Retrieval-Based Word Learning in Young Typically Developing Children and Children with Developmental Language Disorder I: The Benefits of Repeated Retrieval

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

[a]Purdue University, West Lafayette, IN, USA, [b]Louisiana State University, Baton Rouge, LA, USA, [c]Butler University, Indianapolis, IN, USA, [d]George Washington University, Washington, DC, USA

Correspondence to: Laurence B. Leonard, Purdue University, 715 Clinic Drive, Lyles-Porter Hall, West Lafayette, IN 47907, 765-494-3794, xdxl@purdue.edu

Running head: Retrieval-Based Word Learning I

Key Words: developmental language disorder, specific language impairment, language development, word learning, retrieval

Conflict of Interest

There are no relevant conflicts of interest to report.

Funding Statement

This research was supported in part by research grant R01 DC014708 awarded to Laurence B. Leonard. Eileen Haebig was supported by a postdoctoral fellowship on training grant T32 DC00030.

Abstract

Purpose: Scholars have long noted that retention improves significantly when learners frequently test themselves on the new material rather than engage in continuous study with no intermittent testing. In this study, we apply the notion of repeated testing or retrieval to the process of word learning in preschool-aged children with and without developmental language disorder (DLD).

Method: Novel words and their meanings were taught to 10 children with DLD and 10 typically developing (TD) children matched on age (DLD *M* = 63.4 months; TD *M* = 63.2 months). Recall was assessed immediately after the second learning session, and then again one week later.

Results: Both groups showed better retention when they had attempted to retrieve the words during the learning period than when they had simply listened to and studied the words paired with their referents. Relative to their TD peers, the children with DLD seemed to be weaker in their encoding, but these children's retention over a one-week period was indistinguishable from that of their age mates.

Conclusion: Word learning activities that include opportunities for repeated retrieval appear to significantly benefit retention relative to more traditional word learning activities.

In this study, we examine the word learning and retention abilities of preschool-age children with “developmental language disorder” (DLD), often referred to in the research literature as children with “specific language impairment” (SLI). Although much of the DLD literature has emphasized morphosyntax, these children also have major deficits in vocabulary (McGregor, Berns, Owen, Michels, Duff, Bahnsen, & Lloyd, 2012). In fact, longitudinal studies show that individuals with DLD fall further behind their peers in vocabulary ability from preschool to 21 years of age (Rice & Hoffman, 2015). Many studies have shown that the vocabularies of children with DLD have less breadth and depth than those of their same-age peers (Dollaghan, 1998; Kail & Leonard, 1986; Leonard, Nippold, Kail, & Hale, 1983; McGregor, Newman, Reilly, & Capone, 2002; McGregor, Oleson, Bahnsen, & Duff, 2013; McGregor & Waxman, 1998). Studies of novel word learning show that these children require more encounters with a word before learning takes place (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). Novel word learning weaknesses have been documented for nouns, verbs, and adjectives (Kan & Windsor, 2010; Oetting, Rice, & Swank, 1995; Skipp, Windfuhr, & Conti-Ramsden, 2002; Rice, Buhr, & Nemeth, 1990; Windfuhr, Faragher, & Conti-Ramsden, 2002).

Studies of word learning in children with DLD have sometimes made the important distinction between a word’s *form* and its *meaning*. For example, upon learning about pelicans, a child might subsequently have difficulty recalling the details of the actual name (/pɛlɪkən/) – a problem of word form – but successfully remember that the word refers to a bird that eats fish – a detail of word meaning. All children, but especially children with DLD, appear to have greater difficulty with word forms than with word meanings (Gray, 2004; McGregor, Licando et al., 2013).

Another important distinction is that between *encoding* and longer-term retention. For children with DLD, the initial encoding of word forms (e.g., transforming /pɛlɪkən/ in a manner that permits adequate storage) seems to be the greater concern (Bishop & Hsu, 2015; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017; McGregor, Licando et al., 2013).

Given these well documented lexical deficits in DLD, it is perhaps ironic that these children are usually initially identified on the basis of low scores on standardized tests of morphosyntax. Although traditional tests of vocabulary show significant group differences between children with DLD and their typically developing age mates, these tests have been shown to have inadequate sensitivity and specificity (Gray, Plante, Vance, & Henrichsen, 1999; Spaulding, Hosmer, & Schechtman, 2013). Studies assessing the diagnostic accuracy of these tests reveal that most children with DLD score within the normal range, suggesting that these tests are tapping only a relatively shallow level of lexical knowledge. For example, in their report on the weaknesses in both vocabulary breadth and depth in children with DLD, McGregor et al. (2012) noted that 13 of the 14 children with DLD scored within what would be considered the normal range on a standardized test of receptive vocabulary. The same has been true for studies reporting novel word learning deficits in children with DLD; in fact, some studies have shown no correlation between vocabulary test scores and the novel word learning deficits in children with DLD (e.g., Gray, 2003).

In the present study, we focus on the role of *retrieval* in the word learning and retention of preschoolers with DLD. Although the study of retrieval processes has only recently been applied to children’s learning of new words, retrieval has a rich history in the experimental psychology literature. For many years, scholars have noted that long-term retention improves significantly when learners frequently test themselves on the new material rather than engage in continuous study with no intermittent testing. This longstanding observation has recently been the subject of a resurgence in research in the area of memory. For example, using tasks ranging from the learning of science concepts to the learning of fictional stories, Karpicke and his colleagues found that, following an initial study period, repeated practice in retrieving the material resulted in greater long-term retention than consistently studying that material

(Karpicke, 2012; Karpicke & Blunt, 2011; Karpicke & Roediger, 2007b, 2008; Roediger & Karpicke, 2006). The benefits were dramatic; in these studies, long-term retention was 50% to 150% greater in the repeated retrieval condition than in the repeated study condition.

It appears that the advantage of repeated retrieval over repeated study can be enhanced further through the use of *spaced* retrieval. For example, Karpicke and Roediger (2007b) asked young adults to learn and recall a set of word pairs. For one group of participants, testing always occurred immediately after each word pair was studied; for two other groups, testing of each word pair always occurred only after several other word pairs had intervened. Although the immediate retrieval condition was less challenging during the learning period, a recall test two days later showed greater retention of the word pairs learned when there was spacing between study and test trials.

Karpicke and his colleagues have proposed a context-based account to explain the learning benefits derived from repeated retrieval with spacing, hereafter referred to as "repeated retrieval with contextual reinstatement" (Karpicke, Lehman, & Aue, 2014; Karpicke & Zaromb, 2010; Lehman, Smith, & Karpicke, 2014). According to this account, items are experienced and learned in a slowly changing temporal context (Howard & Kahana, 2002; Lehman & Malmberg, 2013). Consequently, features of the item (e.g., semantic and phonological features of a word) and features of the temporal context during study of that item (momentary details internal to the learner) are stored together. During retrieval, people attempt to reinstate the earlier learning context, using features of that context as cues to guide the search process. When retrieval is successful, the context representation associated with the target is updated by adding features of the present context to it. The addition of features to the context representation makes it a more effective cue for retrieving the desired target again because it results in a more restricted search set (Raaijmakers & Shiffrin, 1981); this smaller search set increases the likelihood that the target will be successfully retrieved. Further research suggests that such contextual reinstatement is more likely to occur during spaced retrieval – when the material must be retrieved after other information has intervened between the most recent exposure of the information to be retrieved and the actual retrieval attempt (Karpicke & Roediger, 2007a).

