comparison to our past research with this treatment to determine whether continuing changes to the treatment improve outcomes and (b) to ensure that the treatment is helping children establish the rich word knowledge needed to support a variety of linguistic tasks. Although this task sets a high bar for word learning, our past research shows that it is sensitive enough to detect changes between different treatment conditions. The definition task was administered before treatment, immediately following the conclusion of each treatment, and 2 weeks following the conclusion of each treatment. The delay between the final treatment session and the immediate posttest was 5 days on average for Treatment 1 (SD = 3, range: 1–19) and 6 days on average for Treatment 2 (SD = 5, range: 1–27). The delay between the final treatment session and the 2-week posttest was 21 days on average for Treatment 1 (SD = 14, range: 9–90) and 21 days on average for Treatment 2 (SD = 8, range: 12–49). For a breakout of delay for each treatment condition, see Supplemental Material S6.

Research assistants administered the task across two sessions with 15 Set 1 and 15 Set 2 words tested in each session. Since children learned Set 1 words during Treatment 1 and Set 2 words during Treatment 2, half of the words had been recently taught and the other half were either untreated control words (for testing after Treatment 1) or previously treated words (for testing after Treatment 2). The children first heard three practice words (*bed, ball,* and *candy*), which were words likely to be known by children with DLD. Then, Set 1 (n = 15) and Set 2 (n = 15) words were presented in random order along with familiar words (n = 10, e.g., *chair, teacher, apple*). For each word, the child received the prompt, "Tell me what [word] means." Prompts were prerecorded and presented with computer software so that pronunciations of the words were consistent across tasks. Children's responses were audio-recorded and transcribed for later scoring. The testing script is included as Supplemental Material S7.

Fourteen graduate students in speech-language pathology conducted most of the testing sessions (99.4%). A small number of testing sessions (0.6%) were conducted by a lab coordinator with an out-of-field degree. Of the 15 total testers, 10 also served as treatment providers (i.e., nine graduate students and the lab coordinator). It was intended that the nontreatment providers be the primary testers, but due to scheduling constraints, treatment providers sometimes had to conduct testing. In terms of the graduate student testers who were not treatment providers, 65% of testing sessions were conducted by one PhD student who had not had prior clinical training, and 3% of testing sessions were conducted by four MA students. In terms of the graduate student testers who also were treatment providers, 5% of testing sessions were conducted by two PhD students who were certified speech-language pathologists, 3% of testing sessions were conducted by two PhD students who had completed the clinical MA requirements but had not yet completed the clinical fellowship year, and 24% of testing sessions were conducted by five MA speech-language pathology students. Although it was intended that testers would be unaware of the child's treatment condition (i.e., blind), there were several instances (4% of testing sessions distributed across four participants) where this was not possible, usually for children who had very challenging schedules. Thus, blinding during testing was achieved for 96% of the testing sessions. Supplemental Material S5 provides more details about the distribution of testers relative to treatment providers and highlights cases where blinding was not fully achieved.

In terms of training the testers, all testers were shown how to use the computer software and the recording device. Testers practiced administering the definition test to other trained testers until they were comfortable with the equipment and protocol.

The definition scoring procedures from our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017) were used in the current study. In the prior study, the research team followed the procedures of McGregor, Oleson, Bahnsen, and Duff (2013) and consulted dictionaries to create a scoring rubric that listed common elements of an accurate and complete definition for each word (e.g., pouted = negative emotion + lips/face). Possible scores were as follows: 0 points for an incorrect or absent definition (e.g., pouted = face), 1 point for an appropriate use of the word in a sentence (e.g., pouted = pouted because you don't want to eat and your mom makes you) or for a vague definition (e.g., pouted = cry), 2 points for a conventional definition containing at least one critical element but lacking other critical elements (e.g., pouted = mad), and 3 points for a complete and accurate definition including all critical elements (e.g., pouted = feel mad and push your lips out). Two raters independently scored each response following the rubric guidelines. The raters compared their scores and resolved disagreements through consensus. Raters disagreed on scoring for 2.65% of the data. In rare cases when the two raters could not reach consensus, they consulted a third rater. Although raters were encouraged to formally track their use of a third rater, other raters were often informally consulted because they were present at the time of discussion. Raters formally sought a third rater to resolve disagreements for less than 1% of words.

In the analyses, children's definitions scored as 2 or 3 (i.e., a partially or completely accurate definition) were counted as correct and definitions scored as 0 or 1 (i.e., incorrect definition, absent definition, correct use of a word in a sentence, or vague definition) were counted as incorrect. This was done to afford comparison to prior studies (Justice et al., 2005; Storkel, Voelmle, et al., 2017). However, Table 5 shows the distribution of scores across test points. As in the prior study, scores of 0 predominate the incorrect category and scores of 2 predominate the correct category. Pretreatment and immediate posttreatment score distributions from the current study closely match (i.e., 0- to 2-point difference) those reported in the prior clinical trial (Storkel, Voelmle, et al., 2017; see Table 5, treated words).

Secondary Outcome Measure: Interim Definition and Naming Tasks

Learning also was tracked during treatment. The research assistant who provided the treatment prompted children to provide definitions or name the target words at four points during each treatment. The exact number of exposures for three of the four tests varied by treatment condition, as shown in Table 6. The fourth and final test always occurred during the last treatment session for a given book, which corresponded to 36 exposures for all treatment conditions. The words were assessed in a fixed order while the child viewed the prereading or postreading pictures for each word. For definition prompts, the research assistant asked, "What does [word] mean?" Specific feedback was not provided, but the correct definition always was provided after the child's response regardless of the accuracy of the response. This is shown in the scripts provided in Supplemental Material S4. Definition responses were scored by two independent raters unaware of children's treatment assignment following the procedures described above. For interim definition scoring, raters disagreed on the score for 5.38% of the data. Again, although disagreements were not always formally tracked, raters formally consulted a third rater in less than 1% of scoring disagreements. The rate of disagreements was likely higher for interim definitions than pre/post definitions because children had more correct responses during treatment than they did before or after treatment.

Naming data are not reported here because they showed the same pattern as the definition data. Moreover, the definition data were more complete because definitions were gathered pretreatment and posttreatment, whereas naming was only tested during treatment. However, we describe the naming task for completeness. For naming

Table 5. Percentage (%) of definition responses receiving a score of 0, 1, 2, or 3 at each test point and a summary of overall correct responding based on scoring 2 and 3 as correct (as in the reported analysis) or 1, 2, and 3 as correct.

Test point	0	1	2	3	2 + 3 = correct	1 + 2 + 3 = correct
Pretreatment[a]	95	3	2	0	2	5
Final treatment session (36 exposures)[b]	55	11	20	14	34	45
Immediate posttreatment[b]	77	9	9	6	15	24
2-week posttreatment[c]	81	8	7	4	11	19

Note. Scoring was as follows: 0 points for an incorrect or absent definition, 1 point for appropriate use of the word in a sentence or a vague definition, 2 points for a conventional definition that lacks one or more critical elements, and 3 points for a complete and accurate definition including all critical elements.

[a]Percentage out of 2,040 responses (60 words × 34 children). [b]Percentage out of 2,040 responses for target words only (i.e., excluding control words) summed across Treatments 1 and 2 (60 words × 34 children). [c]Percentage out of 2,009 responses for target words only (i.e., excluding control words) summed across Treatments 1 and 2 (60 words × 34 children; one child did not complete the second 2-week posttest, and data were missing for one word).

Table 6. Number of elaborated exposures to target words received at each interim treatment testing point.

Interim testing point	Dose 4 × Dose Frequency 9	Dose 6 × Dose Frequency 6	Dose 9 × Dose Frequency 4
1	8	6	9
2	20	18	18
3	28	30	27
4	36	36	36

prompts, the research assistant showed the child the post–book reading picture without the orthographic label and asked a question meant to elicit the phonological form of the target word (e.g., "What is the lightning doing?" to elicit *flashing*). Specific feedback was not provided, but the correct orthographic label and context sentence always were provided after the child's response regardless of the accuracy of the response. This is shown in the scripts provided in Supplemental Material S4. Naming responses were scored as correct or incorrect. Correct responses included the word itself (e.g., *glared*), the word in its bare form or with a different ending (e.g., *glare, glaring*), or an acceptable phonological substitution of the word (e.g., *gwaring,* judged on an individual basis). One independent judge scored naming responses, and scores were verified by another judge.

