JSLHR

Research Article

Adjective Learning in Young Typically Developing Children and Children With Developmental Language Disorder: A Retrieval-Based Approach

Laurence B. Leonard,[a] Patricia Deevy,[a] Jeffrey D. Karpicke,[b] Sharon Christ,[c] Christine Weber,[a] Justin B. Kueser,[a] and Eileen Haebig[d]

Purpose: There are strong retention benefits when learners frequently test themselves during the learning period. This practice of repeated retrieval has recently been applied successfully to children's word learning. In this study, we apply a repeated retrieval procedure to the learning of novel adjectives by preschool-age children with developmental language disorder (DLD) and their typically developing (TD) peers. We ask whether the benefits of retrieval extend to children's ability to apply the novel adjectives to newly introduced objects sharing the same characteristics as the objects used during the learning period.
Method: Fourteen children with DLD (M_{age} = 62.64 months) and 13 TD children (M = 62.54 months) learned novel adjectives in 2 sessions. For each child, half of the adjectives were learned in a repeated spaced retrieval condition, and half were learned in a repeated study–only condition. Recall was assessed immediately after the second learning session and 1 week later. A recognition test was also administered at the 1-week mark.
Results: On the recall tests, for both groups of children, recall was better for adjectives learned in the repeated spaced retrieval condition. Adjectives learned by the 2nd day were retained 1 week later. Every adjective correctly applied to an object used during the learning period was also extended accurately to new objects with the same characteristics. On these recall tests, the children with DLD did not differ from the TD group in the number of items recalled, though their phonetic accuracy was lower. On the recognition test, the DLD group showed greater accuracy for adjectives that had been learned in the repeated spaced retrieval condition than for those learned in the repeated study condition, whereas the TD group performed at high levels in both conditions.
Conclusion: Repeated spaced retrieval appears to provide an effective boost to word learning. Because its benefits are seen even when a word must be extended to new objects, the application of this procedure seems well suited for learning new language material rather than being limited to item-specific memorization.

When we test our memory of some detail that we have learned, we are not just seeing if we remember it correctly. It is more than a test. The act of retrieving something from memory actually alters learning. Although this basic fact has been part of the scientific literature for decades, only recently has its contribution to daily learning activities been the focus of systematic research (Karpicke, 2017).

In this study, we examine the role of retrieval in facilitating children's word learning. Our focus is on children with developmental language disorder (DLD). These children —frequently referred to as *children with specific language impairment*—show a significant deficit in language ability that cannot be explained by any subtle deficits that may also be present in cognitive, motor, or sensory areas. Although diagnosed during childhood, DLD is a long-standing

[a]Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN
[b]Department of Psychological Sciences, Purdue University, West Lafayette, IN
[c]Department of Human Development and Family Studies, Purdue University, West Lafayette, IN
[d]Department of Communication Sciences & Disorders, Louisiana State University, Baton Rouge

Correspondence to Laurence B. Leonard: xdxl@purdue.edu

Editor-in-Chief: Sean Redmond
Editor: Filip Smolík

Received May 20, 2019
Revision received July 26, 2019
Accepted August 13, 2019
https://doi.org/10.1044/2019_JSLHR-L-19-0221

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

condition. Its chief symptoms change with development and through intervention. Nevertheless, even in adulthood, individuals diagnosed with DLD at a young age will usually show signs of continued weakness in language (Law, Tomblin, & Zhang, 2008; Rice et al., 2010; Tomblin, Zhang, Buckwalter, & O'Brien, 2003).

Among the lifelong language weaknesses in DLD is the learning of new words; in fact, individuals with DLD fall further behind their peers in vocabulary skill over time, even into adulthood (Rice & Hoffman, 2015). Even when children with DLD earn age-appropriate scores on standardized tests of vocabulary, close inspection reveals significantly weaker vocabulary skills than their typically developing (TD) peers (e.g., Leonard et al., 2019; McGregor et al., 2012). These weaknesses are seen not only in a more limited vocabulary size but also in a shallower knowledge of the words that are already in the children's lexical inventories (Dollaghan, 1998; Kail & Leonard, 1986; Leonard, Nippold, Kail, & Hale, 1983; McGregor, Newman, Reilly, & Capone, 2002; McGregor, Oleson, Bahnsen, & Duff, 2013; McGregor & Waxman, 1998). This difference is seen in both children and young adults.

When children and adults with DLD are presented with novel words to learn, they require more exposures to meet the same learning criterion as their age mates (Alt, 2011; Alt & Plante, 2006; Alt, Plante, & Creusere, 2004; Gray, 2003, 2004; Gray, Pittman, & Weinhold, 2014; Leonard et al., 1982; McGregor, Licandro, et al., 2013; Rice, Oetting, Marquis, Bode, & Pae, 1994). Limitations are seen across lexical categories; nouns, verbs, and adjectives are all learned with difficulty (Kan & Windsor, 2010; Oetting, Rice, & Swank, 1995; Rice, Buhr, & Nemeth, 1990; Skipp, Windfuhr, & Conti-Ramsden, 2002; Storkel, Voelmie, et al., 2017; Windfuhr, Faragher, & Conti-Ramsden, 2002). Slower, less effective word learning is seen on both measures of comprehension and measures of production (Kan & Windsor, 2010).

The impetus for applying retrieval activities to assist word learning in children with DLD has come from a resurgence of retrieval-based research in the field of psychology. Collectively, this work has demonstrated how learners' long-term retention of material is bolstered by inserting instances of self-testing during the period of study (see reviews in Karpicke, 2017; Rowland, 2014). Relative to the retention seen through studying alone, retrieval-infused study produces gains that are 50%–150% greater. Studies have varied widely in the material that participants have been asked to learn. This material has ranged from science concepts to the English translations of words in Swahili. Participants have most often been adults, but recent work has focused on school-age children learning academically relevant material (Fazio & Marsh, 2019).

What makes retrieval so effective? One facilitative element seems to be the multiple opportunities for retrieval that are provided in these studies. Retrieval trials that are introduced relatively frequently produce greater retention (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Another important element is the spacing of retrieval attempts. Although there are clear benefits when retrieval occurs immediately after an item has been studied, the benefits are still greater when there is spacing between a study trial and the following retrieval trial (Karpicke & Roediger, 2007). Spacing is often created by inserting other items between a target item and the attempt to retrieve that item. Effective spacing will often result in errors early in the learning phase but in increasing success as trials proceed. Previous studies have compared the effects of different retrieval schedules in adults and children. With adult participants, both equal spacing and expanding spacing schedules appear to be effective. For studies with children, extra steps are often included during the initial retrieval trials to promote early success, such as allowing for immediate retrieval, providing retrieval hints, and/or giving feedback (Fazio & Marsh, 2019).

The benefits of repeated spaced retrieval are often attributed to its "effortful" nature. Although *effortful* is an apt term, it is primarily descriptive. Attempts to explain why repeated spaced retrieval is helpful are quite limited (see Carpenter, 2009, and Karpicke, Lehman, & Aue, 2014, for two prominent examples). In the current study, we discuss retrieval effects using the framework of the "episodic context" account of Karpicke and his colleagues (Karpicke, 2017; Karpicke et al., 2014; Karpicke & Zaromb, 2010; Lehman, Smith, & Karpicke, 2014). Although the purpose of our study was to assess the effects of repeated spaced retrieval on word learning rather than to test particular theories of retrieval, it was helpful to use a theoretical framework to specify some of the mechanisms that might be at work when children are asked to retrieve new words.

The episodic context account builds on formal memory models that assume that features of the context become associated with items during encoding (Howard & Kahana, 2002). It is further assumed that, when items are retrieved, the prior episodic context is reinstated (Lehman & Malmberg, 2013). The features of the past context are then combined with those of the present context so that the updated context representation includes a composite of features from the earlier and later contexts. With repeated retrieval, the context representation is again updated, creating a representation reflecting multiple contexts. The increasingly unique composite of features across retrieval trials enables future retrieval to become more targeted, because the search set will be restricted to items that share a greater number of contextual features.

These details suggest how repeated retrieval might be beneficial. This account also provides an explanation for why spaced repeated retrieval is optimal. When there are intervening items between a target item's previous study trial and the time of retrieval, the context changes more than when the retrieval trial immediately follows the study trial. This renders the reinstated prior context more likely to be different from the present context, making the composite of earlier and later contextual features less similar to that of other items.

It should be noted that the assumption of contextual reinstatement is made even when experimental procedures are highly structured, with only subtle changes in the episodic/temporal context when moving from one item to the next. The process of associating features of the context with items during encoding is largely a process that is internal to the learner. These experiments do not involve instructions to pay attention to context. However, the plausibility of contextual effects becomes especially clear with recent studies that have overtly prompted learners to consider context. For example, Whiffen and Karpicke (2017) had participants learn two short lists of words (separated by a distractor task). Next, the participants were presented with the words from the two lists mixed together. During this phase, half of the participants were asked to determine whether each word came from the first list or the second list. In other respects, the procedure was the same for the two groups of participants. When recall of the words was then tested, the participant group asked to consider each word's set membership showed greater recall than the other group of participants. It is important to point out that these participants were told to make judgments about set membership only after the words were mixed together. Thus, although the participants were eventually asked about set membership, to be successful, they must have already registered details of the word's context during the encoding process. Evidently, one of those contextual details was the set to which the word belonged.