According to this contextual reinstatement account, no such benefit occurs when people repeatedly study material with no attempts at retrieval. In such cases, the target material is always present, so memory search is not required and the benefits of contextual reinstatement and updating do not occur.

In this investigation, we make use of repeated retrieval with contextual reinstatement to examine word learning in preschoolers with DLD and their same-age typically developing peers. Although the study of children's word learning has a rich history, the role of repeated retrieval during this process has been the subject of relatively few studies (e.g., Chen & Liu, 2014; Fritz, Morris, Nolan, & Singleton, 2007; McGregor et al., 2017). Our limited understanding of the contribution of retrieval can be attributed in part to the fact that much of young children's word learning occurs in situations where there are only occasional requests for children to retrieve a new word, and children's spontaneous attempts to retrieve a word – especially when unsuccessful – are difficult to document. With increased recognition of the importance of word exposure to children's language and academic achievement (Hart & Risley, 1995), and with active encouragement to provide such exposure (Suskind & Suskind, 2015), it is imperative to understand the ways in which retrieval may function during the learning process. In this study, we have developed a word learning procedure that incorporates repeated retrieval with contextual reinstatement to examine the facilitative role that this procedure plays in children's learning and retention of new words.

In most studies of word learning by children with DLD, retrieval has been used as an assessment tool to determine what children have learned, but this process is otherwise viewed as a neutral event in the learning process. The role that retrieval can play in *altering* learning is only now receiving attention. The investigation by McGregor et al. (2017) is one such study. McGregor et al. asked young adults with DLD and their typical-language age mates to learn novel words paired with uncommon objects in each of three conditions. Nine words were learned in a repeated study condition with no requests for retrieval during the learning period. Another nine novel words were tested after each word-referent presentation by having the participant retrieve the name of the object. The remaining condition resembled the previous one except that the experimenter provided the participant with the first syllable of the novel word during the retrieval request. The participants were tested on their recall of the novel words the next day.

During the early trials of the learning period, the young adults with DLD were less accurate than their peers in the two conditions involving retrieval. However, on the recall test the next day, the two groups no longer differed for the novel words learned in these particular conditions. In contrast, the participants with DLD had poorer recall than their peers for the novel words learned in the repeated study condition. McGregor et al. (2017) interpreted these findings as indicating that the participants with DLD had word form encoding weaknesses but that retrieval processes allowed them to close the gap with their peers for novel words assigned to the retrieval conditions. Retention of adequately encoded words did not appear to be a problem for the individuals with DLD.

In the present study, we ask whether preschoolers with DLD along with their typically developing age mates retain novel words more successfully if they engage in repeated retrieval with contextual reinstatement during learning than if they engage in repeated study. A within-participant design is used, with each child tested on novel words learned in a repeated retrieval condition and novel words learned in a repeated study condition. The novel words in the two conditions are presented an equal number of times, thus controlling for amount of exposure. We ask the children to learn both the word forms and their meanings. Shorter- and longer-term (one week) retention are examined.

Based on the previous literature, we expect that both groups of children will have more difficulty with word forms than with meanings but the children with DLD will be more likely to lag behind their peers in the learning of word forms. Any differences between the groups in this regard will be evident early in the learning process, implicating relatively weak encoding. For words adequately encoded during the learning period, longer-term retention should be on par with that of the typically developing children.

Method

Participants

The research was approved by the first author's Institutional Review Board; parents provided informed written consent and children provided verbal assent. We recruited 10 children with DLD (4 female, 6 male) and an equal number of typically developing (TD) children (4 female, 6 male). Six of the children in the DLD group were White, 3 were of mixed race (African-American/White) and 1 child was Asian-American. For the TD group, 9 children were White (including one child who was Hispanic), and 1 child was African-American. Each child in the TD group was within 2 months of age of a child in the DLD group, resulting in similar age ranges and mean ages (DLD *M* = 63.40, *SD* = 6.20; TD *M* = 63.20, *SD* = 4.89), t (18) = 0.08, p = .937. As noted by parent report, mothers of the children with DLD had an average of 15.10 years of formal education (*SD* = 2.23); the average for the mothers of the TD children was 16.90 (*SD* = 2.56), t (18) = 1.67, p = .111. Prior to their participation in the study, all children in the DLD group were enrolled in an intervention program and/or their parents had expressed concern about their children's language development. Their scores on the *Structured Photographic Expressive Language Test – Primary 2* (SPELT-P2; Dawson, Stout, Eyer, Tattersall, Fonkalsrud, & Croley, 2005) were below 87 (*M* = 74.70; *SD* = 12.48), the empirically-derived cutoff reflecting good sensitivity and specificity (Greenslade, Plante, & Vance, 2009). These children passed a hearing screening, and scored in the "nonautistic" range on the *Childhood Autism Rating Scale – Second Edition* (CARS-2; Schopler, Van Bourgondien, Wellman, & Love, 2010). No history of neurological impairment was reported. All children with DLD scored within one standard deviation of the mean or higher on a test of nonverbal intelligence (*M* = 108.40, *SD* = 12.14). For five of the children, the test instrument was the *Primary Test of Nonverbal Intelligence* (PTONI; Ehrler & McGhee, 2008). For the remaining children, the instrument was the *Kaufman Assessment Battery for Children: Second Edition* (KABC-II; Kaufman & Kaufman, 2004). Although existing vocabulary tests have not demonstrated adequate sensitivity and specificity to serve as a diagnostic indicator of language impairment (Gray, Plante, Vance, & Henrichsen, 1999; Spaulding, Hosmer, & Schechtman, 2013), we administered the *Peabody Picture Vocabulary Test – Fourth Edition* (PPVT-IV; Dunn & Dunn, 2007) to obtain an estimation of the children's receptive vocabulary. One child did not receive this test. All but one of the other nine children scored within one standard deviation of the mean on this test (M = 97.67, SD = 9.70), a common finding in the literature on word learning, as noted earlier. As also observed in previous studies, the children in our DLD group scored significantly lower on this test than the children in our TD group (see below), t (17) = 3.26, p = .005, with a large effect size d of 1.51.

The children in the TD group scored within normal limits on the SPELT-P2 (M = 118.90; SD = 7.48) and the nonverbal intelligence measure (PTONI or KABC-II; M = 121.60, SD = 17.06). They also passed a hearing screening. No history of neurological impairment was reported. Scores on the PPVT-IV averaged 115.20 (SD = 13.21). The CARS-2 was not administered to the children in the TD group.

Procedure

The children learned eight novel consonant-vowel-consonant (CVC) words from Storkel and Lee (2011) in two sets of four words. The novel words were /dɔɪk/, /paɪb/, /gɪf/, /nɛp/, /faʊn/, /jʌt/, /wæd/, and /bog/. No two novel words shared the same initial consonant, vowel, or final consonant. Within each set, half of the words were assigned to the repeated retrieval with contextual reinstatement (RRCR) condition and half to a repeated study (RS) condition, with the assignments counterbalanced across the children in each group. Words from the two conditions appeared in alternating order. The novel words were chosen from Experiment 2 of the Storkel and Lee study, which compared words that came from sparse and dense neighborhoods, but that matched on phonotactic probability (measured as positional segment sum and biphone sum). We assigned these words to the two conditions in our study so that they matched on both phonotactic probability and neighborhood density (combining two words from “dense” and two words from “sparse” neighborhoods in each condition). The referents for the novel words were stimuli used by McGregor (2014). These were color photographs of rare animals and exotic plants whose actual names are generally unfamiliar even to adults.