Summary of the Design

Figure 1 provides a summary of the design. Upon meeting the selection criteria, children were randomized to one of four treatment arms, as shown in Figure 1. Pretreatment definition testing was conducted for all 60 words, as indicated by the white box in Figure 1. Treatment 1 was then initiated for 30 words (Set 1 words). Set 1 words were tested via naming and definition tasks four times during Treatment 1, as indicated by asterisks in Figure 1. Upon completion of Treatment 1, all 60 words (Sets 1 and 2) were tested immediately and 2 weeks after treatment via the definition task. Note that Set 1 words have received treatment but Set 2 words have not yet received treatment and serve as untreated control words. Treatment 2 was then initiated for the remaining 30 words (Set 2 words). Set 2 words were tested via naming and definition tasks four times during Treatment 2, as indicated by the asterisks in Figure 1. Upon completion of Treatment 2, all 60 words (Sets 1 and 2) were tested immediately and 2 weeks after treatment via the definition task. Note that Set 2 words have just received treatment, whereas Set 1 words received treatment much earlier. Thus, for Set 1 words, these test points indicate longitudinal maintenance of learning following 5–12 weeks of treatment on a different word set.

Data Analysis Strategy

A series of multilevel longitudinal models were evaluated using SAS Proc Mixed to describe the change in number of Set 1 and Set 2 words accurately defined over seven test points: pretreatment, 36 exposures in Treatment 1 (i.e., last Treatment 1 session), immediate post–Treatment 1, 2-week post–Treatment 1, 36 exposures in the second treatment (i.e., last Treatment 2 session), immediate post–Treatment 2, and 2-week post–Treatment 2. Because the number of exposures varied for the interim definition tests, only the 36-exposure interim definition test data, which was consistent for all treatment conditions, were used. We expected significant slopes for Set 1 words during Treatment 1 when they were the target of intervention. We had no hypotheses about whether Set 1 words would decline or would be maintained after Treatment 1 ended and monitoring continued during Treatment 2. We expected nonsignificant slopes for Set 2 words during Treatment 1, when those words were not the focus of intervention, and significant slopes for Set 2 words during Treatment 2 when they were the target of intervention. Because the number of words defined were counts, the natural log of the words defined was modeled resulting in a model equivalent to a generalized linear mixed model with a negative binomial link function.

The significance of fixed effects was evaluated using Wald tests, and the significance of random effects was evaluated using likelihood ratio tests and information criteria between models with the same fixed effects. We evaluated models with time coded as days since the start of intervention as well as categorically by intervention point. The categorical model fitted the data better. The final categorical coding of time is shown in Table 7. Because there were two treatment phases (Treatment 1 vs. Treatment 2) with different expectations for change in Set 1 versus Set 2 words in each phase, we examined a piecewise model of change in which the change related to the first treatment phase (i.e., pretreatment, 36 exposures in Treatment 1, immediate post–Treatment 1) was described by one linear slope and change related to the second treatment phase (i.e., 2-week post–Treatment 1, 36 exposures in the second treatment, and immediate post–Treatment 2) was described by a second linear slope. This allows the slope for each word set to differ across each phase, as predicted.

Results

Overall Learning

Prior to examining differences in slopes, we fit a saturated means, unstructured variance model and examined differences in words defined across time. Table 8 shows the results for Set 1 and Set 2 words. There was a significant increase in the log of the number of Set 1 words correctly defined from pretreatment to 36 exposures at the end of

Table 7. Coding of the seven test points for Treatment Phases 1 and 2 in the statistical model.

Test point	Treatment Phase 1	Treatment Phase 2
Pretreatment	0	0
36 exposures in Treatment 1	1	0
Immediate post–Treatment 1	2	0
2-week post–Treatment 1	2	0
36 exposures in Treatment 2	2	1
Immediate post–Treatment 2	2	2
2-week post–Treatment 2	2	2

Treatment 1. Thus, children learned Set 1 words when they were the focus of treatment. Then, there was a significant decrease in the log of the number of Set 1 words correctly defined at 36 exposures at the end of Treatment 1 to the immediate post–Treatment 1 test, indicating that children forgot some of the learned words in the interval between the last treatment session and the posttreatment test. However, there were no further changes as Set 1 words continued to be monitored without treatment, indicating that no further forgetting occurred. Although we refer to this as forgetting, it is important to note that, during treatment, children are only tested on the six words that are currently a focus of treatment, whereas posttreatment testing sessions include a larger set of words. Thus, words that were previously defined correctly could have been forgotten due to the passage of time (a more traditional conceptualization of forgetting) or due to the larger number of words being tested (perhaps a less traditional conceptualization of forgetting).

Turning to Set 2 words in Table 8, there were no significant differences during Treatment 1 (pretreatment, immediate post–Treatment 1, and 2-week post–Treatment 1), when Set 2 words were being monitored. Once treatment of Set 2 words was initiated, there was a significant increase in the log of the number of Set 2 words correctly defined at the last Treatment 2 session after 36 exposures had accumulated. Thus, children learned Set 2 words when they were the focus of treatment. Then, there was a significant decrease in the log of the number of Set 2 words correctly defined from the last Treatment 2 session to the immediate post–Treatment 2 test, indicating that children forgot some of the learned words in the interval between the last treatment session and the posttreatment test. However, there were no further significant changes in Set 2 words, indicating that no further significant forgetting occurred.

The best-fitting piecewise growth models for both sets of words included random intercepts and fixed slopes for both treatment phases. Final parameter estimates are shown in Supplemental Material S8. During Treatment 1 for Set 1 words (which were the focus of treatment), there was a significant positive slope of .26, $SE = .09$, $p = .002$, indicating significant learning of Set 1 words during Treatment 1. During Treatment 2, the slope for Set 1 words (which were just being monitored) was a significant negative slope of −.28, $SE = .07$, $p < .001$, indicating significant forgetting. Turning to Set 2 words, during Treatment 1, the slope for Set 2 words (which were not the focus of treatment) was .14, which was not significant, $p = .065$, indicating that Set 2 words were not being learned when they were not the focus of treatment. However, once Set 2 words were treated during Treatment 2, there was a significant positive slope of .48, $SE = .06$, $p < .0001$, indicating significant learning. Taken together, these data show that there are changes in the number of words accurately defined as treatment is initiated or withdrawn for each set of words. This replicates the effectiveness of interactive book reading at an intensity of 36 exposures for kindergarten children with DLD but also highlights the need to further examine and, possibly, boost long-term retention of taught words.

Table 8. Results for the saturated means, unstructured variance model used to examine differences in Set 1 and Set 2 words defined across time.

	Set 1 words				Set 2 words			
Test point	Estimate	*SE*	Lower	Upper	Estimate	*SE*	Lower	Upper
Pretreatment	0.37	0.07	0.21	0.52	0.41	0.08	0.26	0.57
36 exposures in Treatment 1	2.24[a]	0.15	1.92	2.56	N/A[b]	N/A[b]	N/A[b]	N/A[b]
Immediate post–Treatment 1	1.29[a]	0.13	1.02	1.57	0.46	0.09	0.27	0.65
2-week post–Treatment 1	1.00	0.12	0.75	1.25	0.44	0.08	0.27	0.61
36 exposures in Treatment 2	N/A[c]	N/A[c]	N/A[c]	N/A[c]	2.20[a]	0.12	1.96	2.44
Immediate post–Treatment 2	0.83	0.13	0.58	1.09	1.45[a]	0.13	1.17	1.72
2-week post–Treatment 2	0.85	0.12	0.60	1.10	1.36	0.14	1.08	1.65

[a]Significant change compared to the immediately prior test point. [b]Not applicable. Set 2 words were not tested at the last Treatment 1 session. Only treated words were tested in treatment sessions. [c]Not applicable. Set 1 words were not tested at the last Treatment 2 session. Only treated words were tested in treatment sessions.