Contextual features are not considered to be inherent features of a word comparable to a word's semantic or phonological features any more than contextual features become part of a science concept that is being learned. In fact, contextual reinstatement has been assumed in the learning of a wide range of material that cannot be readily described verbally (see the review in Karpicke, 2017). Of course, to the extent that contextual reinstatement through repeated retrieval can ensure a new word's presence and maintenance in a child's vocabulary, that word can subsequently benefit from further elaboration of semantic or phonological details. However, contextual reinstatement does not itself provide the semantic or phonological information.

In this study, we employed a repeated spaced retrieval condition and compared children's retention of novel words in this condition to the children's retention when they studied words only, without retrieval. We refer to these conditions as the repeated retrieval with contextual reinstatement (RRCR) condition and the repeated study (RS) condition, respectively. We expected greater retention in the RRCR condition. In conditions involving continuous study only (as in the RS condition), the lack of retrieval provides no opportunity for contextual reinstatement. Consequently, when at some point the material must finally be recalled, there is no distinctive composite of contextual features to assist in constraining the search.

There have been several studies of retrieval effects on word learning in children or young adults with DLD (Chen & Liu, 2014; Haebig et al., 2019; Leonard et al., 2019; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017). Methods and retrieval schedules have differed across studies, but each investigation has reported advantages for retrieval over the comparison conditions. In the study with the design most similar to this study, Leonard et al. (2019) asked 5-year-olds with DLD and their same-age TD peers to learn a set of nonwords (hereafter referred to as *novel words*) referring to exotic plants and animals. For each child, half of the novel words were learned in an RRCR condition, and half were learned in an RS condition. All words were heard 48 times over 2 days. The two conditions differed only in whether retrieval opportunities were also provided. At the end of the second day and 1 week later, the children's retention of the novel words was tested. For both groups of children, retention scores for the words in the RRCR condition more than doubled the retention scores seen for words in the RS condition. Scores 1 week later were no different from those earned immediately after the learning period.

Haebig et al. (2019) conducted a similar study with novel words representing exotic plants and animals. Instead of asking if RRCR was superior to RS, Haebig et al. asked whether the spacing of retrieval trials (as occurs in RRCR) was more beneficial than the retrieval trials that immediately followed study trials with no intervening items. As in the Leonard et al. (2019) study, large differences in recall favoring the RRCR condition were found. This difference was also apparent when measured at a neural level via event-related potentials. After the learning period, when children heard the novel words paired with either the correct picture or a mismatched picture, they showed a stronger neural response (an N400) to the mismatch if the novel word had been learned in the RRCR condition. Together, the Leonard et al. (2019) and Haebig et al. findings suggest that both the repeated opportunity for retrieval and the spacing of retrieval assist children's learning and retention of new words.

Although the findings of these studies are very encouraging, they constitute only a first step. The referents in these studies were plants and animals and, as whole objects, are easily individuated (Gentner, 1982; Gentner & Boroditsky, 2001). As a result, the link from phonological word form to meaning was relatively transparent for learning these novel words. The phonological forms had to be discerned, learned, and retained, but the referents for these forms were probably never in doubt.

For a lexical category such as adjectives, the mapping is not as straightforward. Adjectives are properties that can vary within a basic noun category (e.g., a red ball, a blue ball, a green ball) yet operate across basic noun categories (e.g., a red ball, a red house, a red dress). It seems that young children need to understand the connection between nouns and object categories before they can establish a connection between adjectives and the properties of objects (e.g., Booth & Waxman, 2009). Successful learners must infer that adjectives are not the names of the objects but rather refer to something about the objects. For preschoolers with DLD faced with learning new adjectives, this insight may not come easily.

According to Mintz and Gleitman (2002), there are two provisions that can make this insight more attainable for young children. The first is to illustrate the attribute using multiple exemplars across basic noun categories (e.g., a ball, a horse, and a boat). The second is to use the adjective with the actual name of the object (e.g., a "stoof" dog) rather than with a more general term (e.g., a "stoof" one). By incorporating these provisions into their procedures, Mintz and Gleitman found that TD children as young as 24 months of age learned to associate newly learned adjectives with objects that had never been seen before, yet shared the same attribute.

In this study, we examine nonword (hereafter, novel) adjective learning by preschool-age children with DLD and their same-age TD peers. Adjectives are among the words that are difficult for these children, though, in previous studies of word learning, they have usually been included with nouns and verbs in a single target vocabulary list (e.g., Oetting et al., 1995; Storkel, Komesidou, Fleming, & Romine, 2017). Here, we focus exclusively on adjectives and adopt the key guidelines followed by Mintz and Gleitman (2002) to highlight the fact that the novel words refer to attributes, not the objects themselves. Of special interest is whether RRCR is more effective than RS in promoting both the learning of the adjective names and their generalization to new objects with the same attribute. Previous studies have examined how retrieval promotes "transfer" of learning to new material (Karpicke, 2017). However, to our knowledge, this is the first effort to determine whether retrieval does more than assist the learning of new words; here, we ask if words can be generalized to new referents—a communicatively important process. Adjectives appear to be an ideal vehicle through which to examine this question.

Method

Participants

The first author's institutional review board approved the research described here. Parents gave informed written consent, and verbal assent was provided by the children. Twenty-seven children participated in the study: 14 children who met the criteria for DLD and 13 children who demonstrated typical language development. The children with DLD (eight boys, six girls) ranged in age from 53 to 71 months (M = 62.64, SD = 5.41). These children were already enrolled in language intervention at the time of study or had been referred by speech-language pathologists and/or parents for possible intervention. The 13 children in the TD group (seven boys, six girls) were similar in age to the children in the DLD group, $t(25) = 0.05$, $p = .96$. Each child in the TD group was within 2 months of age of a child in the DLD group. The TD group ranged in age from 51 to 71 months (M = 62.54, SD = 6.34). The mothers of the two groups differed in years of education, $t(25) = 2.54$, $p = .018$. The mothers of the children with DLD averaged 14.79 years of education (SD = 2.19), whereas the mothers of the TD children averaged 16.69 years of schooling (SD = 1.65).

A two-step process was employed to select children for the DLD group. First, the children were administered the Structured Photographic Expressive Language Test–Preschool 2 (SPELT-P2; Dawson et al., 2005). Children were included in the DLD participant group if they scored at 87 or below, 87 being the empirically derived cutoff reflecting good sensitivity and specificity (Greenslade, Plante, & Vance, 2009). For children whose SPELT-P2 scores were between 87 and 90, the children's spontaneous language samples were scored using developmental sentence scoring (DSS; Lee, 1974) and the finite verb morphology composite (FVMC; Goffman & Leonard, 2000). These children were included in the DLD participant group if they scored below the 10th percentile on DSS and below the cutoff for acceptable sensitivity and specificity on the FVMC reported by Gladfelter and Leonard (2013) and Souto, Leonard, and Deevy (2014). The DSS and FVMC criteria were viewed as an important additional check given that the children were viewed as at risk for DLD and were either enrolled in or referred for intervention. Eleven of the 14 children with DLD met the SPELT-P2 criterion (for N = 11, M = 73.64, SD = 16.37; for N = 14, M = 76.93, SD = 15.78), and the remaining children met the criteria for DSS and the FVMC. All children with DLD scored in the "nonautistic" range on the Childhood Autism Rating Scale–Second Edition (Schopler, Van Bourgondien, Wellman, & Love, 2010) and passed a hearing screening. These children scored above 80 (M = 99.21, SD = 12.88) on the Kaufman Assessment Battery for Children, Second Edition (K-ABC2; Kaufman & Kaufman, 2004), an estimate of nonverbal intelligence. No parents reported a history of neurological impairment for their child.

As expected, the children in the TD group scored within or above normal limits on the SPELT-P2 (M = 119.00, SD = 8.03) and the K-ABC2 (M = 114.31, SD = 11.06). These children's K-ABC2 scores were significantly higher than the corresponding scores for the DLD group, $t(25) = 3.25$, $p = .003$. The children in the TD group were not administered the Childhood Autism Rating Scale–Second Edition. No history of neurological impairment was reported for these children, and all passed a hearing screening.

The standardized tests of vocabulary that are prominent in the field have not demonstrated adequate diagnostic accuracy (Gray, Plante, Vance, & Henrichsen, 1999; Spaulding, Hosmer, & Schechtman, 2013). For this reason, we did not employ them as part of our selection criteria. However, to obtain an estimate of the children's receptive vocabulary, we administered the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007). Most children in the DLD group scored within an average range on this test (M = 102.57, SD = 11.33), though significantly lower than the TD group (M = 118.62, SD = 13.62), $t(25) = 3.34$, $p = .003$. For an estimate of expressive vocabulary, we administered the Expressive Vocabulary Test–Second Edition (EVT-2; Williams, 2007) to all children. Results were similar to those for the PPVT-4; the children with DLD scored within the average range (M = 100.79, SD = 9.25) but significantly below the level of the TD group (M = 115.92, SD = 8.16), $t(25) = 4.50$, $p = .001$.

Finally, the children were administered an 18-item short-sentence repetition task in which the final words in the sentences reflected the syllable structures and phonetic content of the novel words used as adjectives. For example, the items "My coat has a *zipper,*" "Use the *soap,*" and "Pet the *doggie*" gave us an estimate of the child's ability to produce the novel adjective /zogi/. We did not use this task as a selection criterion; errors were allowed. Instead, the task served as an indicator of any unusual productions of segments in particular contexts that might occur during the novel adjective learning task, in case these needed to be taken into account during scoring (see below). All children were highly accurate in producing the correct syllable structure. However, the children with DLD were less accurate (M = 88.71% correct, SD = 11.29) than the TD children (M = 98.85%, SD = 2.19) in producing the speech sounds targeted on this task, $t(25) = 3.18$, $p = .004$.