One set was learned at a time, with one week separating the completion of testing for one set and the introduction of the second set. Using computer presentation of pictures and audio files, the four novel words in each set were presented in four blocks, with two blocks presented on each of two consecutive days and a five-minute break between the two blocks presented on the same day. The duration of each block was approximately 10 minutes. An illustration of a block appears in Figure 1a (for a word in the repeated retrieval condition) and Figure 1b (for a word in the repeated study condition). In each study trial (applying to words in both conditions), the child saw the picture and heard a pre-recorded three-sentence sequence that provided the novel word and its “definition,” that is, what the animal or plant “liked,” as in “This is a /nɛp/. It’s a /nɛp/. A /nɛp/ likes birds.” In each retrieval trial (applying to words in the repeated retrieval condition only), the child saw the picture and heard the question “What’s this called? What do we call this?” Upon responding, the child then heard (with the picture still present) “What does this one like? What does it like?” We ensured that the “definitions” could not be related to the physical characteristics of the referent or any other detail in the photograph. Both animals and plants were described as liking particular types of weather or particular animals. A study trial for the same word immediately followed every retrieval trial (e.g., “This is a /nɛp/. It’s a /nɛp/. A /nɛp/ likes birds.”).

To strike an ideal balance between retrieval success and contextual reinstatement, we varied the spacing so that, in each block, for words in the RRCR condition, the first retrieval attempt at each word occurred immediately after the first study trial for that word, that is, with no other novel words intervening (see Figure 1a). Thereafter, subsequent retrieval of each word in the RRCR condition occurred after three other novel words (each with its own study, or study and retrieval trial, according to condition) had been presented (Figure 1a). We refer to this spacing schedule as 0-3-3, reflecting the number of other words that had intervened between study trials and retrieval trials for the same word. The “0” trial was used to minimize forgetting on the first retrieval attempt, and the equal spacing of retrieval trials thereafter (3-3) was used because such spacing after the first retrieval trial appears to benefit longer-term retention as much as or more than an expanding retrieval schedule (e.g., 2-4) (Karpicke & Roediger, 2007a). Encouragement was provided but the children were not told if their retrieval attempts were correct. The words in the RS condition followed the same study schedule as those in the RRCR condition but included no retrieval trials (see Figure 1b).

Across the four blocks, each word in each condition was heard a total of 48 times (a three-sentence sequence in each of 16 study trials), thus equating the two conditions in terms of the number of times the novel words (48) and definitions (16) were heard. The children were tested five minutes after the fourth block by being shown each picture on the screen and being asked what it is called, and then what it likes, following the same wording as used during retrieval trials. Two test items were used for each word and definition, with the two items always separated by several intervening items assessing other words and definitions. Following the recall test, the children were tested on a receptive, multiple-choice task, referred to here as a form-referent link recognition task (see Gordon & McGregor, 2014). The child was asked to point to the picture on the screen that matched the novel word (e.g., “Where’s the /nɛp/?”) from an array of three pictures. The foils were pictures of other novel word referents. One week later, the children were tested again on both the recall test and the form-referent link recognition task, in that order.

Scoring and Reliability

Novel word productions on the recall tests were scored in terms of the number of items in each condition for which the child was judged to have responded with the correct word. Several criteria were used in the scoring. First, for the production to be scored as an attempt at the target, it could not resemble a real word that might plausibly be used for the novel referent. Second, the production had to be subjectively judged to be a plausible attempt at the target. At this stage, we consulted the children's errors on our real-word probes that were designed to resemble our novel words in segment and syllable shape composition, in case there were unusual patterns that should be taken into account in making this decision. Then, we adapted the Edwards et al. (2004) scoring system of crediting each consonant with one point for each of place, manner, and voicing, and each vowel for the dimensions of backness, height, and length. An additional point was awarded for preserving prosodic shape (CVC). We then required that the production have a higher score than the score that would be credited if the production was actually an attempt at any other novel word in the set. For example, the production /gɪp/ for the target /gɪf/ would earn 9 points (3+3+2+1) and it if it were an attempt at the wrong word, /paɪb/, it would earn 6 points (1+2+2+1). Because we chose novel words that were quite different from one another, novel words that were subjectively judged as attempts at the correct word were only rarely scored as incorrect based on the Edwards et al. scoring system. Phonetic criteria were not applied to the scoring of meaning, because the words constituting the "definitions" (e.g., *birds, butterflies, rain*) were readily interpretable even when productions fell short of adult standards. For data analysis, we combined the number of words (and, separately, the number of meanings) recalled by each child across the two sets, as preliminary analyses revealed no interactions involving set. In a separate analysis, we also scored the children's retrieval attempts for words and definitions in the RRCR condition during the learning period. The same scoring criteria were applied as during recall testing.

To assess scoring reliability for word forms, two judges independently scored the five-minute and one-week recall tests of four children in each participant group. Reliability was computed by comparing the judgments of all responses scored as correct by at least one of the two judges. Agreement was 94%. To assess scoring reliability for meanings, the "definitions" provided by four children in each participant group on the five-minute and one-week recall tests were scored by two independent judges. Each response was scored as correct and incorrect by the two judges and the percentage of items on which agreement occurred was calculated. Agreement was 99%.

Data Analysis

A series of mixed-effects models were estimated with a random intercept at the child level, with repeated measures nested within child. Random slopes for time and learning condition were included when they were not close to zero. For each measure (word form recall, meaning recall, form-referent link recognition), a model with participant group (DLD vs. TD), learning condition (RRCR vs. RS) and time (five minutes vs. one week) were included as variables. We also present a model that included PPVT-IV standard scores and maternal education as covariates along with the three main study variables. In each analysis of the data, a random slope for time was not significant; however, the random slope for condition was statistically significant, indicating that the learning condition effect varied substantially between children. We tested additional models that included two- and three-way interactions with and without the covariates. Only those that proved informative are presented here; additional models are presented in tables appearing in Supplemental Materials. In the tables presented here and in Supplementary Materials, Model A examines main effects; Model B deals with main effects and covariates; Model C includes main effects and two-way interactions; Model D examines main effects, two-way interactions, and covariates; Model E is concerned with main effects, two-way interactions, and the three-way interaction; and Model F includes main effects, two-way interactions, the three-way interaction, and covariates. For the models presented in Supplementary Materials, the added terms were not significant and did not improve model fit. In all tables, effect sizes are reported in terms of partially standardized beta coefficients (b_{std}).

Results

Word Form Recall

An illustration of the children's five-minute and one-week recall of word forms is provided in Figure 2. Table 1 provides a summary of the models applied to word form recall. A learning condition effect revealed that scores were approximately 2.6 points higher for the RRCR condition than for the RS condition (with a large effect size, $b_{std} = 1.00$). This effect was 2.8 points when controlling for PPVT-IV scores and maternal education ($b_{std} = 1.09$). There were no two- or three-way interactions, indicating that the learning condition effect did not differ by group or across time. For both five-minute and one-week recall, all but one of the children with DLD recalled more words in the RRCR condition than in the RS condition (the remaining child showed the reverse pattern). For the TD group, seven of the 10 children showed greater recall in the RRCR condition (one child showed the reverse). From Figure 2 and Table 1 it can be seen that the numerical trend toward higher scores for the TD group was not statistically significant. (Indeed, if the observed numerical trend were to continue with additional participants, approximately 65 children in each participant group would be required to reach significance.) Figure 2 also reveals considerable stability in the children's recall from the five-minute to the one-week test. Further inspection of the data indicated that the recall differences for word form favoring the RRCR over the RS condition held true for each set (Set 1: t (19) = 3.45, p = .003, d = 0.95; Set 2: t (19) = 4.09, p < .001, d = 1.09).

