Retrieval-Based Word Learning in Young Typically Developing Children and Children with Development Language Disorder II: A Comparison of Retrieval Schedules

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

[a] Louisiana State University, Baton Rouge, LA, USA [b]Purdue University, West Lafayette, IN, USA, [c]Boston University, Boston, MA, USA, [d]MGH Institute of Health Professions, Boston, MA, USA, [e]Butler University, Indianapolis, IN, USA, [f]The George Washington University, Washington, DC, USA

Running head: Retrieval-Based Word Learning II

Key Words: developmental language disorders, specific language impairment, word learning, retrieval, electrophysiology

Correspondence to: Eileen Haebig, Louisiana State University, Hatcher Hall Rm 80 Baton Rouge, LA 70803, 225-578-5356, ehaebig1@lsu.edu.

Conflict of Interest

There are no relevant conflicts of interest to report.

Funding Statement

This research was supported by the National Institute of Deafness and Other Communication Disorders, National Institutes of Health, by research grant awarded to Laurence B. Leonard R01-DC014708; Eileen Haebig was supported by a postdoctoral fellowship on training grant T32-DC00030 (PI: Leonard).

Abstract

Purpose: Retrieval practice has been found to be a powerful strategy to enhance long-term retention of new information; however, the utility of retrieval practice when teaching young children new words is largely unknown and even less is known for young children with language impairments. The current study examined the effect of two different retrieval schedules on word learning at both the behavioral and neural levels.

Method: Participants included 16 typically developing children (M_{TD} = 61.58 mo) and 16 children with developmental language disorder (DLD; M_{DLD} = 59.60 mo). Children participated in novel word learning sessions in which the spacing of retrieval practice was manipulated: some words were retrieved only after other words had been presented (i.e., repeated retrieval that required contextual reinstatement; RRCR); others were taught using an immediate retrieval schedule (IR). In Experiment 1, children's recall of the novel word labels and their meanings was tested after a five-minute delay and a one-week delay. In Experiment 2, event-related brain potentials (ERP) were obtained for a match-mismatch task utilizing the novel word stimuli.

Results: Experiment 1 findings revealed that children were able to label referents and to retain the novel words more successfully if the words were taught in the RRCR learning condition. Experiment 2 findings revealed that mismatching picture-word pairings elicited a robust N400 ERP only for words that were taught in the RRCR condition. Additionally, children were more accurate in identifying picture-word matches and mismatches for words taught in the RRCR condition, relative to the IR condition.

Conclusion: Retrieval practice that requires contextual reinstatement through spacing results in enhanced word learning and long-term retention of words. Both typically developing children and children with DLD benefit from this type of retrieval procedure.

Introduction

Children with developmental language disorder (DLD) experience language difficulties that cannot be attributed to hearing loss, intellectual disability, or other neurodevelopmental disorders (Tomblin et al., 1997). In the literature, children with DLD often have been referred to as children with specific language impairment (SLI). Although language profiles within DLD can be heterogeneous, word learning difficulties have been well-documented (Gray, 2004; Haebig, Saffran, & Ellis Weismer, 2017; Kan & Windsor, 2010; Oetting, Rice, & Swank, 1995). Given this, it is important to develop evidence-based clinical practices to optimally teach words to children with DLD. These techniques should be based on current learning theory. Such approaches are supported by previous research and have the potential to help us advance our theoretical and clinical understanding of learning in children with atypical development. The current study presents behavioral and neural findings from a larger project examining the effectiveness of retrieval practice on word learning in preschool children with DLD.