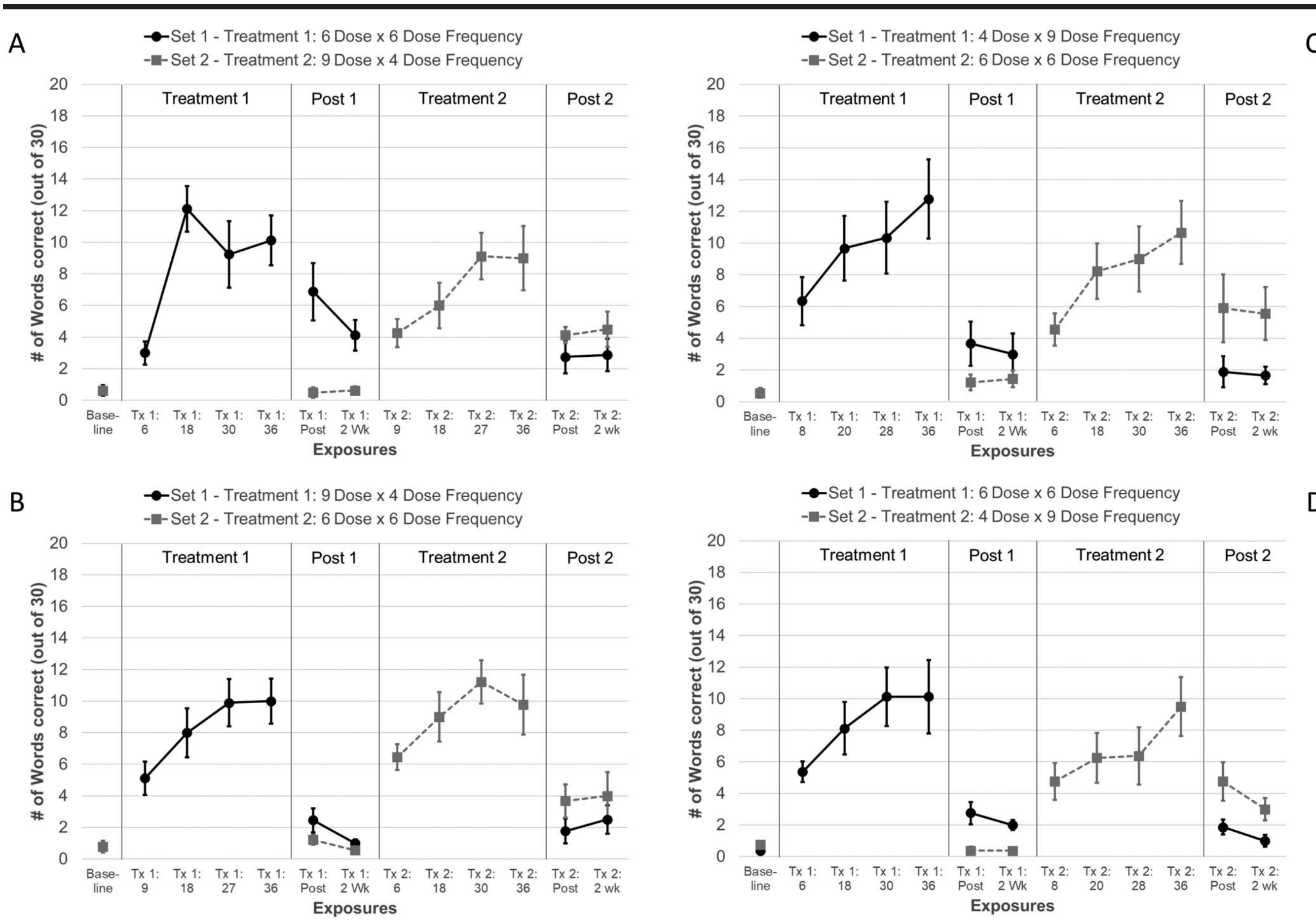
Dose × Dose Frequency Comparisons

To address differences between the Dose × Dose Frequency combinations as well as individual differences, we added predictors to the model to determine if the Dose × Dose Frequency and pretreatment learner characteristics were significantly related to growth in word defining within each treatment phase. All continuous learner characteristics were grand mean centered.

For Set 1 words, Dose × Dose Frequency did not significantly affect slope in either treatment phase, $p = .40$ and $p = .63$, respectively. For Set 2 words, Dose × Dose Frequency did not significantly impact slopes during either phase, $p = .37$ and $p = .19$, respectively. Raw trends in differences between the Dose × Dose Frequency conditions were explored to determine whether any trends were apparent.

Figure 2 illustrates the trends in the full data set for each treatment arm. As shown in Figure 2A, definition accuracy for Set 1 words improves when they are the focus of the Dose 6 × Dose Frequency 6 treatment during the Treatment 1 phase. Children define, on average, 10 words correctly (*SD* = 4, range: 5–18) at 36 exposures in Treatment 1. Once treatment ends, performance declines with children defining an average of seven words correctly (*SD* = 5, range: 0–15) at immediate post–Treatment 1 and four words correctly (*SD* = 3, range: 0–9) at 2-week post–Treatment 1. Further decline is observed during the post–Treatment 2 phase with approximately three words (range: 0–8) defined correctly at each test. Definition accuracy for Set 2 words is relatively low at pretreatment with approximately one word defined correctly (range: 0–2) and remains low in post–Treatment 1 phase with approximately one word defined correctly (range: 0–2), which is expected given that Set 2 words were not the focus of treatment. When treatment is initiated for Set 2 words with the Dose 9 × Dose Frequency 4 treatment, definition accuracy improves with children defining nine words defined correctly (*SD* = 6, range:

Figure 2. Mean number of Set 1 (black circles) and Set 2 (gray squares) words defined correctly at pretreatment baseline, during Treatment 1 (6, 18, 30, and 36 exposures), immediately post–Treatment 1, 2 weeks post–Treatment 1, during Treatment 2 (9, 18, 27, and 36 exposures), immediately post–Treatment 2, and 2 weeks post–Treatment 2. Error bars indicate standard errors. (A) Treatment 1 is Dose 6 × Dose Frequency 6, and Treatment 2 is Dose 9 × Dose Frequency 4. (B) Treatment 1 is Dose 9 × Dose Frequency 4, and Treatment 2 is Dose 6 × Dose Frequency 6. (C) Treatment 1 is Dose 4 × Dose Frequency 9, and Treatment 2 is Dose 6 × Dose Frequency 6. (D) Treatment 1 is Dose 6 × Dose Frequency 6, and Treatment 2 is Dose 4 × Dose Frequency 9.



3–19) at 36 exposures in Treatment 2 but then drops to four words correct (*SD* = 1, range: 2–7) at immediate post–Treatment 2 and five words correct (*SD* = 3, range: 0–9) at 2-week post–Treatment 2. In general, outcomes appear relatively similar across Dose 6 × Dose Frequency 6 and Dose 9 × Dose Frequency 4 with similar numbers of words defined correctly at 36 exposures (10 vs. 9), immediate post (7 vs. 4), and 2-week post (4 vs. 5).

Figure 2b shows the same two treatments but in the reverse order. Here, Dose 9 × Dose Frequency 4 results in children accurately defining (on average) 10 words correctly (*SD* = 4, range: 1–16) at the last treatment session, two words correctly (*SD* = 2, range: 0–8) at the immediate post, and one word correctly (*SD* = 1, range: 0–2) at the 2-week post. Turning to the second treatment, Dose 6 × Dose Frequency 6 results in children accurately defining (on average) 10 words correctly (*SD* = 6, range: 2–19) at the last treatment session, four words correctly (*SD* = 3, range: 0–9) at the immediate post, and four words correctly (*SD* = 5, range: 0–11) at the 2-week post. Again, the number of words defined correctly in Dose 6 × Dose Frequency 6 versus Dose 9 × Dose Frequency 4 is relatively similar at 36 exposures (10 vs. 10), immediate post (4 vs. 2), and 2-week post (4 vs. 1).