Procedure

Novel Adjectives and Their Referents

Eight novel words representing adjective names were created. The children learned them in two sets of four words. In each set, two words were monosyllabic, and two contained two syllables. The novel words were /fɪm/, /taɪmɪk/, /zogi/, /beɪp/, /næfi/, /mok/, /kudɪp/, and /paɪt/. No two words shared the same initial consonant.

For each child, half of the words in each set were assigned to the RRCR condition, and half were placed in the RS condition. Thus, learning condition served as a within-subject variable. The words were randomly counterbalanced so that each word appeared in the RRCR condition for approximately half of the children in each group and in the RS condition for the other half of the children in the group. The words in the two conditions were matched on phonotactic probability (using both the positional segment sum and the biphone frequency sum). In addition, half the words in each condition came from dense phonological neighborhoods, and half came from sparse phonological neighborhoods. These values were obtained from Storkel and Hoover (2010).

Drawings were used as the visual referents for the novel adjectives. Prior to the study, the coauthors discussed unusual characteristics that could be applied to common objects and animals through custom-made drawings. The criteria were that the characteristics should be visually salient and sufficiently unusual to be referred to using a novel adjective name. Some of the characteristics could be described using actual adjective names (e.g., "bumpy"), though they were applied to objects that would not ordinarily possess that attribute. On the basis of the list of tentative characteristics selected by the coauthors, an artist provided samples for further discussion. These samples were included in pilot testing. A final set of characteristics and corresponding drawings was then selected.

To promote the interpretation that each word described a characteristic that applied across different objects, drawings of four different objects were used for each novel adjective. Two were used during the learning period, and two were reserved for testing, as "generalization" items. For example, during the learning period, the child heard /fɪm/ associated with a drawing of a cat and a drawing of a tree, and during testing, the child was tested on /fɪm/ not only on those two items but also on a drawing of a pig and a drawing of an apple. The drawings used for the learned and generalization items for two other novel adjectives are shown in Figure 1.

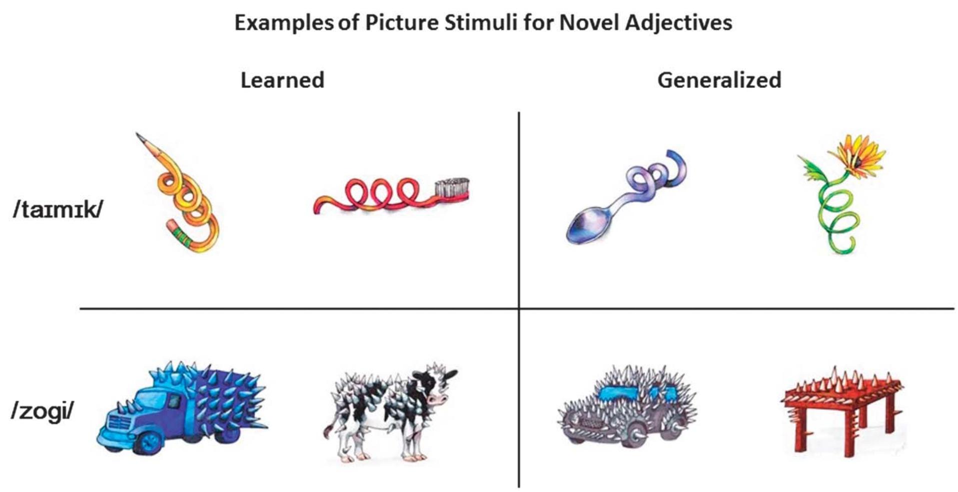
Learning Period

The two sets were learned sequentially, with 1 week separating the final testing of one set and the first learning session of the second set. The pictures and corresponding audio files were presented via a laptop computer. The Appendix provides the sequence used during the learning period for one of the sets. The first learning session began with familiarization items. The two objects for each adjective (e.g., a cat and a tree) were presented side by side, and the child heard, "Look! These are very (e.g., /fɪm/)" or (e.g., referring to a coat and a house) "Look! These are really (e.g., /næfi/)." By hearing the same word associated with two clearly different objects, it was hoped that, from the outset, the child would view the word as referring to a characteristic rather than an object. These familiarization items were repeated at the beginning of the second day.

After the four words in the set had been introduced in this way, the first of four blocks commenced. Two blocks were presented per day for 2 consecutive days, with a brief break between the blocks on the same day. Each block was approximately 10 min in duration. Within each block, the words from the two conditions appeared in alternating order. On the second day, the order of the words was reversed (e.g., /beɪp/-/zogi/-/taɪmɪk/-/fɪm/ instead of /fɪm/-/taɪmɪk/-/zogi/-/beɪp/). This meant that if, on the first day, the first word (e.g., /fɪm/) was in the RRCR condition, then the first word on the second day (e.g., /beɪp/) was in the RS condition and vice versa. We also alternated the referent picture each time a word reappeared in the sequence (e.g., first appearance: a /fɪm/ cat; second appearance: a /fɪm/ tree). All words were heard an equal number of times during each block, for a total of 44 exposures across the 2 days of learning. Thus, the words in the two conditions were matched for exposure and differed only in whether retrieval opportunities were provided. An example of a first block appears in Figure 2, with Figure 2a illustrating a word in the RRCR condition and Figure 2b highlighting a word in the RS condition.

All study trials had the same format. The picture appeared on the screen, and the child heard a prerecorded three-sentence sequence, as in "This cat is /fɪm/. It's very /fɪm/. This cat is really /fɪm/." Thus, there were three exposures of the novel adjective for each study trial. For the retrieval trials, the child saw the picture and heard a prerecorded request to supply the appropriate adjective to complete the sentence, as in "Tell me about the cat. The cat is very _____." The word *very* was prolonged to signal that the child should supply the next word. A study trial for the same word always immediately followed a retrieval

Figure 1. Examples of the pictures used for learned and generalization items for two of the nonwords used for novel adjectives. Copyright © Stephanie Funcheon. Reprinted with permission.



trial. The children were encouraged to listen to the novel words and look at the corresponding picture for all trials. For retrieval trials, the children were provided verbal praise for responding, regardless of accuracy.

Contextual reinstatement depends on retrieval success. To increase the likelihood that retrieval would be successful when spacing was instituted, we arranged for the retrieval trial for the first four items in the RRCR condition (e.g., /fɪm/ cat, /zogi/ truck, /fɪm/ tree, /zogi/ cow) to be immediately preceded by a study trial. Because there are no intervening words between a study trial and a retrieval trial in this instance, we refer to this kind of retrieval trial as a "0" retrieval trial. Figure 2a shows an example. After these four items went through two "0" trials, they appeared in a "3" retrieval trial. That is, as shown in Figure 2a, three items involving other words intervened between a retrieval trial and the preceding study trial for the same word. For the remaining blocks of the learning period, all retrieval trials were "3" retrieval trials.

Because the first four items in the RRCR condition involved two study trials (study, retrieval, study; see Figure 2a), we provided two consecutive study trials (study, study) for the first four items in the RS condition. This served to ensure that the words in the RRCR and RS conditions were heard the same number of times. Thereafter, only a single study trial was used for each item in the RS condition. These details can be seen in Figure 2b.

Recall and Recognition Testing

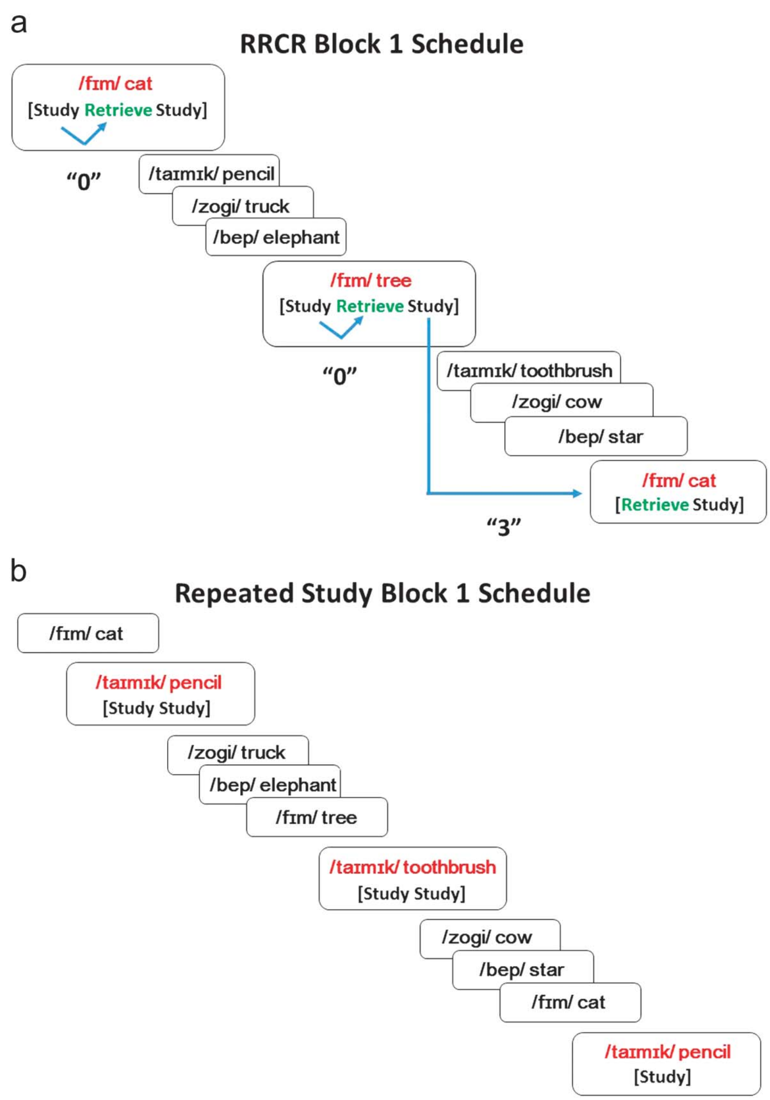
Five minutes following the fourth block, the first recall test was administered. Hereafter, this test will be referred to as the "5-min test." A third order was used for this test, again with the words from the two conditions alternating in order. All items of the test involved presenting the picture on the laptop, and the child was asked to complete the sentence, in the same manner as in the retrieval trials used during the learning period, as in "Tell me about the truck. The truck is very ______." For this test, Items 1–4 and 9–12 referred to drawings used during the learning period, whereas Items 5–8 and 13–16 referred to drawings never seen before (generalization items).