Although there have been a fair number of studies examining word learning in children with DLD, there is a sparsity of word learning interventions to guide clinical practice (Storkel, Voelmle, et al., 2017). Evidence-based word learning procedures include: interactive book reading (Justice, Meier, & Walpole, 2005; Storkel, Voelmle, et al., 2017; Storkel, Komesidou, Fleming, & Swinburne Romine, 2017) and cross-situational statistically-based word learning (Alt, Meyers, Oglivie, Nicholas, & Arizmendi, 2014). These word-learning interventions are still in their infancy and have not all been tested in preschool-age children with DLD. One clinical trial examined an interactive book reading intervention that targeted word learning in school-age children with DLD (Storkel, Voelmle, et al., 2017; Storkel, Komesidou, et al., 2017). In addition, one efficacy study has examined the effectiveness of a cross-situational statistically-based word learning intervention for late talking toddlers (Alt et al., 2014).

Although it is important to examine the effects of these interventions across different developmental stages, it is particularly necessary to study their effectiveness in preschool-age children with DLD. Language impairments are most frequently diagnosed during the fourth and fifth years of life (Leonard, 2014). Furthermore, early deficits in word knowledge often do not resolve with development, but instead persist into early adulthood (Rice & Hoffman, 2015). This is notable because word knowledge is a key predictor of reading and academic success (Catts, Fey, Tomblin, & Zhang, 2002; Lucas & Norbury, 2015; Ouellette, 2006; Quinn, Wagner, Petscher, & Lopez, 2015). Furthermore, word knowledge is associated with social development, with low vocabulary knowledge being linked to low popularity with peers (Gertner, Rice, & Hadley, 1994). Given the strong evidence for early intervention, it is important that we carefully target skills that have been found to be highly associated with positive child outcomes.

Notably, although we have begun to see the promise of the intervention approaches mentioned above, these interventions focus on manipulating the input that children receive during word learning opportunities. In contrast, there also has been promising research that emphasizes the importance of *retrieving* recently-taught information to facilitate learning and longer-term retention (Karpicke & Roediger, 2008; Landauer & Bjork, 1978). For instance, Karpicke, Blunt, and Smith (2016) documented that retrieval-based practice yielded robust learning in typically developing school-age children, regardless of child abilities in reading comprehension and processing speed. Additionally, Goossens et al. (2014) demonstrated that retrieval practice was more effective than study and elaborative re-study for word learning in school-age children. According to Karpicke and Blunt (2011), the act of retrieval is believed to *enhance learning,* instead of merely prompting a report of the knowledge that has been encoded during teaching.

However, all acts of retrieval are not the same. Learning seems most successful when retrieval requires “contextual reinstatement” (Karpicke, Lehman, & Aue, 2014). That is, when retrieving a recently-taught item, one attempts to reconstruct the learning context. Each successful retrieval allows individuals to update the context representation, resulting in an enhanced representation that incorporates features of the prior learning context as well as the current context of retrieval (Karpicke et al., 2014; Lehman, Smith, & Karpicke, 2014). This repeated retrieval process allows an individual to develop an enriched context representation, wherein the features of the item are stored together with features of the unfolding temporal context, which includes the learning context and subsequent study or experience. The more features that are available in the context representation, the more restricted the search set is; thus, the representation can more effectively cue future retrieval.

Research also has indicated that the specific retrieval schedule can influence the effectiveness of retrieval practice. Karpicke and Bauernschmidt (2011) found that varying the relative spacing in schedules (e.g., gradually increasing the spacing between retrieval trials) had no impact on retention; however, absolute spacing mattered. That is, when additional items intervened between retrieval opportunities, a greater benefit was seen for retention. According to the context-based account, the more the context has changed since last retrieval (quantified by number of intervening items), the more likely that new features will be added to the context representation. As previously described, these additional features increase the effectiveness of the context cues. In contrast, immediate retrieval practice schedules, which require little to no contextual reinstatement, show little retention benefit. Although these retrieval studies have been promising, our understanding of the retention benefits of this procedure – referred to here as “repeated retrieval with contextual reinstatement” (RRCR) — is limited, especially in preschool children, with typical and atypical language development. To our knowledge, only one study has examined the benefits of RRCR on word learning in typically developing preschool children (Fritz, Morris, Nolan, & Singleton, 2007). Therefore, in order to strengthen the evidence base of educational and therapeutic practices, it is necessary to extend previous studies to examine the usefulness of RRCR to enhance learning in young children.