Although phonetic criteria were applied to crediting the children with the production of each novel word, we also determined whether, for the novel words judged as having been produced by the children, there was a difference in degree of phonetic accuracy, again with 10 constituting the highest possible score for each word credited. We found only a difference for participant group – the TD children (M = 9.70, SD = 0.37) were more accurate than the children with DLD (*M* = 9.32, *SD* = 0.70), F (1, 18) = 8.82, p = .008, d = 0.68. Mean accuracy was never below 92% for any combination of learning condition, testing time, and participant group.

In Table 2, we provide descriptive data reflecting the children's overall word form accuracy during RRCR retrieval trials over the course of the learning period. The children's productions of the appropriate novel word were much more likely during a "0" trial than during a "3" trial, owing, no doubt, to the fact that for "0" trials, retrieval occurred immediately after the preceding study trial. Table 2 also shows the increase in performance from the earliest to the latest "3" trials. As in the recall tests, accuracy was slightly higher for the TD children, though the two groups of children seemed quite similar in the pace of their accuracy gains once retrieval trials were repeated. It was also observed that the degree of change in retrieval accuracy from the fourth "3" trial (occurring on the first day) to the fifth "3" trial (occurring on the second day) was not appreciably different from the same-day change from the second "3" trial to the third "3" trial, that also had an intervening "0" trial. This was true for both groups of children.

Meaning Recall

An illustration of the children's five-minute and one-week recall of meaning is provided in Figure 3. Table 3 provides a summary of the model analyses for meaning recall. As we found for word form recall, there was a significant learning condition effect for meaning recall, where scores for the RRCR condition were about 1.1 points higher than scores in the RS condition (b_{std} = 0.66). This difference was 1.2 after controlling for the two covariates (b_{std} = 0.70). There were no interactions, indicating that the learning condition effect did not differ by participant group or time. Six of the 10 children with DLD had better recall in the RRCR condition than in the RS condition at both testing periods. This was true for four of the 10 children in the TD group. All exceptions were cases of equivalent scores for the two conditions (usually maximum scores); there were no children in either group who showed higher scores in the RS condition than in the RRCR condition.

Recall scores were considerably higher for meaning than for word form, and it is possible that learning condition effects (though statistically significant) were minimized due to ceiling effects. This seems especially true for the RRCR condition, where eight children in the DLD group and six children in the TD group earned the maximum score on both the five-minute and one-week test. An additional child in the DLD group and three additional children in the TD group earned the maximum score at one of these test periods.

Table 4 provides descriptive data reflecting the children's overall accuracy for meaning during RRCR retrieval trials over the course of the learning period. As expected, the children were more accurate during a "0" trial than during a "3" trial. Although the children's accuracy across the trials for meaning was higher than their accuracy for word forms (see Table 2) during the same trials, the first trial (a "0" trial) is a notable exception. The DLD and TD groups' mean accuracy levels were 2.60 and 2.90, respectively, for the first meaning trial but 3.60 and 3.90, respectively, for the first word form trial. Note that there were phonetic criteria for crediting the children for word form accuracy, but for meaning accuracy, no such criteria were used, as rough approximations were readily identifiable (e.g., /bʌwəfaɪz/ for *butterflies*). The lower accuracy for meaning on this first trial seems most likely due to the children's primary attention being drawn to the novel name, which probably rendered what the plant/animal "liked" quite secondary in salience. That is, hearing "This is a /nɛp/. It's a /nɛp/. A /nɛp/ likes birds." the children might well have focused on /nɛp/ and paid less attention to "birds." Judging from the remaining trials, however, it is clear that remembering what each plant/animal liked was less burdensome than remembering the name of the plant/animal itself. A comparison of Tables 2 and 4 also reveals that the difference favoring the TD group during the word form trials was much narrower during the meaning trials.

Form-Referent Link Recognition

The relative difficulty experienced by the children in word form recall was not present when testing shifted to a multiple choice task requiring the children to recognize the referent associated with each word form. There were no differences due to learning condition, time, or participant group, and no interactions (*p*s $\geq$ .189). It is possible that our recognition task was insensitive to potential differences due to ceiling effects. Five of the 10 children with DLD and eight of the 10 children in the TD group had 100% accuracy at both time points, and in several other cases, only a single error was observed.

Discussion

Word Form

The main goal of this inquiry was to determine if RRCR provided benefits to word learning and retention that exceeded the benefits seen with the more customary method of repeated study. This was clearly the case for word forms; recall was dramatically better in the RRCR condition. This was true for the recall test five minutes following the learning period and for the test one week later.

The advantage of RRCR is presumed to lie in the enhanced representation created through retrieval with slightly changing contexts – a condition created when retrieval must occur after some change from a preceding encoding context. This was operationalized in the present study by the insertion of three other items between each word's study trials and its retrieval trials (hence, 0-3-3).

Although large effects were seen for word form recall, differences were not found for performance on the form-referent link recognition task. Ceiling effects obstructed our ability to test learning condition differences on this task. It appeared that the recognition task required only the ability to store a rough approximation of each novel word's representation in order to succeed, given that the novel words differed considerably in their phonological composition.

Earlier word learning studies (using procedures different from ours) indicated that children with DLD require more word exposures than their peers before reaching a designated criterion level (e.g., Alt, 2011; Alt & Plante, 2006; Alt et al., 2004; Gray, 2003, 2004; Gray et al., 2014; Leonard et al., 1982; McGregor, Licandro et al., 2013; Rice et al., 1994). In our study, children in both groups heard each word the same number of times, yet we did not observe any group differences. Although these results are encouraging, we do not believe that our DLD and TD groups were equivalent in ability. In particular, we suspect that the children with DLD had somewhat weaker word form encoding skills.

In Table 2, we saw how, at each retrieval trial, the children with DLD were slightly less accurate than the TD children. The fact that this difference was apparent from the first trial is suggestive of weaker encoding. Furthermore, even for those novel words judged to have been successfully recalled, the DLD group was less accurate in their productions than their same-age typically developing peers.

In RRCR, a word's representation is enriched through the accumulation of distinctive contextual features through successful retrieval. Despite their encoding weaknesses, the children with DLD were able to take advantage of this type of enhancement. In fact, the general lack of participant group by learning condition interactions in our experiments indicates that this facilitative process operated in the children with DLD as effectively as in their peers.

The children with DLD were also as successful as their peers in retaining word forms from five-minute to one-week testing. As Figure 2 illustrates, these children's retention was quite stable from one testing point to the next. We do not interpret these findings as indicating age-appropriate long-term retention on the part of the children with DLD. First, we did not test the children beyond one week. Second, and crucially, our task was quite simple; relatively few novel words had to be learned, all novel words were monosyllabic, and each novel word was associated with an easily depicted referent. Thus, although we suspect that encoding was the weakest ability in these children, additional manipulations might well have produced group differences in the stability of the children's retention across time.

Meaning

Based on previous work, we expected better recall for meanings than for word forms (e.g., McGregor, Licandro et al., 2013). However, we assumed that the pattern of results would resemble the pattern seen for word forms, with RRCR yielding better recall than RS. This proved true, though the magnitude of the learning condition difference was somewhat smaller for meaning than for word forms. As we found for word forms, the two groups of children were not significantly different, and retention from five minutes to one week was very stable.