One week later, the identical recall test was administered. This was followed by the form–referent link recognition test. Sixteen items were used for this test, corresponding to the two learned drawings and two generalization drawings for each of the four words. The items were in random order with the provision that items testing the same word were separated by at least two other items. For each item, four drawings were shown and the child was asked, "Show me the one that is (e.g., /zogi/)." (The sentence structure of this prompt allowed the adjective to appear in final position. Although this structure also permitted the final word to be interpreted as a noun—comparable to "water" or "clothing"—the recognition test was administered at the end of the children's research participation after they had heard multiple examples of the word being used in an adjective context.) The three distractor drawings for each item came from the same set. Two of the four drawings were learned items, and two were generalization drawings. The position of the correct drawing was systematically varied across the four spatial positions on the screen.

Scoring and Reliability

For the recall tests, the data took the form of the number of items judged to have been recalled correctly, tallied separately for each participant group (DLD, TD), condition (RRCR, RS), time (5 min, 1 week), and item type (learned, generalization). The judgments were based on several criteria. First, a response could not have resembled an actual word (e.g., "bumpy") or an adaptation of an actual word (e.g., "stairy" = stair-step-like) that could potentially apply to the referent picture. Second, based on

Figure 2. (a) An example of the first block showing a novel word /fɪm/ assigned to the repeated retrieval with contextual reinstatement (RRCR) condition. In this block, /fɪm/ is retrieved in three instances. Retrieval is immediate in the first two retrieval trials. These are designated "0" because there are no words intervening between the retrieval trial and the preceding study trial. For the third retrieval trial for /fɪm/, three other words intervened between the retrieval trial and the preceding study trial. For this reason, this retrieval trial is designated "3." (b) An example of the first block showing a novel word /taɪmɪk/ assigned to the repeated study condition. In this instance, three other words intervened between appearances of each word, but only study trials are employed.



our subjective, holistic interpretation, the response had to be a plausible attempt at the correct word. For this determination, we also considered the child's phonological patterns in the 18-item production test of real words resembling the syllable shapes and phonetic content of the novel words. Responses meeting these criteria were then subjected to the third criterion. We scored each response according to the scoring system developed by Edwards, Beckham, and Munson (2004). Each consonant in the production was given 1 point each for correct place, manner, and voicing. Each vowel was awarded 1 point each for correct backness, height, and length. An additional point was given if the production retained the correct syllable shape (e.g., consonant–vowel–consonant, consonant–vowel–consonant–vowel, consonant–vowel–consonant–vowel–consonant). These points were totaled for each response and compared to the total that would be awarded if the child's production was actually an attempt at one of the other novel adjectives. For example, if a child responded to a /fɪm/ item with the production of /bɪm/ instead of /fɪm/, it would earn 8 points if regarded as an attempt at /fɪm/ (1 + 3 + 3 + 1 = 8), whereas it would earn 7 points if regarded as an attempt at the (incorrect) word /beɪp/ (3 + 2 + 1 + 1 = 7). If the production earned more points when treated as an attempt at the correct adjective, the child was credited with a correct response. Note that, because the novel words were quite phonologically distinct, judgments based on this scoring system rarely diverged from subjective judgments.

Scoring of the children's responses on the form–referent link recognition task was based on the accuracy of the children's pointing. The four drawings on the screen for each item were sufficiently spaced so that the specific drawing selected by the child could be determined quite easily. Self-corrections were accepted if they were immediate.

To assess reliability for scoring the children's responses on the recall tests, a second judge independently scored the 5-min and 1-week recall tests for one set from 12 children, six from each participant group. Item-by-item interjudge agreement for accuracy was 93% for the responses of the DLD group and 97% for the responses of the TD group.

Data Analysis

To examine the children's novel word recall and form–referent link recognition, a series of mixed-effects models were estimated, with and without the covariates of PPVT-4 standard score, EVT-2 standard score, K-ABC2 standard score, and maternal education. Diagnostic group (DLD, TD) was a between-participants variable; within-participant variables were learning condition (RRCR, RS), time (5 min, 1 week, for word recall only), and item type (learned, generalization). Random slopes for learning condition, time, and item type were included in the models when they were not close to zero. As a result, only the random slope for the learning condition variable was included in the models for recall. Both learning condition and item type random slopes were relevant in the recognition models.

Main effects models and full factorial models that included all possible two-way, three-way, and four-way interactions were tested hierarchically. We present the main effects models with no interactions to provide baseline, pooled effects of each model variable. We also present the best fitting model, which is the model that includes all statistically significant interactions and subinteractions. This resulted in a model with one 3-way interaction and three 2-way interactions for recall and a model with one 2-way interaction for recognition. Effect sizes are reported as partially standardized beta coefficients (b_{std}), which are comparable to a Cohen's *d,* except they represent conditional, standardized mean differences, conditioned on other variables in the model. Note that, in the text and figures, when we report the mean number of items correctly recalled or recognized, these are the unconditional means.

Results

Word Recall

Summaries of the results for word recall appear in Tables 1–3. Although one of the covariates, the EVT-2, was significant (p = .050), the covariates had no bearing on the main effects or interactions. As can be seen in Table 1, there was a large effect for learning condition (b_{std} = 0.91), pooled over group, time, and item type. Recall scores were 2.14 points higher for words in the RRCR condition (M = 4.54, SD = 1.68) compared to recall of words in the RS condition (M = 2.40, SD = 2.35). The magnitude of this difference can also be seen in the mean percentages correct (out of a maximum score of 8)—57% versus 30%—where

Table 1. Main effects model for the recall test outcome.

Fixed effects	b	95% CI		b_{std}	p
Group (DLD vs. TD)	−0.39	−2.82	2.04	−0.17	.753
Condition (RRCR vs. RS)	2.14	1.18	3.10	0.91	.000
Item (learn vs. gen)	0.29	0.12	0.46	0.12	.001
Time (1 week vs. 5 min)	0.08	−0.09	0.25	0.04	.341
Covariates					
PPVT-4	−0.01	−0.10	0.08		.887
Maternal education	−0.32	−0.82	0.18		.204
EVT-2	0.13	0.00	0.25		.050
K-ABC2	−0.03	−0.11	0.06		.542
Intercept	−2.58	−15.98	10.82		
Random effects	σ^2				
Condition	6.22	3.55	10.91		
Intercept	5.12	2.76	9.49		

Note. N = 27, observations = 216. CI = confidence interval; DLD = developmental language disorder; TD = typically developing; RRCR = repeated retrieval with contextual reinstatement; RS = repeated study; learn = learned; gen = generalization; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; EVT-2 = Expressive Vocabulary Test–Second Edition; K-ABC2 = Kaufman Assessment Battery for Children, Second Edition.

Table 2. Model with Group × Time × Condition interaction for the recall test outcome.

Fixed effects	b	95% CI		b_{std}	p
Group (DLD vs. TD)	−0.53	−2.97	1.92		.672
Condition (RRCR vs. RS)	1.54	0.17	2.91		.028
Item (learn vs. gen)	0.29	0.12	0.45		.001
Time (1 week vs. 5 min)	0.27	−0.06	0.60		.113
Two-way interactions					
Group × Time	0.23	−0.23	0.69		.329
Condition × Time	−0.23	−0.70	0.24		.337
Group × Condition	1.75	−0.16	3.65		.072
Three-way interactions					
Group × Time × Condition	−0.73	−1.39	−0.08	−0.31	.028
Covariates					
PPVT-4	0.01	−0.10	0.08		.888
Maternal education	−0.32	−0.82	0.18		.204
EVT-2	0.13	0.00	0.25		.050
K-ABC2	−0.03	−0.11	0.06		.698
Intercept	−2.66	−16.06	10.75		.698
Random effects	σ^2				
Condition	5.99	3.39	10.63		
Intercept	5.14	2.78	9.51		

Note. N = 27, observations = 216. CI = confidence interval; DLD = developmental language disorder; TD = typically developing; RRCR = repeated retrieval with contextual reinstatement; RS = repeated study; learn = learned; gen = generalization; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; EVT-2 = Expressive Vocabulary Test–Second Edition; K-ABC2 = Kaufman Assessment Battery for Children, Second Edition.

Table 3. Simple effects for the Group × Time × Condition interaction for recall.