The purpose of the present study is to enhance our understanding of retrieval-based learning by providing an important extension to the findings presented in our companion paper, *Retrieval-Based Word Learning in Young Typically Developing Children and Children with*

Developmental Language Disorder 1: The Benefits of Repeated Retrieval (Leonard et al., under review). Leonard et al. conducted an initial investigation of RRCR by first comparing it with repeated study in typically developing preschool children and preschool children with DLD. Children were taught 8 novel words. Four words were taught within a repeated study (RS) condition in which children were exposed to each novel word a total of 48 times and related semantically meaningful information 16 times. An additional 4 words were taught in an RRCR condition, in which the children were exposed to the word form and meaning an equivalent number of times as in the RS condition. The distinction between the RS and RRCR learning conditions was that, in the RRCR condition, children were prompted to retrieve the word and its definition before listening to a study trial. Leonard et al. tested word learning across three tasks: word form recall, meaning recall, and form-referent link recognition. Both children with DLD and typical language development recalled the word form and its meaning significantly more for words that were taught in the RRCR condition, relative to the RS condition. However, there were no significant differences of learning condition in the form-referent link recognition task due to ceiling effects. Leonard et al. (under review) clearly demonstrated that preschool children benefit from retrieval practice and that this benefit is similar for typically developing children and children with DLD. The next natural extension is to explore which retrieval schedules optimize learning.

Current Study

The current study is the first to compare two retrieval-based learning schedules in typically developing preschool children and preschool children with DLD. Although Leonard et al. (under review) demonstrated clear evidence that retrieval practice enhances word learning, relative to repeated study, it is necessary to test different retrieval schedules to design more effective intervention procedures for clinicians. Therefore, in the current study, children were presented novel words in two learning conditions. Half of the words presented to each child involved RRCR. The other half were presented the same number of times and allowed for the same number of opportunities for production but with minimal or no contextual reinstatement. We refer to this condition as the immediate retrieval (IR) condition.

In addition to testing these two retrieval schedules, the current study expands the Leonard et al. (under review) study, by incorporating additional tasks to more thoroughly test for learning differences between the groups and learning conditions. Experiment 1 included behavioral measures of production and comprehension of the newly taught words that align with the measures used in Leonard et al (i.e., picture naming and identification tasks). Experiment 2 examined the underlying neural processes associated with matching and mismatching picture-label pairings of the newly taught words. Specifically, the match-mismatch task included child judgments of the appropriateness of the picture-label pairings and online electroencephalographic (EEG) recordings to measure the underlying neural patterns associated with matching and mismatching picture-label pairs. Of importance, the N400 event-related brain potential (ERP) has been an effective measure to examine the strength of an association between a prime (e.g., picture) and a target (e.g., label) stimulus (Kutas & Hillyard, 1980). Several studies have demonstrated that the amplitude of the N400 captures the frankness of a semantic violation, or the degree of semantic incongruity between stimuli (Federmeier & Kutas, 2001; Kutas & Federmeier, 2011). Therefore, the ERP correlates of semantic processing from our sample of children with typical language development and children with DLD offer insight into the depth of learning novel labels in the two conditions. Given that children with DLD have deficits not only in breadth but also depth of word knowledge (McGregor, Oleson, Bahnsen, & Duff, 2013), it is important to use a multi-level approach to examine how children with DLD learn new words and the ways in which word learning can be enhanced in targeted interventions.

Across the two experiments that we present in the current study, we asked: Does repeated retrieval with contextual reinstatement enhance novel word learning to a greater degree than immediate retrieval practice? Is this advantage seen for longer- as well as shorterterm retention? Do preschool children with DLD resemble typically developing preschoolers in their learning patterns across the RRCR and IR learning conditions and across time?