Figures 2C and 2D show the patterns for the Dose 6 × Dose Frequency 6 treatment and the Dose 4 × Dose Frequency 9 treatment. In Figure 2C, Dose 4 × Dose Frequency 9 results in children accurately defining (on average) 13 words correctly (*SD* = 7, range: 0–21) at 36 exposures, four words correctly (*SD* = 4, range: 0–13) at the immediate post, and three words correctly (*SD* = 4, range: 0–12) at the 2-week post. Turning to the second treatment, Dose 6 × Dose Frequency 6 results in children accurately defining (on average) 11 words correctly (*SD* = 6, range: 2–19) at 36 exposures, six words correctly (*SD* = 6, range: 0–16) at the immediate post, and six words correctly (*SD* = 5, range: 0–14) at the 2-week post. As with the prior comparisons, the number of words defined correctly in Dose 6 × Dose Frequency 6 versus Dose 4 × Dose Frequency 9 is relatively similar at 36 exposures (11 vs. 13), immediate post (6 vs. 4), and 2-week post (6 vs. 3).

Figure 2D shows the same two treatments but in the reverse order. Here, Dose 6 × Dose Frequency 6 results in children accurately defining (on average) 10 words correctly (*SD* = 7, range: 0–15) at 36 exposures, three words correctly (*SD* = 2, range: 1–7) at the immediate post, and two words correctly (*SD* = 1, range: 1–4) at the 2-week post. Turning to the second treatment, Dose 4 × Dose Frequency 9 results in children accurately defining (on average) 10 words correctly (*SD* = 5, range: 0–15) at the last treatment session, five words correctly (*SD* = 3, range: 2–11) at the immediate post, and three words correctly (*SD* = 2, range: 1–7) at the 2-week post. Again, the number of words defined correctly in Dose 6 × Dose Frequency 6 versus Dose 4 × Dose Frequency 9 is relatively similar at 36 exposures (10 vs. 10), immediate post (3 vs. 5), and 2-week post (2 vs. 3).

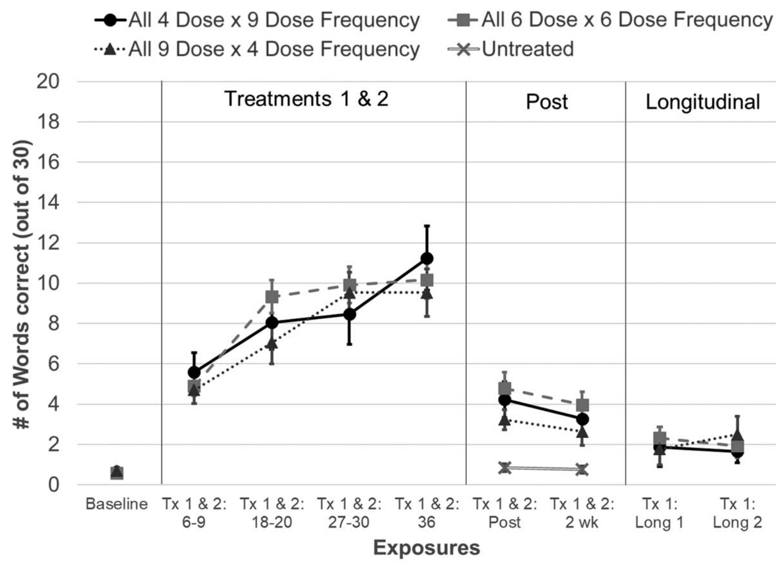
Figure 3 provides a final summary comparison of the Dose × Dose Frequency conditions, collapsing across Treatments 1 and 2. As shown here, the number of words correct is largely overlapping. Although it is possible that a larger sample would identify significant differences between Dose × Dose Frequency conditions, it seems unlikely that this would yield important real-world impact given the relatively small actual difference in number of words learned observed here (i.e., average differences were on the order of one to three words). Moreover, there is large variation in treatment response, with children correctly defining anywhere from 0 to 21 words at 36 exposures. These large differences in treatment response may have more real-world impact than the small differences associated with variation in Dose × Dose Frequency.

Variation in Treatment Response

In terms of variation in learning by pretreatment learner characteristics, the CELF Core Language composite was related to performance on Set 2 words. CELF Core Language scores were significantly related to intercept, $p = .001$, with the log of the words defined correctly increasing by .03 for every point higher on the CELF Core Language. Here, the intercept relates to the average number of Set 2 words defined correctly across all test points (pretreatment, immediate post–Treatment 1, 2-week post–Treatment 1, 36 exposures in the second treatment, immediate post–Treatment 2, 2-week post–Treatment 2). The left panel of Figure 4 shows that children with higher language scores defined more Set 2 words accurately (averaged across all test points) than children with lower language scores. In addition, CELF Core Language scores were related to slopes for Set 2 words during Treatment 2, $p = .02$. During Treatment 2, the slope of Set 2 words increased by 0.013 (*SE* = 0.006) for every point increase in the CELF Core Language score, indicating a steeper learning trajectory for children with better language skills. Figure 4 represents slope as gain scores, which are computed by subtracting the closest pretreatment test point from the immediate posttreatment test point. The center panel of Figure 4 shows steeper pre-to-post gains in the number of Set 2 words defined correctly for children with higher CELF Core Language scores than for children with lower CELF Core Language scores. The effect of the CELF Core Language on Set 1 words was similar during Treatment 1, with slopes increasing by 0.014 (*SE* = 0.007) for every point increase in the CELF Core Language score, but this just missed statistical significance, $p = .06$. The right panel of Figure 4 shows steeper pre-to-post gains in the number of Set 1 words defined correctly for children with higher CELF Core Language scores than for children with lower CELF Core Language scores.

CELF Understanding Spoken Paragraphs scores were related to the slope for Set 2 words during Treatment 2, with the slope for Set 2 words increasing by 0.08 (*SE* = 0.03), $p = .02$, for every point increase in CELF Understanding Spoken Paragraphs scores. As shown in the left panel of Figure 5, children with higher CELF Understanding Spoken Paragraph scores showed steeper pre-to-post gains in the number of Set 2 words defined correctly than children with lower CELF Understanding Spoken Paragraphs scores.

Figure 3. Mean number of words defined correctly in the Dose 4 × Dose Frequency 9 treatment (circles), the Dose 6 × Dose Frequency 6 treatment (squares), and the Dose 9 × Dose Frequency 4 treatment (triangles) prior to treatment (i.e., baseline), during treatment, and posttreatment. Data are collapsed across Treatment 1 and Treatment 2 and usually involve collapsing Set 1 and Set 2 words with two exceptions: (a) Untreated words are Set 2 words tested at baseline, immediately post–Treatment 1, and 2 weeks post–Treatment 1, and (b) longitudinal data are Set 1 words tested immediately post–Treatment 2 and 2 weeks post–Treatment 2. Error bars indicate standard errors.



Similarly, CELF Understanding Spoken Paragraph scores were related to slopes during Treatment 1 for Set 1 words, with slopes increasing by 0.07 (SE = 0.04) for every point increase in CELF Understanding Spoken Paragraphs scores, although the relationship was not significant for Set 1 words because of their larger standard error, p = .08. As shown in the right panel of Figure 5, children with higher CELF Understanding Spoken Paragraph scores showed steeper pre-to-post gains in the number of Set 1 words defined correctly than children with lower CELF Understanding Spoken Paragraphs scores.

Other possible predictors of learning were unrelated to intercepts or slopes. These included DELV Semantic adjusted standard scores, CTOPP Phonological Awareness composite scores, CTOPP Nonword Repetition standard scores, Goldman-Fristoe Test of Articulation scores, age, and gender.