The findings might have been influenced by the way we operationalized “meaning” in our study. The children’s task was not to tell us what, say, a /nɛp/ liked, but rather what the picture representing a /nɛp/ liked. The task was verbal in that the answers such as “birds” are words, but these were definitions of exotic plants and animals appearing in pictures, not definitions of words. Still, we were impressed with how readily the children were able to learn the associations between these very arbitrary definitions and the depicted plants and animals. On the other hand, we do not claim that the children with DLD were free of weaknesses even in this task. In Table 4, we saw that these children had slightly lower accuracy than their TD peers across trials. These differences were less obvious than the ones for word form (Table 2), but they leave open the possibility that encoding weaknesses extended beyond word form, albeit at a more subtle level.

Implications and Next Steps

We believe the main findings of this study make a crucial point. As seen in Figures 2 and 3, the RRCR condition helped the children with DLD achieve levels of word form and meaning retention that were comparable to or exceeded the levels seen for the TD children in the more typical learning condition of repeatedly hearing the word in the presence of the referent (repeated study). Because age-matched TD children represent a type of gold standard, this finding is noteworthy, especially considering that the procedure used in the RRCR condition did not require extra word exposures for the children with DLD.

There are also some practical implications of the work reported here. The poorer recall for words in the repeated study condition relative to those in the RRCR condition may not have been surprising given the historical literature with adult participants. However, this finding takes on added importance in the case of word learning in younger children. For example, the "30-million-word" gap between children receiving different levels of input (see Hart & Risley, 1995; Suskind & Suskind, 2015) refers to words the children hear, not those that they retrieve and produce. Based on the results of this study, it seems reasonable to suspect that, given the same amount of word exposure in the same contexts, children given opportunities to use these words will learn them more readily than those who simply hear them.

The apparent differences between the learning of word forms as opposed to meanings suggest that vocabulary building activities will not be a uniform process. Children may readily learn a meaning and therefore give the impression that the new word has been established in their lexicons. However, the *form* of that word may still be fragile, with a phonological representation that is too weak to permit reliable use. Practitioners will need to keep this distinction in mind when establishing criteria for success.

Much more investigation of RRCR should be pursued to fully understand the retrieval process. The ideal number of words that can to be learned at the same time using RRCR must be discovered, as well as the ideal retrieval schedule (e.g., 0-3-3, 0-2-2, or other) for these words. Although learning and retention were clearly better in the RRCR condition than in the comparison condition, the number of novel words used in the study was rather low. Efforts to increase this number should be made. Future research should also determine whether RRCR is just as effective when new verbs and adjectives are the vocabulary targets rather than nouns. Another important avenue for study is the discovery of whether RRCR facilitates not only the learning of new words but children's ability to apply these words to new referents – their ability to generalize.

There is yet another property of RRCR that requires investigation. Although retrieval seems to be a crucial component for learning, the degree of contextual reinstatement during retrieval can differ according to the task. Contextual reinstatement is assumed to occur when a representation is updated by including features of an earlier context at the time the material is being retrieved again. Such reinstatement is usually promoted by requiring retrieval of material after other material has intervened – the procedure followed for the RRCR condition used in the present study. Note, however, that retrieval can occur with little to no reinstatement if it occurs immediately after the previous exposure. In this case, the context has changed very little, and therefore such immediate retrieval will not lead to an enriched representation, minimizing the likelihood of successful longer-term retention. In the companion article that follows – *Retrieval-Based Word Learning in Young Children with Typical Language Development and Children with Developmental Language Disorder II: A Comparison of Retrieval Schedules* (Haebig et al., under review) – we compare these two types of repeated retrieval conditions in an effort to optimize children's word learning.

Although there is still much to understand about RRCR, the study reported here represents an important start. It reveals a type of process that seems more beneficial to word learning than common alternatives such as repeated study, and is applicable to children with typical language development as well as children with DLD. Further refinements in the study of RRCR seem likely to lead to even greater benefits to word learning.

Acknowledgments

The authors express appreciation to the participating children and their families, and to the members of the Child Language Lab for their valuable assistance during this project: Megan Miller, Janell Blunt, Anna Redmaster, and Erin Boyle.

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.

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.

Dunn, L, & Dunn, D. (2007). *Peabody Picture Vocabulary Test – Fourth Edition*. 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.

Ehrler, D., & McGhee, R. (2008). *Primary Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.

Fritz, C., Morris, P., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology*, *60*, 991-1004.

Gordon, K., & McGregor, K. (2014). A spatially supported forced-choice recognition test reveals children's long-term memory for newly learned word form. *Frontiers in Psychology*. doi.org/10.3389/fpsyg.2014.00164.

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.

Hart, B., & RIsley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. 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 25*.

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. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, *21*, 157-163.

Karpicke, J., & Blunt, J. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772-775.

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. (2007a). 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. (2007b). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151-162.

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*. Circle Pines, MN: American Guidance Service.

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., 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. (2014). Deficits in word form encoding characterize developmental learning disability. Paper presented at the Symposium on Research in Child Language Disorders, Madison, WI.

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 and 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.

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.

Raaijmakers, J., & Shiffrin, R. (1981). Search for associate memory. *Psychological Review*, *88*, 93-134.

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.

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

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 and Communication Disorders*, *37*, 253-271.

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., & Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, *26*, 191-211.

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

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

Table 1.

Model Results for the Word Form Recall Outcome

	Model A			Model B		
fixed effects	b	95% CI		b	95% CI	
group (DLD vs. TD)	-0.58	-2.24	1.08	-1.27	-3.60	1.06
condition (RR vs RS)	2.58	1.43	3.72	2.82	1.70	3.93
time (1wk vs. 5min)	-0.23	-0.63	0.18	-0.24	-0.66	0.19
PPVT				-0.04	-0.12	0.05
mother's education				0.05	-0.35	0.46
intercept	3.28	2.09	4.47	6.59	-4.65	17.82
random effects	σ^2			σ^2		
condition	6.01	2.90	12.47	5.28	2.45	11.36
intercept	3.19	1.52	6.70	3.53	1.58	7.91

Note. *N* = 20, observations = 80. Sample size for model with covariates (Model B) = 19 children and 76 repeated observations. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05.

Table 2.

Mean Number of Word Forms Accurately Retrieved on Each Trial in the Repeated Retrieval with Contextual Reinstatement Condition.

Group	Day 1						Day 2					
	0	3	3	0	3	3	0	3	3	0	3	3
DLD	3.60	0.80	1.10	3.70	1.60	1.60	3.80	2.10	2.10	3.90	2.70	2.50
	0.70	*0.92*	*0.99*	*0.48*	*1.17*	*1.43*	*0.42*	*1.20*	*1.60*	*0.32*	*1.42*	*1.51*
TD	3.90	1.00	1.60	4.00	2.50	2.50	4.00	3.10	3.10	4.00	3.20	3.40
	0.32	*1.25*	*1.07*	----	*1.27*	*1.18*	----	*1.10*	*0.99*	----	*0.92*	*0.97*

Note. Standard deviations appear in italics. “0” trials were trials in which retrieval was required immediately after a study trial; “3” trials were trials in which retrieval was required after three other words had intervened since the last study trial for the word to be retrieved. Maximum score = 4.

Table 3.