Comparison	*b*	95% CI		b_{std}	*p*
DLD vs. TD for RS condition at 5 min	−0.53	−2.97	1.92	−0.22	.672
DLD vs. TD for RRCR condition at 5 min	1.22	−1.85	4.28	0.52	.435
DLD vs. TD for RS condition at 1 week	−0.30	−2.74	2.15	−0.13	.812
DLD vs. TD for RRCR condition at 1 week	0.72	−2.35	3.78	0.30	.647
RRCR vs. RS for TD group at 5 min	1.54	0.17	2.91	0.65	.028
RRCR vs. RS for DLD group at 5 min	3.29	1.96	4.61	1.40	.000
RRCR vs. RS for TD group at 1 week	1.31	−0.06	2.68	0.56	.062
RRCR vs. RS for DLD group at 1 week	2.32	1.00	3.64	0.99	.001
1 week vs. 5 min for TD group in RS condition	0.27	−0.06	0.60	0.11	.113
1 week vs. 5 min for DLD group in RS condition	0.50	0.18	0.82	0.21	.002
1 week vs. 5 min for TD group in RRCR condition	0.04	−0.29	0.37	0.02	.821
1 week vs. 5 min for DLD group in RRCR condition	−0.46	−0.79	−0.14	−0.20	.005

Note. CI = confidence interval; DLD = developmental language disorder; TD = typically developing; RRCR = repeated retrieval with contextual reinstatement; RS = repeated study.

the mean score for the RRCR condition was almost double the mean score for the RS condition. Thirteen of the 14 children in the DLD group had higher total recall scores for the RRCR condition than for the RS condition. The remaining child showed the reverse direction. Ten of the 13 children in the TD group also had higher total scores for the RRCR condition. Two additional children's scores were the same for the two learning conditions. The remaining TD child showed a large difference in the opposite direction. Further inspection of the data indicated that the recall differences favoring the RRCR condition over the RS condition held true for each set (Set 1: $t(27) = 4.14$, $p = .0003$, $d = 1.06$; Set 2: $t(27) = 2.39$, $p = .0241$, $d = 0.61$).

There was also a small effect for item type ($b_{std} = 0.12$). Recall scores were 0.29 points higher for items that were included during the learning period ($M = 3.62$, $SD = 1.61$) than for items representing generalization items ($M = 3.33$, $SD = 1.63$).

Although neither a main effect for participant group nor a main effect for time was observed, each variable was involved in a three-way Group × Condition × Time interaction, reflecting a moderate effect size ($b_{std} = 0.31$). This interaction is shown in Table 2 and illustrated in Figure 3. The simple effects shown in Table 3 reveal that, at both time points, each group's recall scores were higher in the RRCR condition than in the RS condition. This effect was moderate for the TD group (b_{std} = 0.56–0.65) and very large for the DLD group (b_{std} = 0.99–1.40). For the TD children, recall scores were 1.54 points higher for RRCR than for RS at 5 min (RRCR: $M = 4.42$, $SD = 1.61$; RS: $M = 2.88$, $SD = 2.26$; *M*s = 55% vs. 36%) and 1.31 points higher at 1 week (RRCR: $M = 4.46$, $SD = 1.84$; RS: $M = 3.15$, $SD = 2.48$; *M*s = 56% vs. 39%). For the children with DLD, the corresponding advantages for RRCR over RS were 3.29 points (RRCR: $M = 4.86$, $SD = 1.61$; RS: $M = 1.57$, $SD = 2.26$; *M*s = 61% vs. 20%) and 2.32 points (RRCR: $M = 4.39$, $SD = 1.84$; RS: $M = 2.07$, $SD = 2.47$; *M*s = 55% vs. 26%). Thus, for the children in the DLD group, scores were more than three times higher in the RRCR condition than in the RS condition at 5 min and more than two times higher at 1 week. The simple effects also revealed that scores for the DLD group declined by 0.46 points from 5 min to 1 week in the RRCR condition ($b_{std} = 0.20$) but increased by 0.50 points in the RS condition ($b_{std} = 0.21$). These were small effects, and, as noted above, even at 1 week, the difference favoring RRCR for these children was very large. No other interactions proved informative.

As noted earlier, there was a small effect for item type when pooled over learning condition, group, and time. However, item type did not enter into any interactions. This is illustrated in Figure 4, where it can be seen that generalized items reflected the same RRCR > RS difference seen

Figure 3. The (unconditional) mean number of items correct on the recall test at 5 min and 1 week for novel adjectives in the repeated retrieval with contextual reinstatement (RRCR) condition and the repeated study (RS) condition by children with developmental language disorder (DLD) and children with typical language development (TD). The figure reflects the Group × Time × Condition interaction, collapsed across item type (learned, generalization). Maximum score = 8. Error bars are standard errors.

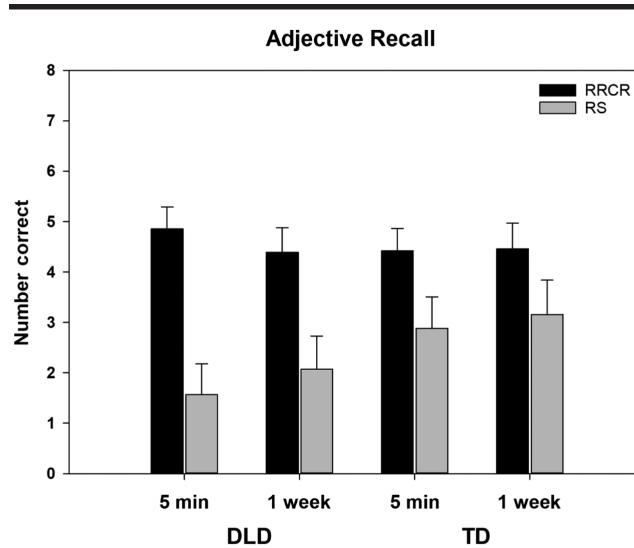
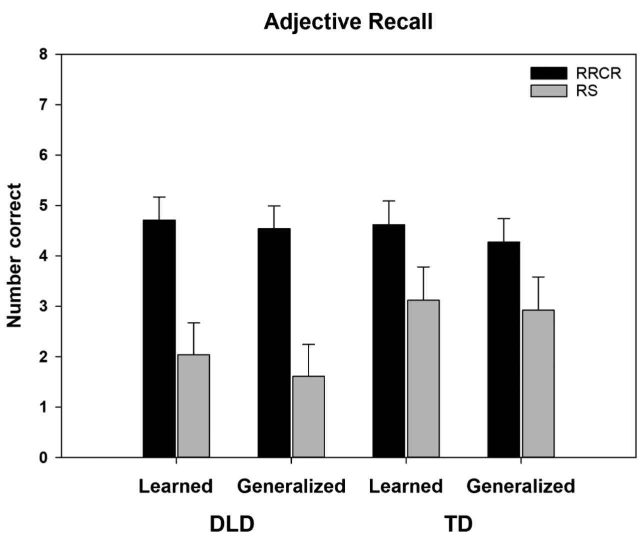


Figure 4. The (unconditional) mean number of novel adjective items correct on the recall test for learned and generalization items in the repeated retrieval with contextual reinstatement (RRCR) condition and the repeated study (RS) condition by children with developmental language disorder (DLD) and children with typical language development (TD). The figure reflects the similar behavior of the learned and generalization items when considered within each learning condition and participant group. Maximum score = 8. Error bars are standard errors.



for learned items, and showed the same pattern as learned items when the two groups of children are compared.

The phonetic criteria used to determine recall accuracy allowed for responses that did not perfectly match the correct form. Two different responses, each judged as correct, could have differed in how faithfully they represented the target word in their phonetic details. For this reason, we examined each child's correct response in terms of the points earned using the Edwards et al. (2004) system. Because the number of points possible varied depending on the syllable structure of the word, we computed percentages correct. Of the words judged as correctly recalled, the productions of the TD children (M = 97.31%, SD = 2.95) were reliably more accurate than the productions of the children with DLD (M = 91.86%, SD = 4.61), $t(25) = 3.63$, $p = .0013$, $d = 1.41$. No other differences in phonetic accuracy were apparent in the data. Recall that the children with DLD were less accurate than the TD children in their productions of the final word of the sentences on the real-word probes administered during the participant selection process. However, their accuracy on the probes was not correlated with their phonetic accuracy in producing the novel adjectives, $r = .10$, $p > .05$. The same was true for the TD children, $r = .38$, $p > .05$.

Form–Referent Link Recognition

The results for recognition are summarized in Tables 4–6 and illustrated in Figure 5. The covariates played no role in the results. Pooled across groups and item type

Table 4. Main effects model for the recognition outcome.

Fixed effects	b	95% CI		b_{std}	p
Group (DLD vs. TD)	−1.13	−2.63	0.36	−0.67	.137
Condition (RRCR vs. RS)	0.79	0.33	1.24	0.46	.001
Item (learn vs. gen)	0.29	−0.11	0.69	0.17	.161
Covariates					
PPVT-4	0.01	−0.07	0.05		.710
Maternal education	−0.18	−0.49	0.12		.245
EVT-2	0.01	−0.07	0.09		.749
K-ABC2	0.03	−0.02	0.09		.205
Intercept	5.48	−2.90	13.87		.200
Random effects	σ^2				
Condition	0.79	0.28	2.26		
Item	0.48	0.12	1.90		
Intercept	1.67	0.80	3.49		

Note. N = 26, observations = 104. CI = confidence interval; DLD = developmental language disorder; TD = typically developing; RRCR = repeated retrieval with contextual reinstatement; RS = repeated study; learn = learned; gen = generalization; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; EVT-2 = Expressive Vocabulary Test–Second Edition; K-ABC2 = Kaufman Assessment Battery for Children, Second Edition.