Item Analysis

We had no a priori hypotheses about how item characteristics would relate to which words were learned, but we explored this post hoc. In terms of our prior work (Storkel,

Figure 4. The left panel shows the average number of Set 2 words defined correctly across all test points (pretreatment, immediate post–Treatment 1, 2-week post–Treatment 1, 36 exposures in the second treatment, immediate post–Treatment 2, 2-week post–Treatment 2) relative to Clinical Evaluation of Language Fundamentals (CELF) Core Language standard scores. The center panel shows gains in the number of Set 2 words accurately defined compared to CELF Core Language standard scores. Gain is computed as the number of Set 2 words accurately defined at 2 weeks post–Treatment 1 (i.e., right before the start of Treatment 2) subtracted from the number of Set 2 words accurately defined at immediate post–Treatment 2. The right panel shows gains in the number of Set 1 words accurately defined compared to CELF Core Language standard scores. Gain is computed as the number of Set 1 words accurately defined at pretreatment subtracted from the number of Set 1 words accurately defined at immediate post–Treatment 1. The line in all three panels is the linear fit line.

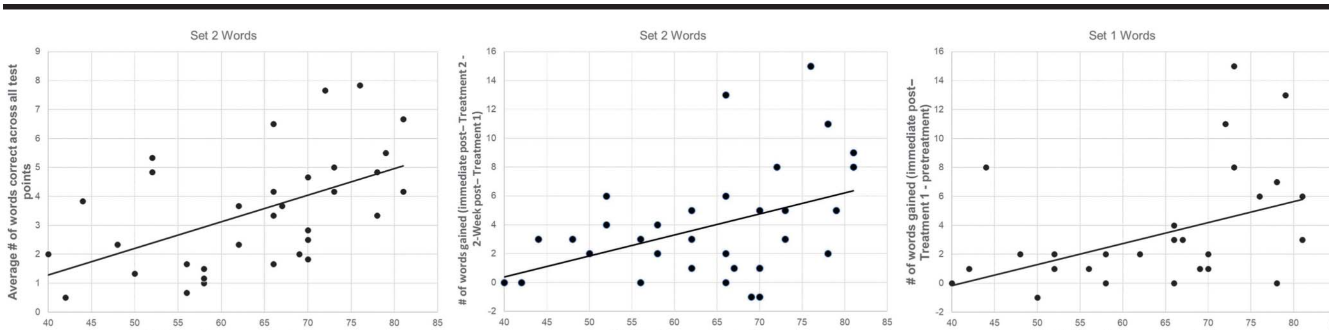
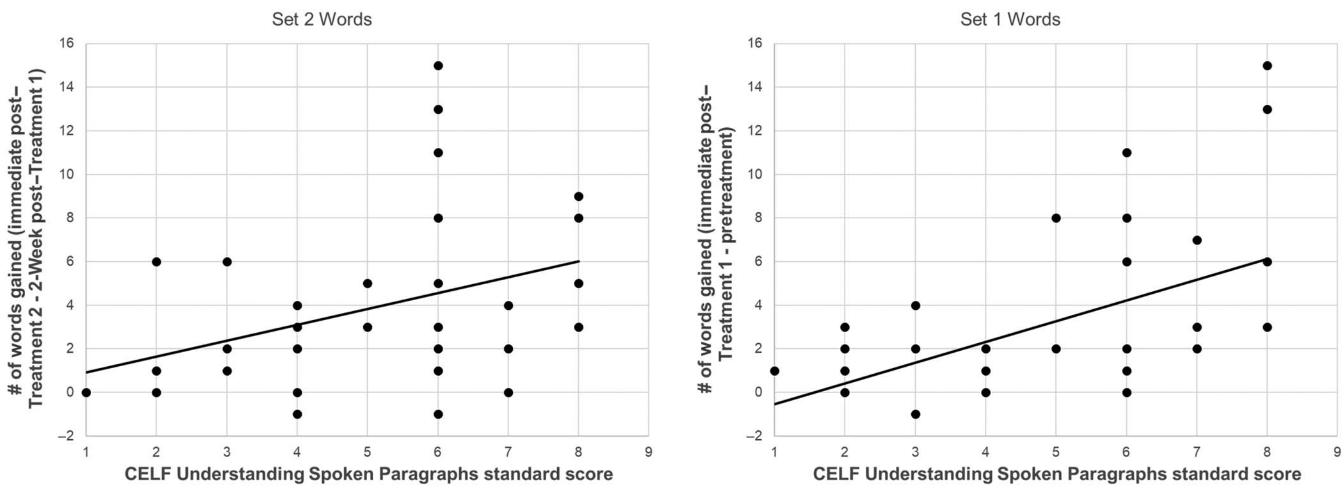


Figure 5. The left panel shows gains in the number of Set 2 words accurately defined compared to Clinical Evaluation of Language Fundamentals (CELF) Understanding Spoken Paragraphs standard scores. Gain is computed as the number of Set 2 words accurately defined at 2 weeks post–Treatment 1 (i.e., right before the start of Treatment 2) subtracted from the number of Set 2 words accurately defined at immediate post–Treatment 2. The right panel shows gains in the number of Set 1 words accurately defined compared to CELF Understanding Spoken Paragraphs standard scores. Gain is computed as the number of Set 1 words accurately defined at pretreatment subtracted from the number of Set 1 words accurately defined at immediate post–Treatment 1. The line in all three panels is the linear fit line.

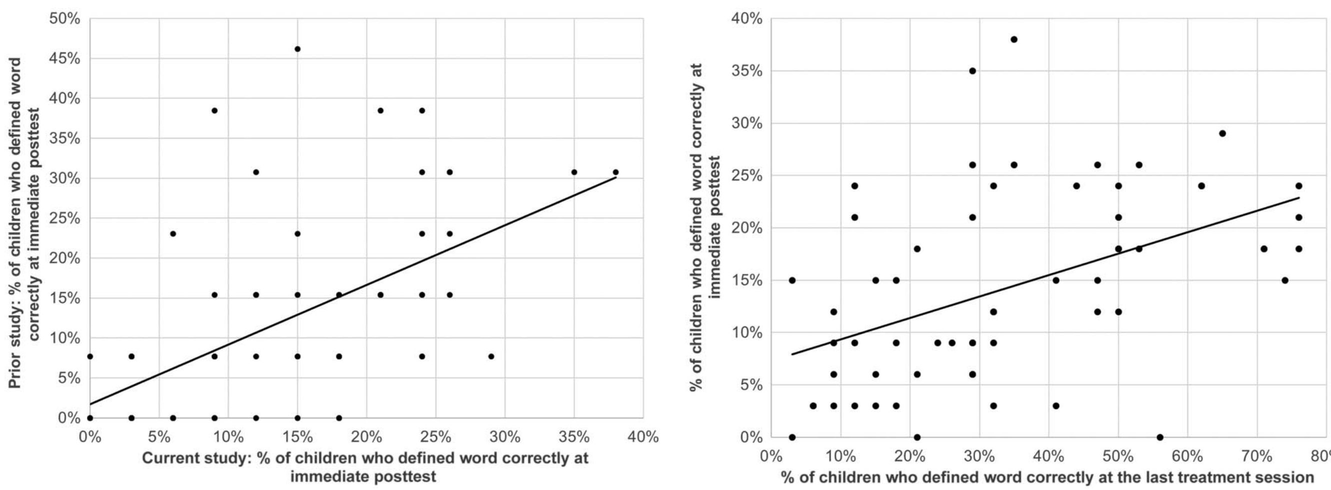


Voelmle, et al., 2017), we examined whether characteristics of the taught words (i.e., part of speech, word frequency) or characteristics of the exposure context (i.e., number of words in the definition used for training, average frequency of the words in the definition used for training, word frequency of the synonym used in training, number of words in the context sentence, average frequency of the words in the context sentence, number of words in the book text sentence, and average frequency of the word in the book text sentence) correlated with which words children learned (see Supplemental Table S2 in Storkel, Voelmle, et al., 2017). The percentage of children who learned a word in the prior preliminary clinical trial (Storkel, Voelmle, et al., 2017) was significantly correlated with data reported by Justice et al. (2005), who taught the same words via the same books (although with slightly different methods). Thus, there was some stability across studies in which words were frequently versus infrequently learned. However, none of the characteristics of words or exposure context was significantly correlated with learning in the prior preliminary clinical trial. In other words, there was no obvious explanation as to why certain words seemed to be easier to learn than others.