Model Results for the Meaning Recall Outcome

	Model A			Model B		
fixed effects	b	95% CI		b	95% CI	
group (DLD vs. TD)	-0.34	-2.00	1.33	-0.41	-2.65	1.82
condition (RR vs RS)	1.10	0.36	1.84	1.16	0.39	1.92
time (1wk vs. 5min)	-0.05	-0.32	0.22	-0.05	-0.34	0.23
PPVT				-0.03	-0.11	0.05
mother's education				0.15	-0.24	0.53
intercept	6.67	5.48	7.85	7.83	-2.94	18.59
random effects	σ^2			σ^2		
condition	2.45	1.16	5.17	2.50	1.15	5.42
intercept	3.42	1.71	6.86	3.43	1.60	7.39

Note. *N* = 20, observations = 80. Sample size for model with covariates (Model B) = 19 subjects with 76 repeated observations. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05.

Table 4.

Mean Number of Meanings Accurately Retrieved on Each Trial in the Repeated Retrieval with Contextual Reinstatement Condition.

Group	Day 1						Day 2					
	0	3	3	0	3	3	0	3	3	0	3	3
DLD	2.60	3.00	3.20	3.60	3.80	3.50	3.60	3.50	3.50	3.60	3.70	3.80
	1.07	*0.81*	*0.92*	*0.70*	*0.42*	*0.97*	*0.70*	*0.97*	*0.85*	*0.70*	*0.67*	*0.42*
TD	2.80	3.20	3.30	3.70	3.80	3.70	3.80	3.80	3.70	3.80	3.90	4.00
	1.23	*0.79*	*0.95*	*0.48*	*0.42*	*0.48*	*0.63*	*0.42*	*0.67*	*0.42-*	*0.32*	----

Note. Standard deviations appear in italics. “0” trials were trials in which retrieval was required immediately after a study trial; “3” trials were trials in which retrieval was required after three other words had intervened since the last study trial for the word to be retrieved. Maximum score = 4.

Figure Captions

Figure 1. (a) An example of a block showing a novel word /nɛp/ assigned to the Repeated Retrieval with Contextual Reinstatement (RRCR) condition. In each block, each novel word is retrieved in three instances. Retrieval is immediate in the first retrieval trial with no words intervening between the retrieval and the preceding study trial (“0”). For the next two retrieval trials of the word, three other words intervened between the retrieval trial and the preceding study trial of the same word (“3”). (b) An example of a block showing a novel word /paɪb/ assigned to the Repeated Study condition. Again, three other words intervene between appearances of each word, but only study trials are employed.

Figure 2. The mean number of word form items correct on the recall test at five minutes and one week for novel words in the Repeated Retrieval with Contextual Reinstatement (RRCR) condition and the Repeated Study (RS) condition by the children with developmental language disorder (DLD) and the children with typical language development (TD). Error bars are standard errors.

Figure 3. The mean number of meaning items correct on the recall test at five minutes and one week for novel words in the Repeated Retrieval with Contextual Reinstatement (RRCR) condition and the Repeated Study (RS) condition by the children with developmental language disorder (DLD) and the children with typical language development (TD). Error bars are standard errors.

Figure 1a

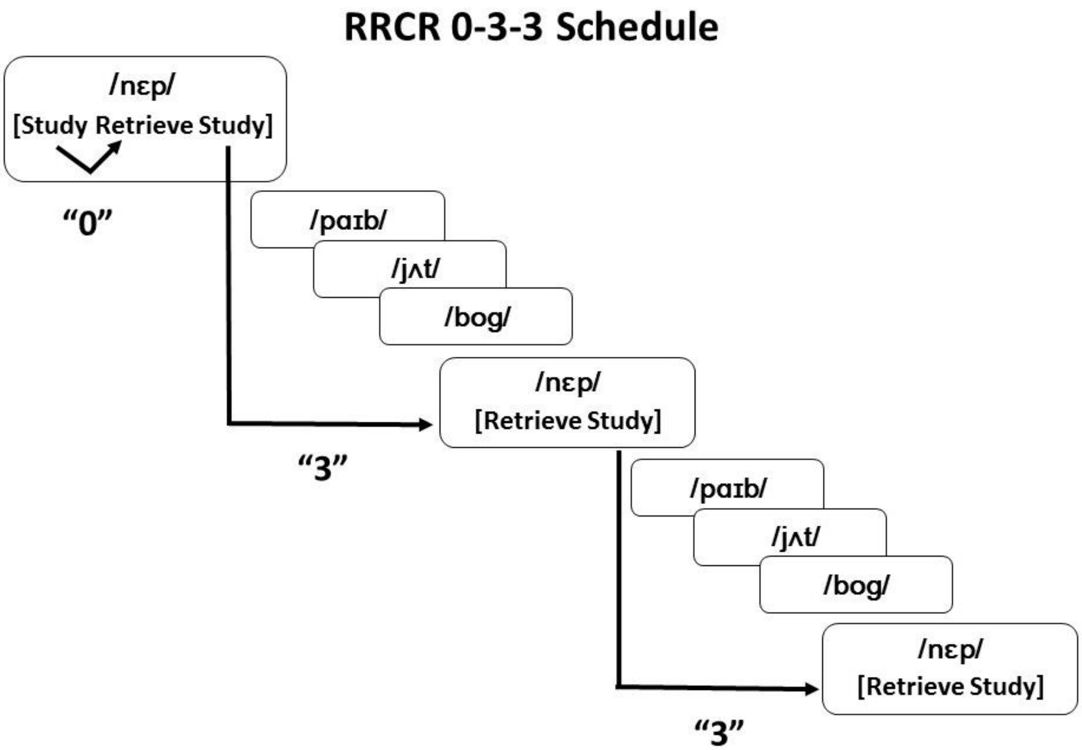

RRCR 0-3-3 Schedule
/nεp/
[Study Retrieve Study]
"0"
/paɪb/
/jʌt/
/bog/
/nεp/
[Retrieve Study]
"3"
/paɪb/
/jʌt/
/bog/
/nεp/
[Retrieve Study]
"3"

Figure 1b

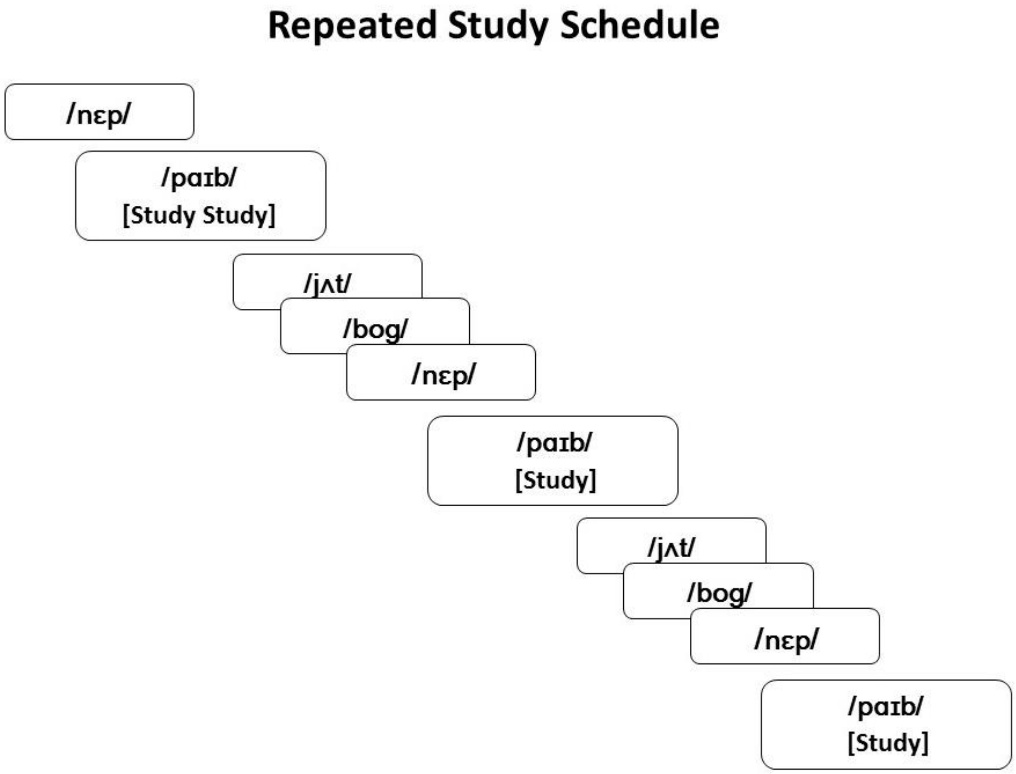

Repeated Study Schedule
/nɛp/
/paɪb/
[Study Study]
/jʌt/
/bog/
/nɛp/
/paɪb/
[Study]
/jʌt/
/bog/
/nɛp/
/paɪb/
[Study]