(see Table 4), children recognized 0.79 more items from the RRCR condition (M = 6.90, SD = 1.47, corresponding to 86%) than from the RS condition (M = 6.11, SD = 1.47, corresponding to 76%; $b_{std} = 0.46$). From Table 5, it can be seen that there was a Group × Condition interaction, which revealed that the advantage for the RRCR condition was 1.12 items larger for the children with DLD than for

Table 5. Model with Group × Condition interaction for the recognition outcome.

Fixed effects	b	95% CI		b_{std}	p
Group (DLD vs. TD)	−1.39	2.89	0.12		.071
Condition (RRCR vs. RS)	0.23	−0.35	0.81		.437
Item (learn vs. gen)	0.29	−0.11	0.69		.160
Two-way interactions					
Group × Condition	1.12	0.29	1.94	0.66	.008
Covariates					
PPVT-4	−0.01	−0.07	0.05		.737
Maternal education	−0.18	−0.49	0.13		.252
EVT-2	0.01	−0.07	0.09		.773
K-ABC2	0.03	−0.02	0.09		.217
Intercept	5.69	−2.67	14.05		.182
Random effects	σ^2				
Condition	0.53	0.15	1.96		
Item	0.48	0.12	1.89		
Intercept	1.67	0.80	3.48		

Note. N = 26, observations = 104. CI = confidence interval; DLD = developmental language disorder; TD = typically developing; RRCR = repeated retrieval with contextual reinstatement; RS = repeated study; learn = learned; gen = generalization; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; EVT-2 = Expressive Vocabulary Test–Second Edition; K-ABC2 = Kaufman Assessment Battery for Children, Second Edition.

Table 6. Simple effects for the Group × Condition interaction for recognition.

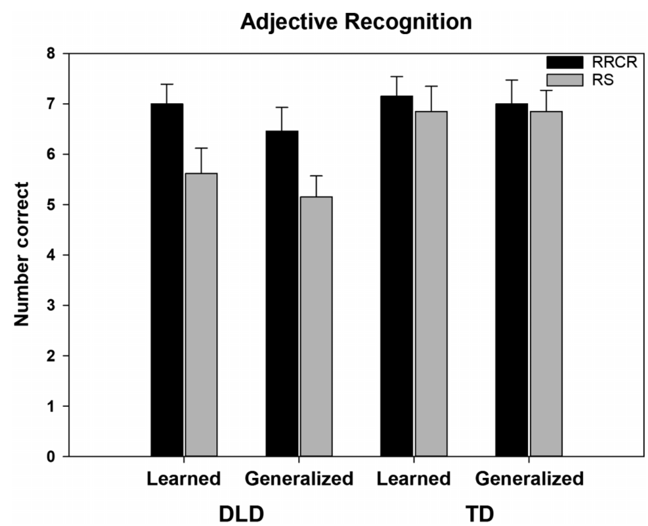
Comparison	b	95% CI		b_{std}	p
DLD vs. TD for RS condition	−1.39	−2.89	0.12	−0.82	.071
DLD vs. TD for RRCR condition	−0.27	−1.88	1.33	−0.16	.739
RRCR vs. RS for TD group	0.23	−0.35	0.81	0.14	.437
RRCR vs. RS for DLD group	1.35	0.76	1.93	0.79	.000

Note. CI = confidence interval; DLD = developmental language disorder; TD = typically developing; RS = repeated study; RRCR = repeated retrieval with contextual reinstatement.

the TD children (b_{std} = 0.66). For the DLD group, RRCR: *M* = 6.73, *SD* = 1.46; RS: *M* = 5.38, *SD* = 1.47. For the TD group, RRCR: *M* = 7.08, *SD* = 1.47; RS: *M* = 6.85, *SD* = 1.48. According to the simple effects (see Table 6), this advantage of the RRCR condition held only for the DLD group (mean percentages of 84% vs. 67%).

Item type was not involved in a main effect or any interactions. As can be seen in Figure 6, learned and generalization items behaved in a similar manner within each learning condition and participant group.

Figure 6. The (unconditional) mean number of learned and generalization items correct on the form–referent link recognition test for novel adjectives in the repeated retrieval with contextual reinstatement (RRCR) condition and the repeated study (RS) condition by children with developmental language disorder (DLD) and children with typical language development (TD). The figure reflects the similar behavior of the learned and generalization items within each learning condition and participant group. Maximum score = 8. Error bars are standard errors.

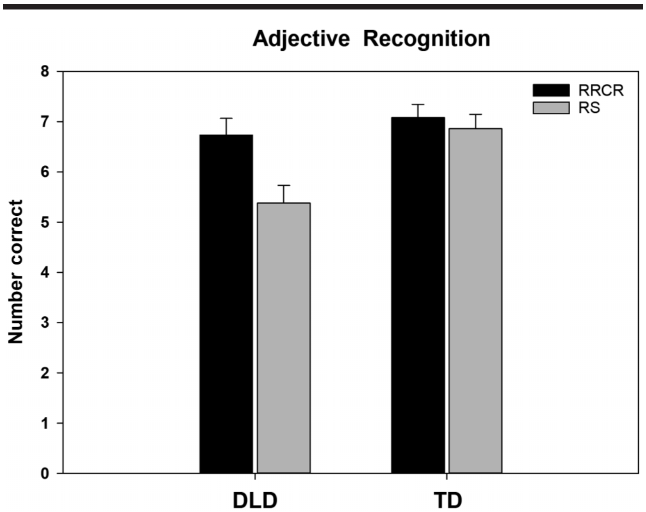


Discussion

Word Recall

The RRCR condition was associated with better recall of the novel adjectives than the RS condition. This was true for both groups of children, though the effect was especially strong for the children with DLD. The RRCR advantage held across time and across learned and generalization items. For the DLD group, the RRCR advantage decreased in magnitude from 5 min to 1 week, but even after 1 week, the effect was large.

Figure 5. The (unconditional) mean number of items correct on the form–referent link recognition test for novel adjectives in the repeated retrieval with contextual reinstatement (RRCR) condition and the repeated study (RS) condition by children with developmental language disorder (DLD) and children with typical language development (TD). The figure reflects the Group × Condition interaction, collapsed across item type (learned, generalization). Maximum score = 8. Error bars are standard errors.



Although the benefits of repeated retrieval over repeated study are well established in the psychology literature, our use of a within-subject design put the process to a stringent test. During the learning period, words in the RRCR condition and those in the RS condition appeared in alternating order. Therefore, upon being asked to retrieve one word, the children might have anticipated having to retrieve the next word appearing in the sequence. We do not know if the children engaged in covert retrieval of the words in the RS condition, but, if they did, it did not provide the boost to recall seen for the words in the RRCR condition.

We did not find a difference between the children with DLD and the TD children in word recall—a finding that mirrors the results of our previous study that compared novel noun learning for RRCR versus RS conditions (Leonard et al., 2019). However, we do not claim that the two groups were comparable in ability. First, a larger number of participants might have revealed a group difference. The number of children who participated was sufficient to reveal large differences between the two learning conditions but might have been insufficient to reveal genuine differences between groups. Second, we can easily imagine how a task with more novel words, fewer exposures, and a retention test administered more than 1 week later could have produced a clear difference between the groups. The TD children were not at ceiling in our recall task; therefore, if any such changes in task resulted in a greater gap between the two groups, it would be a matter of the children with DLD showing a larger drop in performance relative to the decline seen in their peers.

Even though the DLD and TD groups did not differ in their recall scores, a look at the phonetic details of their correct responses indicated that the two groups were not equivalent in how accurately their recall responses matched the target word. This finding accords well with findings from other studies that suggest encoding weaknesses in children with DLD (Bishop & Hsu, 2015; McGregor et al., 2017; McGregor, Licandro, et al., 2013). Even when the children showed evidence of producing particular speech sounds and syllable shapes correctly during our initial real-word probes, the children with DLD were more likely to respond with productions such as /tæmɪk/ for /taɪmɪk/ and /kubɪp/ for /kudɪp/ during the recall tests. These off-the-mark productions did not seem to be the result of degradations in the representations of the words over time, because, for words in the RRCR condition, children with DLD sometimes produced the words in a similar imprecise manner early in the learning period. We do not think these production inaccuracies were readily attributable to problems with specific speech sounds, because even though the children with DLD were less accurate than the TD children on the real-word probes, we found no relationship between their accuracy on the real-world probes and their phonetic accuracy in producing the novel words.

We are confident that the children in both groups were interpreting the novel words as names for attributes rather than names for individual items. Although there was a main effect favoring items presented during the learning period over the generalization items, the effect size was relatively small. Strikingly, for all children—DLD and TD alike—if a word was correctly recalled for a learned item, it was also produced correctly in response to a generalization item for the same word.

Although the difference between the learning conditions held across both testing periods, the comparisons between 5-min and 1-week scores contained some unexpected findings. First, the DLD group's scores for the RRCR condition dropped somewhat from 5 min to 1 week, whereas the TD group's scores were relatively stable. Second, and more surprising, was the finding that the children with DLD showed an increase in scores from 5 min to 1 week for the RS condition. Why would this be the case? We can imagine how this might occur in general, though it is not clear why it would be true only for the DLD group. Specifically, words in the RS condition were heard but not retrieved during the learning period. The children's first opportunity for retrieval occurred during a 5-min testing. Although this put these words at a disadvantage relative to the words in the RRCR condition, the retrieval opportunity during 5-min testing might have permitted contextual reinstatement, leading to higher scores on the children's next attempts to retrieve these words during testing 1 week later. Of course, such limited retrieval practice, though beneficial, would not be sufficient to bring the 1-week RS scores up to the level seen for the RRCR condition. Words in the latter condition were associated with many more retrieval opportunities prior to testing.