We expected that both groups would more successfully learn the words in our RRCR condition. As a result, we hypothesized that children would demonstrate higher accuracy in recalling the RRCR words (form and meaning), relative to the IR words. Furthermore, although we expected both groups of children to benefit from the RRCR condition, as in the Leonard et al (under review) study, group differences seemed possible given the well-documented word form encoding limitations in children with DLD (Alt & Plante, 2006). Thus, we hypothesized that children with DLD may learn fewer words relative to their typically developing peers.

Experiment 1

Methods

Participants. Participants included 16 typically developing (TD) children (10 female, 6 male) and 16 children with DLD (10 female, 6 male), who were matched on gender ($\chi^2 = 0.00$, $p = 1.00$), chronological age ($M_{TD} = 61.58$ mo, $SD_{TD} = 5.16$; $M_{DLD} = 59.60$ mo, $SD_{DLD} = 4.43$; $t(30) = 1.20$, $p = .24$), and maternal years of education ($M_{TD} = 16.63$ years, $SD_{TD} = 1.75$; $M_{DLD} = 15.50$ years, $SD_{DLD} = 1.59$; $t(30) = 1.90$, $p = .07$). None of these children participated in the word learning study presented in Leonard et al. (under review). Fifteen children in the TD group were reported to be white and one parent chose not to report the child's race and ethnicity. Fourteen children in the DLD group were reported to be white, two were reported as biracial (one white and Asian American, and one white and African American). None of the children were reported to be Hispanic. This study was approved by Purdue University's Institutional Review Board. All participants provided verbal assent and a parent or legal guardian provided informed written consent.

The children met several selection criteria to be included in the study. All children passed a hearing screening at 20 dB through headphones at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz (ASHA, 1997). Additionally, all children scored within or above one standard deviation of the mean on the nonverbal cognitive assessment, the *Kaufman Assessment Battery for Children – Second Edition* (KABC-II; Kaufmann & Kaufman, 2004; M_{TD} = 115.81, $range_{TD}$: 96 - 133; M_{DLD} = 101.88, $range_{DLD}$: 87 - 118).

The typically developing children performed within or above one standard deviation from the mean on the *Structured Photographic Expressive Language Test – Preschool Second Edition* (SPELT-P2; Dawson et al., 2005; *M* = 113.06, range: 100 - 128). All but two children with DLD earned a standard score of 87 or below on the SPELT-P2 (*M* = 77.21, range: 56 - 89), which is a score that has been empirically determined to be the cutoff point yielding high sensitivity and specificity for children with language impairments at this age (Greenslade, Plante, & Vance, 2009). The two children in the DLD group who scored slightly above this cutoff (89) were retained in the study because their Developmental Sentence Score (DSS; Lee, 1974), was below the 10th percentile. The language sample that was used to derive the DSS was taken during an examiner-child free-play activity. Child utterances were transcribed by a trained coder and checked by a second trained coder; differences were resolved by consensus. The first 50 complete and intelligible utterances which included both a subject and verb were scored according to the DSS guidelines. Lastly, all children with DLD scored in the “nonautistic” range on the *Childhood Autism Rating Scale – Second Edition* (CARS-2; Schopler, Van Bourgondien, Wellman, & Love, 2010). The CARS-2 was not administered to the TD children.

Although not serving as a selection criterion, the *Peabody Picture Vocabulary Test – Fourth Edition* (PPVT-IV; Dunn & Dunn, 2007) was administered to all children. The typically developing children scored at high levels on this measure (*M* = 121.06, range: 106 - 145). The majority of the children with DLD scored within the normal range on the PPVT-IV; however, not surprisingly, they had lower standard scores, relative to the TD children (*M* = 103.44, range: 83 - 124; $t(30) = 4.45$, $p < .001$, $d = 1.56$). Finally, we determined children’s handedness using an abbreviated assessment that prompts children to perform daily living skills (writing, drawing, throwing a ball, pretending to eat, and pretending to brush his/her teeth; *Edinburgh Handedness*

Inventory; Oldfield, 1971). All children were right-handed except one child with DLD who was left-handed.