Figure 2

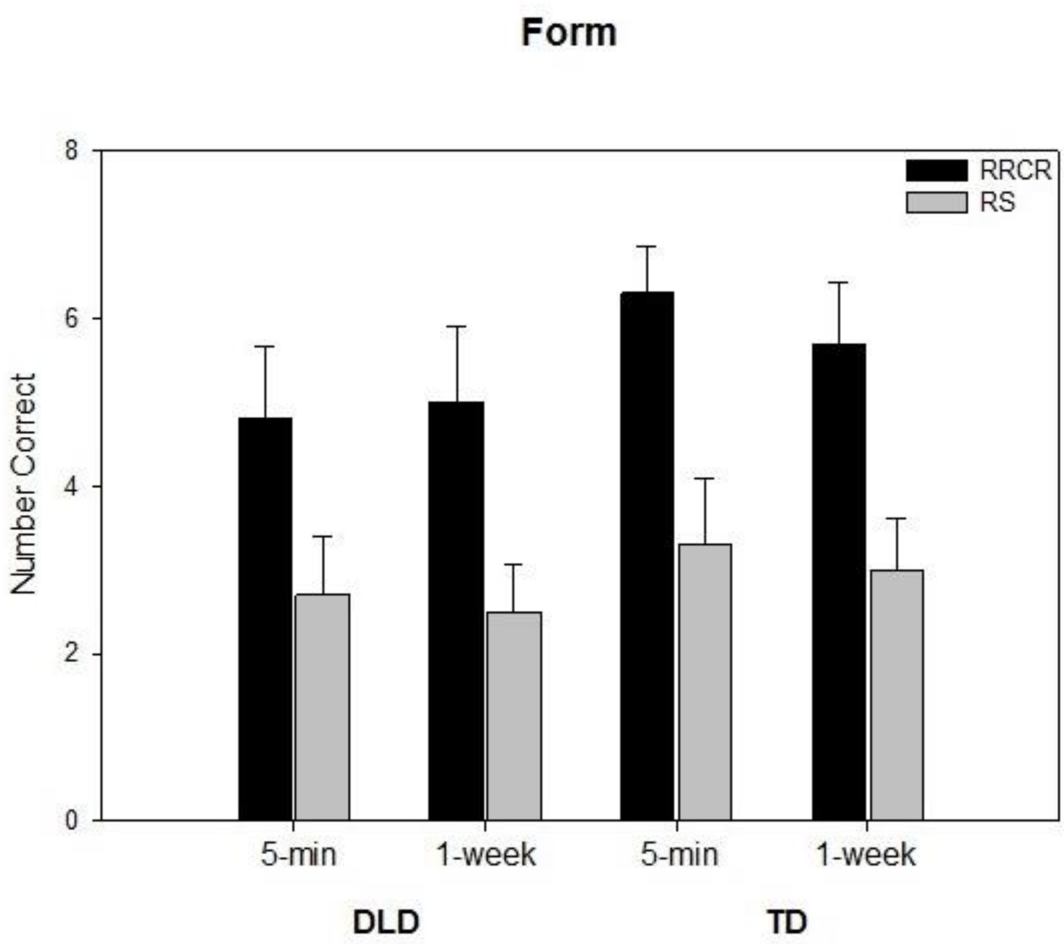
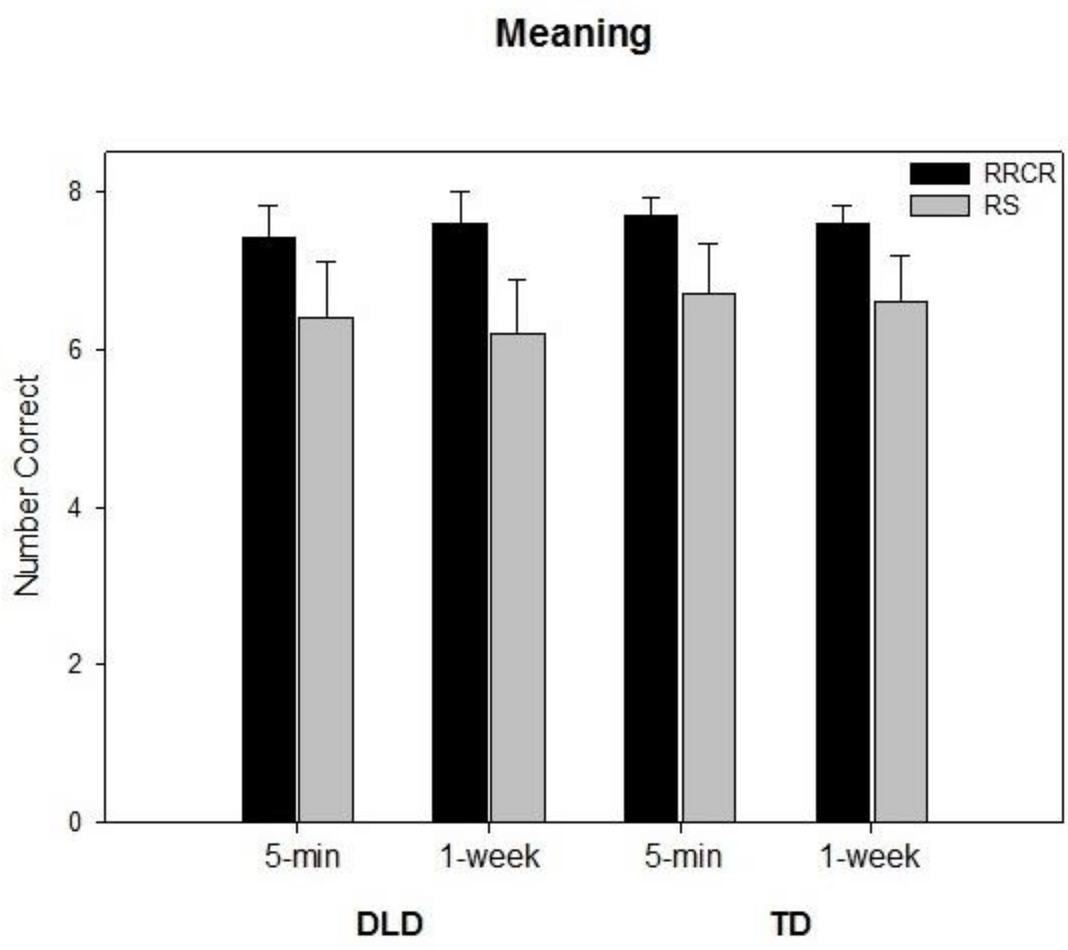

Form
RRCR
RS
Number Correct
0
2
4
6
8
5-min
1-week
5-min
1-week
DLD
TD

Figure 3


Meaning
RRCR
RS
Number Correct
0
2
4
6
8
5-min
1-week
5-min
1-week
DLD
TD

Supplemental Materials

Supplemental Table 1.