Raw data and results for the current data set are shown in Supplemental Materials S9 and S10, respectively. For the current data set, we once again began by examining stability. We found a significant correlation between the percentage of children who defined a word correctly at the immediate posttest in the current study and the percentage of children who defined a word correctly at the immediate posttest in our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017), $r(60) = .56$, $p < .001$, $r^2 = .31$. As shown in the left panel of Figure 6, as the percentage of children correctly defining a word at the immediate posttest in the current study increased, so too did the percentage of children correctly defining the word at the immediate posttest in the prior study. In addition, there was a trend in the same direction for a relationship between the percentage of children defining a word correctly at the final treatment session in the current study and the percentage of children correctly defining the word at the immediate posttest in the prior study, $r(60) = .23$, $p = .08$, $r^2 = .05$. Turning to relationships within the current study, the percentage of children who defined a word correctly at the last treatment session was significantly correlated with the percentage of children who defined the word correctly at the immediate posttest, $r(60) = .47$, $p < .001$, $r^2 = .22$. As shown in the right panel of Figure 6, as the percentage of children who defined a word correctly at the last treatment session increased, so too did the percentage of children who defined the word correctly at the immediate posttest. Thus, once again, the words that were frequently versus infrequently learned were somewhat similar across studies and over time within the current study.

Next, we explored correlations between the percentage of children who defined words correctly at the last treatment session or at the immediate posttest and various word or exposure characteristics. There was a significant relationship between the percentage of children who defined a word correctly at the last treatment session and the average frequency of the content words in the context sentence used in the review activity, $r(60) = -.26$, $p < .05$, $r^2 = .07$. However, this relationship did not match what one would expect and may have been driven by two extreme values. That is, as the average frequency of the words in the context sentence increased, the number of children correctly defining the words at the last treatment session decreased. This is counter to what one would predict: Higher frequency should ease

Figure 6. The left panel shows the percentage of children defining a word correctly at the immediate posttest in the prior preliminary clinical trial (Storkel, Voelmle, et al., 2017) compared to the percentage of children defining a word correctly at the immediate posttest in the current study. The right panel shows the percentage of children defining a word correctly at the last treatment session (in the current study) compared to the percentage of children defining a word correctly at the immediate posttest (in the current study). The line in both panels is the linear fit line.



processing and improve word learning. This relationship appeared to be driven by two outliers (*gathered, grumbling*), which had values that were 3.0 and 2.7 *SD*s above the mean of all items, respectively. All other items were within −1.2 and 1.6 *SD*s of the mean. When these two items were removed, the relationship was no longer significant, $r(58) = -.20$, $p > .10$, $r^2 = .04$.

The only other relationship that even approached significance was between the percentage of children who defined a word correctly at the last treatment session and the number of words in the exposure sentence in the book text, $r(60) = -.24$, $p = .06$, $r^2 = .06$. Specifically, as the number of words in the book exposure sentence increased, the number of children correctly defining the words at the last treatment session tended to decrease. This trend matches what one would expect: Longer sentences should be more difficult to process, hindering word learning.

Discussion

The goal of this research was to determine whether different combinations of dose and dose frequency lead to better word learning by children with DLD during interactive book reading and whether any pretreatment factors were significantly associated with the number of words learned. Results replicated the prior preliminary clinical trial (Storkel, Voelmle, et al., 2017) in showing that 36 exposures during interactive book reading supports significant word learning by children with DLD. However, the results also extend the prior work in demonstrating significant forgetting once treatment is withdrawn. Moreover, manipulation of dose and dose frequency did not lead to significant differences in word learning outcomes. As in the prior preliminary clinical trial, there was wide variation across children with DLD in the number of words learned during interactive book reading. The prior predictors of word learning variation (Storkel, Voelmle, et al., 2017) were not replicated. Instead, two language scores, CELF Core Language and CELF Understanding Spoken Paragraphs, were significantly associated with variation in word learning. Finally, item analysis showed consistency across studies in which words were frequently or infrequently learned by children, but few word characteristics or characteristics of the exposure context predicted this effect.

Replicating Effectiveness of 36 Exposures While Raising Concerns With Forgetting

Children with DLD showed significant slopes for Set 1 and Set 2 words during the treatment phase for each word set. However, the results also show a significant drop in performance for Set 1 and Set 2 words after treatment is withdrawn. The pattern is quite striking. Children (on average) correctly defined 10 words (range: 0–21) in their last treatment session (collapsing across the first and second treatments), and then the number of correctly defined words dropped sharply and significantly to four words (range: 0–16) at the immediate posttest and then dropped slightly further (albeit nonsignificantly) to three words (range: 0–14) at the 2-week posttest. In other words, approximately 40% of the words defined correctly at the end of treatment were defined accurately 5–6 days later, and 30% of the words defined correctly at the end of treatment were defined accurately 21 days later. These findings are consistent with the prior finding (Storkel, Voelmle, et al., 2017), which showed that children with DLD defined five words (range: 0–14) correctly at the immediate posttest following 36 exposures, which is a similar level of accuracy as in the current study. However, by measuring learning during treatment, it became clear that children with DLD learned more words during treatment than they retained posttreatment. Although there may be a need to alter this interactive book reading treatment to further improve the number of words learned during treatment (i.e., achieve closer to 30 words learned by the end of treatment), these results suggest that it may be more important to focus first on altering the treatment to better support retention of the words once treatment has ended. Retaining only a small percentage of learned words after the treatment has ended represents a major barrier to long-term improvement in vocabulary for children with DLD.

What are some of the possible explanations for why children with DLD forgot the words once treatment ended? Prior research and theory suggest that forgetting is potentially related to overall learning and decay rate (Ridgeway, Mozer, & Bowles, 2017). Specifically, stronger learning during training may index stronger, more durable memories that remain robust even when training is withdrawn (Ridgeway et al., 2017; Storkel, 2015), although others argue for the alternative (E. L. Bjork, 2004; R. A. Bjork, 2011; Wojcik, 2013). Applied to the current findings, one could infer that children with DLD learned the words during treatment but did not learn them well enough to buffer forgetting. It also is important to note that the test during treatment may be considered easier than the test posttreatment. Specifically, fewer words are tested during treatment (i.e., only the six that are the focus of the current treatment), and the testing occurs in the context of the training (i.e., picture stimuli and a familiar routine are present). This difference in the testing context also is consistent with the hypothesis that children learned the words well enough to respond accurately in a supportive context but not in a less supportive context. Taken together, these hypotheses suggest that we might need to think more about what it means to have a strong memory of a word and provide treatment until that strong memory has been established.

One avenue for defining a strong memory is to think about the requirements of the task being used to measure learning. A definition task requires a robust semantic representation, so on that front, children should have had a strong semantic representation of the correctly defined words at the end of treatment, although perhaps the supportiveness of the definition task also should be considered. A second avenue for defining a strong memory may be to think about the number of times a child achieves a correct response during treatment (Storkel, 2015). The end of our treatment was defined by the number of exposures rather than by a certain number of correct responses for a given word or a given word set. It is possible that children with DLD may need to establish a consistently correct response before treatment is withdrawn to support long-term retention. Turning to the concept of decay, a slower decay rate leads to more gradual forgetting (Ridgeway et al., 2017). Children with DLD have been shown to differ on measures of decay, showing faster decay rates (Nichols et al., 2004). If a fast decay rate is the issue, then it may be necessary to provide booster treatments after treatment has ended to reinstate learning and support long-term retention.

What could be done during treatment to create a strong memory? It is important to note that the current treatment is primarily receptive. That is, the child passively listens to the input provided by the research assistant. McKeown (2019) notes numerous avenues for increasing child engagement and depth of processing when teaching new vocabulary. These suggestions could easily be incorporated into our book reading protocol to facilitate more active engagement by the child, facilitating creation of strong memories.