Form–Referent Link Recognition

For the children with DLD, the findings for recognition accuracy mirrored those for word recall, with scores higher for the RRCR condition than for the RS condition. We found no condition effects for the TD children. In this case, ceiling effects were likely involved.

For both groups of children in both learning conditions, scores were much higher for the recognition test than for the second recall test that was likewise administered 1 week after the learning period. The recognition test required a representation of each word that was only as detailed as needed to distinguish it from the other novel words. For some of the words, this level of detail was unlikely to be sufficient to permit accurate retrieval. Nevertheless, these rather imprecise representations were sufficient for children to show the same degree of accuracy on generalization items as on learned items.

Implications

The use of spaced retrieval for the RRCR condition was intended to maximize the benefits of retrieval. Specifically, when other items are inserted between each word's study trials and its retrieval trials, the episodic context has changed. Retrieval with slightly changing contexts is assumed to enhance the context representation and make it less similar to that of other words, thus reducing the search space (Karpicke, 2017). Note that spaced retrieval is assumed to increase access to words, a process that applies to comprehension and production. Thus, RRCR is expected to have facilitative effects on receptive tasks as well. For the children with DLD, this was seen in the clear RRCR advantage over RS on the form–referent link recognition task.

However, to argue that the benefits of the RRCR condition were attributable in part to spaced retrieval rather than to the opportunity for retrieval more generally, we can only resort to our previous findings. In a study of novel noun learning, Haebig et al. (2019) found that a spaced retrieval condition produced greater recall than a condition that provided the same number of retrieval opportunities but with no intervening items. With such immediate retrieval, the context changes very little, and hence, the context representation does not become more distinctive.

The advantage of the RRCR condition over the RS condition was the same for the learned and generalization items. This finding suggests that the benefits of retrieval practice went beyond facilitating access to item-specific information. The words whose retrieval was aided in the RRCR condition had sufficient flexibility to be applied to new objects—objects from a different category (e.g., from a drawing of a cow to a drawing of a table). This observation that spaced retrieval provides the same advantage even when generalization is required is probably the most important finding of this study. We do not claim that the RRCR condition uniquely facilitated the generalization process, for the smaller number of words retrieved in the RS condition nevertheless showed a similar degree of generalization. Indeed, we suspect that the features needed for generalization—the perceptual features of the learned items—were available to a similar degree in the two conditions. The difference, we believe, is that the more distinctive context representation brought about by spaced retrieval in the RRCR condition simply increased the likelihood that the words (already equipped for generalization) could be successfully retrieved. That is, we believe that the children recognized the attribute shared by the generalization and learned items regardless of whether the word appeared in the RRCR or the RS condition. However, the RRCR condition was more successful in helping the children come up with the specific word that applied to the attribute.

Our finding of spaced retrieval advantages over repeated study for adjectives mirrors our earlier findings for novel nouns referring to exotic plants and animals (Leonard et al., 2019). Another parallel finding pertains to time; in both studies, children with DLD retained the words over 1 week as successfully as their TD peers. Similar findings were seen for our study comparing spaced retrieval with immediate retrieval (Haebig et al., 2019) and earlier studies of young adults by McGregor et al. (2017) that used a different learning procedure. From these findings, we cannot conclude that individuals with DLD function adequately in their long-term retention. However, the evidence thus far suggests that, once a word has reached the threshold of being recalled reliably immediately after the learning period, retention thereafter does not appear to be a major concern. Accordingly, the problem may be more one of learning in the first place rather than one of forgetting what has been learned.

Limitations and Next Steps

Finding that a spaced retrieval condition yielded better recall than repeated study is still a long way from concluding that the procedures described here should be adopted. Even though all children generalized the words they learned, they only recalled approximately 57% of the words in the RRCR condition; recognition was better for these words, but the 86% accuracy level needs to be weighed against the fact that the recognition task required only a rough approximation of a word's representation. It is likely that a more effective spaced retrieval schedule could be found, one that might result in even better recall. Our selection of the number of words for the children to learn and the number of word exposures to use might well have been far from ideal. Modifications in these details might further increase children's word learning and recall.

Our use of adjectives in this study provided valuable information about retrieval practice and generalization. However, other lexical classes remain to be examined. Discovering the role of spaced retrieval in verb learning would be an important next step. Along with determining whether retrieval aids the learning and generalization of verbs, investigators could assess whether verbs that are better retained through retrieval practice are also more likely to be modulated with appropriate grammatical inflections (e.g., /beɪp/, /beɪps/, /beɪpt/, /beɪpɪŋ/).

The tight controls provided by our experimentally created words and consistent auditory and visual presentations constituted a necessary early step in evaluating retrieval practice effects on word learning. A subsequent phase will necessarily involve real words, perhaps in a more natural presentation mode such as book reading. It is not difficult to imagine how such a future study might be designed if the data warrant moving to this more ecologically relevant stage.

Research is also needed to better isolate some of the learning mechanisms involved when retrieval assists the retention of words. One such mechanism is consolidation (e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). The effect of consolidation is often detected in a design in which participants learn a set of new words in the evening and are tested 12 hr later with intervening sleep or learn the words in the morning and are tested after 12 hr of wakefulness. Recall is usually significantly greater when sleep has intervened. Using this type of design, McGregor, Licandro, et al. (2013) found that a group of young adults with DLD showed consolidation of word meanings but not of the novel word forms themselves. In our study, the learning period was divided into two short sessions held over 2 consecutive days. Pilot work suggested that a single, longer session would lead to fatigue and not be conducive to learning. However, by extending the learning period over 2 days, we were not in a position to compare recall with or without intervening sleep. Our finding that recall 1 week later was as strong as recall immediately after the learning period indicated that retrieval benefits were maintained, but we cannot yet determine whether retrieval practice effects were already evident on the first day. Studies of young adults with typical language reveal advantages of retrieval over study even when recall testing occurs during the first/only day, 15 min after the learning period (Smith, Roediger, & Karpicke, 2013). However, we have not yet incorporated first-day recall testing in our studies of young children. This is an important theoretical question, though it may be a less pressing issue from a clinical perspective, because the goal of word learning activities is for the child to retain a word over the long term.

Although our study was not a comparison of alternative retrieval-based theories, we have discussed repeated spaced retrieval within the framework of the episodic context account (e.g., Karpicke et al., 2014) because it seemed to be a good fit given the nature of our task. The children heard novel words that had no already-established connections to other words, and the auditory presentation of the words and the visual presentation of the objects associated with those words were unchanged throughout the learning period. That is, the semantic and phonological information was constant. What did change was the fact that retrieval occurred at different time points with intervening material. These slight changes in context allowed the features of the newest retrieval context to be blended with those of the previous contexts, which allowed the item to be more distinctive. With a unique composition of contextual features, the search set was presumably more restricted, and retrieval could be more effective.

Future work might test the assumption of this account by manipulating episodic retrieval in a more overt way. Earlier, we discussed the study by Whiffen and Karpicke (2017) in which half of the participants were asked to judge which of two lists each word had appeared in. Recall proved to be superior in that condition, indicating that list membership must have been one of the contextual features that became associated with the word during encoding. An adaptation of this type of task might be created that is suitable for children.

Even with the shortcomings of our procedure and the need for additional research, we are heartened by three key observations. First, thus far, the use of RRCR seems to assist the learning and retention of new words for TD children and children with DLD alike. Therefore, this type of procedure might have general applicability beyond its use as a clinical tool for children with language deficits.

Second, for both groups, the RRCR advantage was seen even for items requiring generalization of the novel adjectives to new objects. This finding suggests that the value of RRCR goes beyond learning and retaining an association between a new word and a specific referent. Rather, the benefits of RRCR seem to apply to true word learning, where new words must extend to referents that were never involved in the original learning of the words.

Third, to the extent that presenting children with numerous exposures to each new word and its referent constitutes an appropriate method of teaching children vocabulary (Hart & Risley, 1995; Suskind & Suskind, 2015), adding opportunities for the children to retrieve the word seems likely to improve learning further. In fact, when we inspect Figure 3, we see that the recall of the children with DLD in the RRCR condition was as good as, if not better than, the recall of the TD children using the more conventional RS procedure of only continuously hearing the word and seeing its referent. Through further refinement of this retrieval-based approach, children with DLD might be in a better position to participate in word learning activities with their peers without significant additional assistance.

Acknowledgments

This research was supported in part by Research Grant R01 DC014708 awarded to Laurence B. Leonard. Eileen Haebig was supported by a postdoctoral fellowship Training Grant T32 DC00030.

References

Alt, M. (2011). Phonological working memory impairments in children with specific language impairment. *Journal of Communication Disorders, 44,* 173–185.

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,* 941–954.

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research, 47,* 407–420.

Bishop, D. V. M., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology, 3,* 3.

Booth, A., & Waxman, S. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development, 80,* 15–22.

Carpenter, S. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569.

Chen, Y., & Liu, H.-M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities, 35,* 10–20.

Dawson, J., Stout, C., Eyer, J., Tattersall, P., Fonkalsrud, J., & Croley, K. (2005). *Structured Photographic Expressive Language Test–Preschool 2*. DeKalb, IL: Janelle Publications.

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics, 19,* 193–207.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35–39.

Dunn, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4)*. Minneapolis, MN: AGS/Pearson.

Edwards, J., Beckham, M., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47,* 421–436.

Fazio, L., & Marsh, E. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science, 28,* 111–116.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), *Language development. Volume 2: Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, United Kingdom: Cambridge University Press.