Word learning task. Children were taught 12 novel words as labels for exotic plants and animals. To prevent fatigue and to promote learning, the 12 words were taught across two sets of similarly structured word learning tasks. The novel words were: /bog/, /nɛp/, /paɪb/, /jʌt/, /daɪbo/, /fumi/, /gine/, /tomə/, /kodəm/, /meləp/, /pobɪk/, and /tɛkət/. Eight of the 12 novel words were disyllabic with syllable-initial stress and four were monosyllabic. Together, these two word types represent approximately 90% of the word tokens that children from two to six years of age hear based on child-directed speech in the CHILDES database (Roark & Demuth, 2000). An even number of each syllable shape (CVCV, CVCVC, CVC) was used, shapes that are well represented in the speech of 5-year-old children in home and preschool contexts based on Hall, Nagy, and Linn (1984). No novel words with the same syllable shape had the same word-initial phoneme. Additionally, the consonants within the novel words consisted of early-emerging sounds that could be easily produced by most preschoolers. Within each word learning set, children were taught six novel words that corresponded to six unfamiliar pictures. Each set consisted of three novel words that were taught in the IR condition and three novel words that were taught in a RRCR condition. Novel word assignments were counterbalanced for learning condition (IR vs. RRCR) across children. Within each learning condition, children learned one of each syllable structure: CVC, CVCV, CVCVC. Stimuli were matched between the learning conditions on syllable shape, phonotactic probability (average biphone frequency), and neighborhood density using the Storkel and Hoover (2010) child language corpora database. Picture referents consisted of colored photographs of exotic plants and animals (used by McGregor, 2014), whose real names are typically unknown by adults. Eight of the pictures and four of the CVC novel words also were used in the Leonard et al. (under review) word learning study.

We presented the stimuli using a computer presentation program wherein a block design was used to present words within each learning condition. Within each set, children completed four blocks; two blocks (IR block and RRCR block) were completed on the first day and an additional two blocks (IR and RRCR) were completed on the second consecutive day. Each block presentation lasted approximately 10 minutes; we provided a five-minute break between each block. The order in which the blocked learning conditions were presented was counterbalanced across children. See Figure 1 for a depiction of IR/0-0-0 and RRCR/0-2-2 blocks.

For all words in both conditions, there were “study” trials and “retrieval” trials. In study trials, the child saw a picture and heard the novel word and its definition (what it “liked”), as in “This is a /daɪbo/. It’s a /daɪbo/. A /daɪbo/ likes rocks.” Thus, for each study trial, the child heard the word form (e.g., /daɪbo/) three times and the definition (e.g., “rocks”) once. The words selected for the semantic information (e.g., “rocks”) were early-acquired words. What was “liked” was arbitrarily paired with a target object; no information contained in the referent picture provided a clue as to what the depicted referent “liked.” In retrieval trials, the child saw the picture and was asked for its name and what it liked, as in “What’s this called? What do we call this?” and (after the child responded with the picture still present), “What does this one like? What does it like?” After each retrieval trial, another study trial was presented that served as feedback (regardless of the child’s accuracy on the retrieval trial). This second study trial was identical to the study trial that preceded the retrieval trial. Each novel word appeared in one block per day; within each block, there were four study trials that presented the novel word three times each within the script. Therefore, the total number of exposures of each novel word was 24 and each word meaning (i.e., what it likes) was heard 8 times; as described below, each word form and meaning had six retrieval opportunities.