Although these hypotheses for understanding forgetting by children with DLD and for improving the treatment effectiveness of interactive book reading require confirmation, the immediate implication for clinical practice is that word learning by children with DLD needs to be monitored beyond the end of treatment. A clinician may have evidence that a child has learned a word and thus switches treatment targets. Our data suggest that it is quite possible that, if those learned words were tested 5–6 days later in a less supportive task, the child with DLD may no longer show evidence of knowing the word. Consequently, clinicians may want to periodically monitor treatment targets after they are no longer the focus of treatment and may want to consider establishing a home or classroom program to provide regular encounters with target words once treatment has ended.

Dose × Dose Frequency Comparisons

In general, we did not observe strong effects from our manipulation of dose and dose frequency during treatment. This finding suggests that the overall cumulative number of exposures likely has greater impact on word learning during interactive book reading (Storkel, Voelmle, et al., 2017) than more fine grained aspects of how the cumulative number of exposures is achieved through different combinations of dose and dose frequency. This conclusion is consistent with a similar conclusion by Meyers-Denman and Plante (2016), who found no impact of dose schedule on treatment outcomes in a conversational recast treatment for children with DLD. Although laboratory studies show robust effects of massed versus distributed practice (e.g., Childers & Tomasello, 2002; McGregor et al., 2007; Riches et al., 2005), this does not appear to translate to clinical practice where less extreme combinations tend to be used. One way to marry these findings is to assume that some minimum dose and dose frequency are needed to achieve learning and that a dose frequency of 1 (as in massed practice) may not meet the minimum threshold. However, once the minimum dose and dose frequency are achieved, specific combinations are less influential on learning. Furthermore, our minimum of 4 for both dose and dose frequency seems to be at or above threshold, but further exploration of lower doses and dose frequencies is needed to determine the exact threshold.

Another difference between the results of the current study and those of prior studies is that prior research has suggested that children with DLD may experience greater difficulty with encoding than memory consolidation. The current study suggests the opposite. Here, children with

DLD appeared to show strong encoding but poor retention potentially due to poor memory consolidation. There are numerous differences between the current study and prior research. For example, Storkel, Voelmle, et al. (2017) varied the intensity so not all children received the adequate intensity. Likewise, McGregor, Licandro, et al. (2013) provided a set number of exposures, rather than a number of exposures that was empirically determined to be adequate. It is possible that difficulty with encoding is observed mainly when the overall number of exposures is inadequate. That is, when encoding has been adequately supported by providing sufficient exposures, difficulties with memory consolidation may be revealed.

A further conclusion is that the range of doses and dose frequencies tested provided relatively similar support for encoding and memory consolidation by children with DLD because differences in Dose × Dose Frequency did not translate into differences in word learning. It seems that whatever support is gained for encoding and memory consolidation through simple repetition of information is similar within the range of doses and dose frequencies studied here. Consequently, encoding and/or memory consolidation may need to be supported in a way that goes beyond simple repetition so that children with DLD can learn and remember more words from treatment. Again, McKeown's (2019) tutorial provides a review of vocabulary instruction strategies that promote active engagement and the formation of robust memories.

In terms of application to clinical practice, these findings suggest that clinicians can select whichever Dose × Dose Frequency combination best aligns with practical considerations. As shown in Table 3, the different combinations of Dose × Dose Frequency are associated with different session durations (13–16 min) and different numbers of sessions (10–23 sessions) but do not appear to lead to substantially different learning. There may be scenarios where shorter sessions are ideal, such as when a child who may have difficulty focusing for longer periods or when scheduling constraints and caseload size within a school setting work against longer sessions. In this case, Dose 4 × Dose Frequency 9, which (on average) requires 13-min sessions, may be the best match to practical considerations. In complement, there may be situations when fewer sessions are ideal, such as when a family is paying for each session or when the number of reimbursed sessions is capped. In this case, Dose 9 × Dose Frequency 4, which equates to 10 sessions, may be optimal.

Variation in Treatment Response

Our prior preliminary clinical trial (Storkel, Voelmle, et al., 2017) also noted wide variation in the number of words learned by children with DLD at the end of interactive book reading treatment. This variation in outcomes was associated with variation in scores on the DELV Semantic subtest, the CTOPP Nonword Repetition, and the CTOPP Phonological Awareness. None of these test scores was replicated as significant predictors of treatment outcomes in this study. In our prior study, we assumed that these tests indexed semantic and phonological processing and that both types of processing were crucial to word learning. There are several limitations in the current study that affect the ability to draw conclusions about why these differences arose. One important difference between the two studies is that the prior study tested a variety of treatment intensities, whereas the current study tested only the adequate intensity of 36 exposures. It may be that the adequate intensity helped children overcome phonological and semantic processing difficulties, minimizing the contribution of these skills to word learning during interactive book reading. A second important difference between the two studies is the method of analysis. The prior study examined the relationship between pretreatment characteristics and performance at the immediate posttest. The current study used a more nuanced approach by examining how pretreatment characteristics related to intercept and slope in a multilevel model, which index average performance across time and change in performance across time, respectively. Thus, the current analysis captures dynamic learning rather than a static outcome of learning.

Overall language skills, as indexed by the CELF Core Language score and the CELF Understanding Spoken Paragraphs score, were related to treatment outcomes in the current study but not in the prior study (Storkel, Voelmle, et al., 2017). It is unclear why the predictors of word learning change across studies, but that also has been observed in other studies (Alt & Plante, 2006; Blewitt et al., 2009; Elley, 1989; Gray, 2006; Hansson et al., 2004; Jackson et al., 2016; Justice et al., 2005; Penno et al., 2002; Robbins & Ehri, 1994). It is possible that differences in findings relate to methodological differences, as previously noted for this study. However, it also could be that differences in findings across studies relate to how the treatment supports the child's strengths and weaknesses. As noted previously, the adequate intensity of 36 exposures may have better supported phonological and semantic processing, allowing a different weakness to emerge as a predictor of treatment outcomes. Specifically, weaknesses in language may now be the limiting factor because these weaknesses index how well the child understands the story context and the other grammatical relationships that help to establish the meaning of the words. This hypothesis also is consistent with the nonsignificant trend in the item analysis that children tended to be less successful learning words contained in longer sentences than words contained in shorter sentences in the book. This finding also seems to suggest that language skills may be limiting learning. Taken together, children who have stronger language skills may be better able to take advantage of the treatment. Either children with weaker general language skills need a different treatment or the interactive book reading treatment needs to be altered to better support deficits in these areas.

Conclusions

This study provided further evidence that 36 exposures to target words during interactive book reading is an adequate intensity to support word learning by children with DLD. Moreover, the study showed that variations in the dose and dose frequency to achieve 36 exposures did not have a significant impact on the number of words learned by children with DLD during interactive book reading. Thus, clinicians can choose a dose and dose frequency that best match practical considerations. The current version of interactive book reading appeared to sufficiently support phonological and semantic processing but may not have been sufficient to support variation in general language skills, indicating that further development of the treatment is needed to better serve all children with DLD. A final critical finding is that children with DLD showed strong learning during treatment but poor retention after treatment ended, indicating that retention of newly learned words needs to be monitored posttreatment. This finding also indicates the need to continue to develop interactive book reading treatment to better support long-term retention of newly learned words posttreatment. In particular, adding more interactive elements to the treatment may be a promising avenue to pursue. Although interactive book reading continues to show promise as an effective word learning intervention for children with DLD, further development is needed to enhance the effectiveness of this treatment approach, with a specific focus on improving long-term retention of newly learned words and improving treatment response across children with DLD and different language profiles.

Acknowledgments

This research was supported by National Institutes of Health Grants DC012824 to Holly Storkel (supported all aspects of the research) and HD090216 to John Colombo (provided infrastructure for research design and analysis) and NCATS-UL1TR002366 to Richard Barohn (provided infrastructure for data management via Research Electronic Data Capture).