Additional Model Results for the Word Form Outcome

	Model C			Model D			Model E			Model F		
fixed effects	b	95% CI		b	95% CI		b	95% CI		b	95% CI	
group (DLD vs. TD)	-0.78	-2.49	0.94	-1.49	-3.87	0.89	-0.60	-2.36	1.16	-1.31	-3.72	1.11
condition (RR vs RS)	2.83	1.12	4.53	2.82	1.19	4.46	3.00	1.25	4.75	3.00	1.31	4.69
time (1wk vs. 5min)	-0.48	-1.18	0.23	-0.48	-1.20	0.25	-0.30	-1.12	0.52	-0.30	-1.14	0.54
group X time	0.45	-0.36	1.26	0.45	-0.41	1.31	0.10	-1.05	1.25	0.08	-1.14	1.29
cond X time	0.05	-0.76	0.86	0.05	-0.80	0.91	-0.30	-1.45	0.85	-0.30	-1.48	0.88
group X cond	-0.55	-2.89	1.79	-0.07	-2.37	2.23	-0.90	-3.38	1.58	-0.44	-2.90	2.01
group X cond X time							0.70	-0.93	2.33	0.74	-0.97	2.46
PPVT				-0.04	-0.12	0.05				-0.04	-0.12	0.05
mother's education				0.05	-0.35	0.46				0.05	-0.35	0.46
intercept	3.39	2.16	4.62	6.70	-4.54	17.96	3.30	2.05	4.55	6.62	-4.63	17.87
random effects	σ^2			σ^2			σ^2			σ^2		
condition	6.280	2.98	13.25	5.61	2.56	12.28	6.26	2.97	13.24	5.59	2.55	12.28
intercept	3.184	1.51	6.71	3.53	1.57	7.92	3.18	1.51	6.71	3.52	1.57	7.92

Note. *N* = 20, observations = 80. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Sample sizes for models with covariates (Model D, Model F) = 19 children and 76 repeated observations.

Supplemental Table 2.

Additional Model Results for the Meaning Outcome

	Model C			Model D			Model E			Model F		
fixed effects	b	95% CI		b	95% CI		b	95% CI		b	95% CI	
group (DLD vs. TD)	-0.40	-2.09	1.29	-0.48	-2.74	1.77	-0.30	-2.01	1.41	-0.37	-2.65	1.90
condition (RR vs RS)	0.90	-0.20	2.00	0.90	-0.22	2.01	1.00	-0.14	2.14	1.00	-0.15	2.15
time (1wk vs. 5min)	-0.20	-0.68	0.28	-0.21	-0.70	0.29	-0.10	-0.65	0.45	-0.10	-0.67	0.47
group X time	0.10	-0.45	0.65	0.10	-0.48	0.68	-0.10	-0.88	0.68	-0.12	-0.95	0.70
cond X time	0.20	-0.35	0.75	0.21	-0.37	0.79	0.00	-0.78	0.78	0.00	-0.80	0.80
group X cond	0.20	-1.31	1.71	0.33	-1.23	1.90	0.00	-1.61	1.61	0.11	-1.56	1.78
group X cond X time							0.40	-0.70	1.50	0.44	-0.72	1.61
PPVT				-0.03	-0.11	0.05				-0.03	-0.11	0.05
mother's education				0.15	-0.24	0.53				0.15	-0.24	0.53
Intercept	6.75	5.55	7.95	7.92	-2.85	18.68	6.70	5.49	7.91	7.86	-2.90	18.62
random effects	σ^2			σ^2			σ^2			σ^2		
condition	2.57	1.20	5.53	2.61	1.18	5.79	2.56	1.19	5.53	2.60	1.17	5.78
intercept	3.42	1.70	6.87	3.43	1.59	7.40	3.41	1.70	6.86	3.42	1.58	7.38

Note. N = 20, observations = 80. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Sample sizes for models with covariations (Model D, Model F) = 19 children and 76 repeated observations.

Supplemental Table 3a.

Model Results for the Recognition Outcome

	Model A			**Model B**		
fixed effects	b	95% CI		b	95% CI	
group (DLD vs. TD)	-0.66	-1.64	0.32	-0.42	-1.39	0.55
condition (RR vs RS)	-0.23	-0.91	0.46	-0.03	-0.54	0.49
time (1wk vs. 5min)	-0.38	-1.02	0.27	-0.18	-0.65	0.28
group X time						
cond X time						
group X cond						
group X cond X time						
PPVT				-0.01	-0.04	0.02
mother's education				0.08	-0.09	0.25
intercept	15.82	15.08	16.56	15.54	10.87	20.20
random effects	σ^2			σ^2		
time	0.89	0.20	3.88	0.21	0.01	4.59
condition	1.19	0.35	4.01	0.45	0.08	2.59
intercept	0.66	0.17	2.58	0.34	0.07	1.61

Note. *N* = 20, observations = 80. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Sample size for the model with a covariate (Model B) = 19 children and 76 repeated observations.

Supplemental Table 3b.

Additional Model Results for the Recognition Outcome

	Model C			**Model D**			**Model E**			**Model F**		
fixed effects	b	95% CI		b	95% CI		b	95% CI		b	95% CI	
group (DLD vs. TD)	-0.28	-1.39	0.84	-0.18	-1.28	0.92	-0.30	-1.53	0.93	-0.42	-1.60	0.75
condition (RR vs RS)	0.03	-1.06	1.11	0.23	-0.61	1.07	0.00	-1.20	1.20	0.00	-0.93	0.93
time (1wk vs. 5min)	-0.18	-1.21	0.86	0.03	-0.76	0.82	-0.20	-1.36	0.96	-0.20	-1.08	0.68
group X time	-0.65	-1.94	0.64	-0.28	-1.24	0.68	-0.60	-2.24	1.04	0.20	-1.08	1.48
cond X time	0.25	-0.74	1.24	-0.16	-1.02	0.70	0.30	-1.12	1.72	0.30	-0.88	1.48
group X cond	-0.75	-2.12	0.62	-0.37	-1.42	0.68	-0.70	-2.40	1.00	0.11	-1.24	1.46
group X cond X time							-0.10	-2.11	1.91	-0.97	-2.68	0.74
PPVT				-0.01	-0.04	0.02				-0.01	-0.04	0.02
mother's education				0.08	-0.09	0.24				0.08	-0.09	0.25
intercept	15.69	14.86	16.51	15.38	10.71	20.05	15.70	14.83	16.57	15.50	10.83	20.17
random effects	σ^2			σ^2			σ^2			σ^2		
Time	0.88	0.19	4.09	0.23	0.01	4.83	0.88	0.18	4.19	0.23	0.01	4.70
condition	1.16	0.33	4.12	0.45	0.07	2.80	1.15	0.32	4.17	0.45	0.07	2.79
intercept	0.66	0.17	2.63	0.33	0.07	1.46	0.66	0.16	2.68	0.33	0.07	1.66

Note. N = 20, observations = 80. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Sample sizes for models with covariates (Model D, Model F) = 19 children and 76 repeated observations.