Gladfelter, A., & Leonard, L. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. *Journal of Speech, Language, and Hearing Research, 56,* 542–552.

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology, 9,* 151–161.

Gray, S. (2003). Word learning by preschoolers with specific language impairment: What predicts success. *Journal of Speech, Language, and Hearing Research, 46,* 56–67.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research, 47,* 1117–1132.

Gray, S., Pittman, A., & Weinhold, J. (2014). Effect of phonotactic probability and neighborhood density on word-learning configuration by preschoolers with typical development and specific language impairment. *Journal of Speech, Language, and Hearing Research, 57,* 1011–1025.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30,* 196–206.

Greenslade, K., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the Structured Photographic Expressive Language Test–Preschool: Second Edition. *Language, Speech, and Hearing Services in Schools, 40,* 150–160.

Haebig, E., Leonard, L., Deevy, P., Karpicke, J., Christ, S., Usler, E., . . . Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder II: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research, 62,* 932–943.

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Howard, M., & Kahana, M. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299.

Kail, R., & Leonard, L. (1986). Word-finding problems in language-impaired children (ASHA Monographs, No. 25). Rockville, MD: American Speech-Language-Hearing Association.

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 53,* 739–756.

Karpicke, J. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, Volume 2. Learning and memory: A comprehensive reference* (pp. 487–514). Oxford, United Kingdom: Academic Press.

Karpicke, J., Lehman, M., & Aue, W. (2014). Retrieval-based learning: An episodic context account. In B. Ross (Ed.), *Psychology of learning and motivation: Volume 61* (pp. 238–284). Waltham, MA: Elsevier.

Karpicke, J., & Roediger, H. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704–719.

Karpicke, J., & Roediger, H. (2008). The critical importance of retrieval for learning. *Science, 319,* 966–968.

Karpicke, J., & Zaromb, F. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62,* 227–239.

Kaufman, A., & Kaufman, N. (2004). *Kaufman Assessment Battery for Children, Second Edition (K-ABC2)*. Circle Pines, MN: AGS.

Law, J., Tomblin, J. B., & Zhang, X. (2008). Characterizing the growth trajectories of language-impaired children between 7 and 11 years of age. *Journal of Speech, Language, and Hearing Research, 51,* 739–749.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Lehman, M., & Malmberg, K. (2013). A buffer model of encoding and temporal correlations in retrieval. *Psychological Review, 120,* 155–189.

Lehman, M., Smith, M., & Karpicke, J. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1787–1794.

Leonard, L., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., . . . Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder I: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research, 62,* 932–943.

Leonard, L., Nippold, M., Kail, R., & Hale, C. (1983). Picture naming in language-impaired children. *Journal of Speech and Hearing Research, 26,* 609–615.

Leonard, L., Schwartz, R., Chapman, K., Rowan, L., Prelock, P., Terrell, B., . . . Messick, C. (1982). Early lexical acquisition in children with specific language impairment. *Journal of Speech and Hearing Research, 25,* 554–564.

McGregor, K., Berns, A., Owen, A., Michels, S., Duff, D., Bahnsen, A., & Lloyd, M. (2012). Association between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders, 42,* 35–47.

McGregor, K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research, 60,* 2891–2905.

McGregor, 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,* 1845–1856.

McGregor, K., Newman, R., Reilly, R., & Capone, N. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 45,* 998–1014.

McGregor, 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,* 307–319.

McGregor, K., & Waxman, S. (1998). Object naming at multiple hierarchical levels: A comparison of preschoolers with and without word-finding deficits. *Journal of Child Language, 25,* 419–430.

Mintz, T., & Gleitman, L. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition, 84,* 267–293.

Oetting, J., Rice, M., & Swank, L. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech, Language, and Hearing Research, 38,* 434–445.

Rice, M., Buhr, J., & Nemeth, M. (1990). Fast mapping word learning abilities of language delayed preschoolers. *Journal of Speech and Hearing Disorders, 55,* 33–42.

Rice, M., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58,* 345–359.

Rice, M., Oetting, J., 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,* 106–122.

Rice, M., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53,* 333–350.

Roediger, H., & Karpicke, J. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Rowland, C. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432–1463.

Schopler, E., Van Bourgondien, M., Wellman, G., & Love, S. (2010). *Childhood Autism Rating Scale–Second Edition*. Los Angeles, CA: Western Psychological Services.

Skipp, A., Windfuhr, K., & Conti-Ramsden, G. (2002). Children's grammatical categories of verb and noun: A comparative look at children with specific language impairment (SLI) and normal language (NL). *International Journal of Language & Communication Disorders, 37,* 253–271.

Smith, M., Roediger, H., & Karpicke, J. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1712–1725.

Souto, S., Leonard, L., & Deevy, P. (2014). Identifying risk for specific language impairment with narrow and global measures of grammar. *Clinical Linguistics & Phonetics, 28,* 741–756.

Spaulding, T., Hosmer, S., & Schechtman, C. (2013). Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *International Journal of Speech-Language Pathology, 15,* 453–462.

Storkel, H., & Hoover, J. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods, 42,* 497–506.

Storkel, H., Komesidou, R., Fleming, K., & Romine, R. (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.

Storkel, H., Voelmie, K., Fierro, V., Flake, K., Fleming, K., & Romine, R. (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.

Suskind, D., & Suskind, B. (2015). *Thirty million words: Building a child's brain.* New York, NY: Dutton.

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research, 46,* 1283–1296.

Whiffen, J., & Karpicke, J. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1036–1046.

Williams, K. (2007). *Expressive Vocabulary Test–Second Edition.* Minneapolis, MN: NCS Pearson.

Windfuhr, K., Faragher, B., & Conti-Ramsden, G. (2002). Lexical learning skills in young children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 37,* 415–432.

Appendix (p. 1 of 2)

Sequence Used During the Learning and Testing Period for One Set

Set 1, Day 1

I. Familiarization: /fɪm/, /taɪmɪk/, /zogi/, /bep/

II. Learning phase

Novel adj.	Noun	Condition	Exposure
Block 1			
/fɪm/	cat	RRCR	Study–Retrieval–Study
/taɪmɪk/	pencil	RS	Study–Study
/zogi/	truck	RRCR	Study–Retrieval–Study
/bep/	elephant	RS	Study–Study
/fɪm/	tree	RRCR	Study–Retrieval–Study
/taɪmɪk/	toothbrush	RS	Study–Study
/zogi/	cow	RRCR	Study–Retrieval–Study
/bep/	star	RS	Study–Study
/fɪm/	cat	RRCR	Retrieval–Study
/taɪmɪk/	pencil	RS	Study
/zogi/	truck	RRCR	Retrieval–Study
/bep/	elephant	RS	Study
Block 2			
/fɪm/	tree	RRCR	Retrieval–Study
/taɪmɪk/	toothbrush	RS	Study
/zogi/	cow	RRCR	Retrieval–Study
/bep/	star	RS	Study
/fɪm/	cat	RRCR	Retrieval–Study
/taɪmɪk/	pencil	RS	Study
/zogi/	truck	RRCR	Retrieval–Study
/bep/	elephant	RS	Study
/fɪm/	tree	RRCR	Retrieval–Study
/taɪmɪk/	toothbrush	RS	Study
/zogi/	cow	RRCR	Retrieval–Study
/bep/	star	RS	Study

Appendix (p. 2 of 2)

Set 1, Day 2

I. Familiarization: /bep/, /zogi/, /taɪmɪk/, /fɪm/

II. Learning phase

Novel adj.	Noun	Condition	Exposure
Block 3			
/bep/	elephant	RS	Study
/zogi/	truck	RRCR	Retrieval–Study
/taɪmɪk/	pencil	RS	Study
/fɪm/	cat	RRCR	Retrieval–Study
/bep/	star	RS	Study
/zogi/	cow	RRCR	Retrieval–Study
/taɪmɪk/	toothbrush	RS	Study
/fɪm/	tree	RRCR	Retrieval–Study
/bep/	elephant	RS	Study
/zogi/	truck	RRCR	Retrieval–Study
/taɪmɪk/	pencil	RS	Study
/fɪm/	cat	RRCR	Retrieval–Study
Block 4			
/bep/	star	RS	Study
/zogi/	cow	RRCR	Retrieval–Study
/taɪmɪk/	toothbrush	RS	Study
/fɪm/	tree	RRCR	Retrieval–Study
/bep/	elephant	RS	Study
/zogi/	truck	RRCR	Retrieval–Study
/taɪmɪk/	pencil	RS	Study
/fɪm/	cat	RRCR	Retrieval–Study
/bep/	star	RS	Study
/zogi/	cow	RRCR	Retrieval–Study
/taɪmɪk/	toothbrush	RS	Study
/fɪm/	tree	RRCR	Retrieval–Study

III. 5-min break

IV. Recall

Novel adj.	Noun	Condition	Item type
/zogi/	truck	RRCR	Learned
/bep/	elephant	RS	Learned
/fɪm/	cat	RRCR	Learned
/taɪmɪk/	pencil	RS	Learned
/zogi/	car	RRCR	Generalized
/bep/	cup	RS	Generalized
/fɪm/	pig	RRCR	Generalized
/taɪmɪk/	spoon	RS	Generalized
/zogi/	cow	RRCR	Learned
/bep/	star	RS	Learned
/fɪm/	tree	RRCR	Learned
/taɪmɪk/	toothbrush	RS	Learned
/zogi/	table	RRCR	Generalized
/bep/	refrigerator	RS	Generalized
/fɪm/	apple	RRCR	Generalized
/taɪmɪk/	flower	RS	Generalized