References

Adi-Japha, E., & Abu-Asba, H. (2014). Learning, forgetting, and relearning: Skill learning in children with language impairment. *American Journal of Speech-Language Pathology, 23,* 696–707. https://doi.org/10.1044/2014_AJSLP-13-0031

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(5), 941–954. https://doi.org/10.1044/1092-4388(2006/068)

American Speech-Language-Hearing Association. (1997). Guidelines for screening for hearing impairment-school-age children, 5 through 18 years. *American-Speech-Languge-Hearing Association, 4,* IV-74ff–IV-74hh.

Bishop, D. V. M. (2004). Specific language impairment: Diagnostic dilemmas. In L. Verhoeven & H. van Balkom (Eds.), *Classification of developmental language disorders: Theoretical issues and clinical implications* (pp. 309–326). Mahwah, NJ: Erlbaum.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS ONE, 11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Bjork, E. L. (2004). *Research on learning as a foundation for curricular reform and pedagogy*. Paper presented at the Reinvention Center Conference, Washington, DC.

Bjork, R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 1–22). New York, NY: Psychology Press.

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology, 101*(2), 294–304. https://doi.org/10.1037/A0013844

Brown, H., Weighall, A., Henderson, L. M., & Gareth Gaskell, M. (2012). Enhanced recognition and recall of new words in 7- and 12-year-olds following a period of offline consolidation. *Journal of Experimental Child Psychology, 112*(1), 56–72. https://doi.org/10.1016/j.jecp.2011.11.010

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Cirrin, F. M., & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools, 39*(1), S110–S137. https://doi.org/10.1044/0161-1461(2008/012)

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793–795. https://doi.org/10.1056/NEJMp1500523

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803–820. https://doi.org/10.1162/jocn.2009.21059

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London, Series B, Biological Science, 364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Dempster, F. (1996). Distributing and managing conditions of encoding and practice. In E. Bjork & R. Bjork (Eds.), *Handbook of perception and cognition* (2nd ed., pp. 317–344). New York, NY: Academic Press.

Elley, W. B. (1989). Vocabulary acquisition form listening to stories. *Reading Research Quarterly, 24,* 174–187. https://doi.org/10.2307/747863

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2)*. Circles Pines, MN: AGS.

Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research, 46*(1), 56–67. https://doi.org/10.1044/1092-4388(2003/005)

Gray, S. (2006). The relationship between phonological memory, receptive vocabulary, and fast mapping in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*(5), 955–969. https://doi.org/10.1044/1092-4388(2006/069)

Hansson, K., Forsberg, J., Löfqvist, A., Mäki-Torkko, E., & Sahlén, B. (2004). Working memory and novel word learning in children with hearing impairment and children with specific language impairment. *International Journal of Language & Communication Disorders, 39*(3), 401–422. https://doi.org/10.1080/13682820410001669887

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research Electronic Data Capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42,* 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Henderson, L. M., Weighall, A. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674–687. https://doi.org/10.1111/j.1467-7687.2012.01172.x

Henderson, L. M., Weighall, A. R., & Gaskell, M. G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology, 116*(3), 572–592. https://doi.org/10.1016/j.jecp.2013.07.004

Hunsberger, S., Rubinstein, L. V., Dancey, J., & Korn, E. L. (2005). Dose escalation trial designs based on a molecularly targeted endpoint. *Statistics in Medicine, 24*(14), 2171–2181. https://doi.org/10.1002/sim.2102

Jackson, E., Leitao, S., & Claessen, M. (2016). The relationship between phonological short-term memory, receptive vocabulary, and fast mapping in children with specific language impairment. *International Journal of Language & Communication Disorders, 51*(1), 61–73. https://doi.org/10.1111/1460-6984.12185

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17–32. https://doi.org/10.1044/0161-1461(2005/003)

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. https://doi.org/10.3102/0034654310377087

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419–457. https://doi.org/10.1037/0033-295x.102.3.419

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research, 56*(6), 1845–1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders, 48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

McGregor, K. K., Sheng, L., & Ball, T. (2007). Complexities of expressive word learning over time. *Language, Speech, and Hearing Services in Schools, 38,* 353–364. https://doi.org/10.1044/0161-1461(2007/037)

McKeown, M. G. (2019). Effective vocabulary instruction fosters knowing words, using words, and understanding how words work. *Language, Speech, and Hearing Serivces in Schools, 50,* 466–476. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0126

Meyers-Denman, C. N., & Plante, E. (2016). Dose schedule and enhanced conversational recast treatment for children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 47*(4), 334–346. https://doi.org/10.1044/2016_LSHSS-15-0064

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*(2), 979–1007. https://doi.org/10.3102/0034654309332561

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added valuate of dialogic parent–child book readings: A meta-analysis. *Early Education & Development, 19*(1), 7–26. https://doi.org/10.1080/10409280701838603

Nichols, S., Jones, W., Roman, M. J., Wulfeck, B., Delis, D. C., Reilly, J., & Bellugi, U. (2004). Mechanisms of verbal memory impairment in four neurodevelopmental disorders. *Brain and Language, 88*(2), 180–189. https://doi.org/10.1016/S0093-934X(03)00097-X

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110*(4), 611–646. https://doi.org/10.1037/0033-295x.110.4.611

Oetting, J. B. (1999). Children with SLI use argument structure cues to learn verbs. *Journal of Speech, Language, and Hearing Research, 42,* 1261–1274. https://doi.org/10.1044/jslhr.4205.1261

O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus, 10*(4), 389–397. https://doi.org/10.1002/1098-1063(2000)10:4<389::aid-hipo5>3.0.co;2-p

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23–33. https://doi.org/10.1037//0022-0663.94.1.23

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales (RIAS)*. Lutz, FL: Psychological Assessment Resources.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37*(1), 106–121. https://doi.org/10.1044/jshr.3701.106

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research, 48*(6), 1397–1411. https://doi.org/10.1044/1092-4388(2005/097)

Ridgeway, K., Mozer, M. C., & Bowles, A. R. (2017). Forgetting of foreign-language skills: A corpus-based analysis of online tutoring software. *Cognitive Science, 41,* 924–949. https://doi.org/10.1111/cogs.12385

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86*(1), 54–64. https://doi.org/10.1037/0022-0663.86.1.54

Schwartz, R. G., & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language, 10*(1), 57–64. https://doi.org/10.1017/S0305000900005134

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Editiona (CELF-4)*. Bloomington, MN: Pearson.

Seymour, H. N., Roeper, T. W., de Villers, J., & de Villers, P. A. (2005). *Diagnostic Evaluation of Language Variation–Norm Referenced (DELV-Norm Referenced)*. San Antonio, TX: The Psychological Corporation.

Small, D. (1985). *Imogene's antlers.* New York, NY:Crown Publishers.

Storkel, H. L. (2015). Learning from input and memory evolution: Points of vulnerability on a pathway to mastery in word learning. *International Journal of Speech-Language Pathology, 17,* 1–12. https://doi.org/10.3109/17549507.2014.987818

Storkel, H. L. (2016). *Interactive book reading to accelerate word learning by kindergarten children with specific language impairment (SLI)*. Retrieved from http://hdl.handle.net/1808/20313

Storkel, H. L., Komesidou, R., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying adequate progress and successful learning patterns. *Language, Speech, and Hearing Services in Schools, 48,* 108–124. https://doi.org/10.1044/2017_LSHSS-16-0058

Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech, and Hearing Services in Schools, 48,* 16–30. https://doi.org/10.1044/2016_LSHSS-16-0014

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology, 66*(5), 1001–1025. https://doi.org/10.1080/17470218.2012.724694

Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review, 68,* 229–247.

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology, 3,* 46. https://doi.org/10.3389/fpsyg.2012.00046

Wagner, R. K., Torgesen, J., Rashotte, C., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2)*. Austin, TX: Pro-Ed.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 70–77. https://doi.org/10.1002/mrdd.20139

Wojcik, E. H. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in Psychology, 4*, 151. https://doi.org/10.3389/fpsyg.2013.00151