For words in the RRCR condition, the first retrieval trial occurred immediately after the first study trial for that word. However, subsequent retrieval trials for that word occurred only after two other words had been presented. This schedule is referred to as a 0-2-2 schedule, which reflects the number of words intervening between study trials and retrieval trials of the same word. Because intervening words create a change in temporal context during “2” trials, the 0-2-2 condition was assumed to promote contextual reinstatement (see Karpicke & Roediger, 2007). For words in the IR condition, all retrieval trials immediately followed a study trial of the same word. This schedule is referred to as a 0-0-0 schedule, because no words intervened between study trials and retrieval trials of the same word. Because the 0-0-0 schedule involved no change in temporal context, limited (or no) contextual reinstatement was assumed. The two conditions were equivalent in both the number of times the word was heard and the number of retrieval opportunities provided for that word.

We should note that having the 0-0-0 and 0-2-2 conditions in separate blocks was not equivalent to the “massed” versus “distributed” learning conditions often described in the memory literature. The blocks for each condition were presented on the first day and repeated on the second day in the same order. Testing did not occur until the end of the learning period on the second day. Thus, the blocks representing the two conditions alternated and blocks representing the same condition never appeared consecutively, as might be expected for a “massed” condition. In addition, some massed-versus-distributed studies emphasize the spacing of study trials in particular (the “inter-study interval”; see review in Cepeda, Pashler, Vul, Rohrer, & Wixted, 2006). In the present experiment, study trials were separated by retrieval trials. This was true for both “0” trials and “2” trials (“0” refers to the absence of intervening words between a study trial and a retrieval trial for the same word, not to consecutive study trials for the same word).

After children completed the learning blocks on the second day, they received an additional five-minute break and then they completed a recall test. Word form recall and meaning recall were tested (e.g., Form: “What’s this called? What do we call this?”, Meaning: “What does this one like? What does it like?”). One week later, children returned and completed the word form and meaning recall tests again; they also completed a form-referent link recognition test. During the form-referent link recognition test, children were presented with an array of four pictures and the child was asked to point to the correct picture (e.g., “Where’s the /pobɪk/?”). One week after completing the recall and recognition tests for set 1, children were introduced to the second set of six words, with procedures identical to those of set 1.

Scoring and reliability. Child responses to the word form recall tests were scored according to the number of accurate responses that children provided within each learning condition. We used several criteria when coding accuracy. To score child attempts to produce the target, we first confirmed that the child production did not resemble a real word that could be used as a plausible label for the novel referent. Next, each production was subjectively judged as being a plausible or implausible attempt at the target. While making this judgment, we consulted each child’s speech errors on our real-word probes that were designed to resemble our novel words in segment and syllable shape composition. Following this, we applied an adapted version of the Edwards, Beckman, and Munson (2004) scoring system, wherein each consonant was assigned a range of 0 to 3 points for the accuracy of its place, manner, and voicing, and each vowel was assigned one point for the dimensions of backness, height, and length. One additional point was given if the child production preserved the prosodic shape of the target (e.g., CVC). Lastly, we required the child production to have a higher score than the score that would have been assigned if the production was an attempt at any of the other novel words in the set. For instance, the production /topɪk/ for the target word /pobɪk/ would earn 14 points (2+3+2+3+3+1), whereas if this production were considered an attempt at the incorrect word, /kodəm/, 9 points would be awarded (2+3+1+2+0+1). Total scores were based on combining each child’s score across the two sets, as preliminary analyses revealed no interactions involving set. Two judges with experience in the phonetic transcription of child speech independently scored the five-minute and one-week recall responses of four children from each participant group to assess reliability. We computed reliability by comparing the judgments of all responses scored as correct by at least one of the two judges. Agreement was

97%.

Child productions for the meaning recall items were scored based on accuracy. Because the focus was on the semantic content of the word, mispronunciations of the target meaning (e.g., “wocks” for “rocks”) were accepted as accurate. Inter-judge agreement was 99%.

In a separate analysis, we also scored the children’s retrieval attempts during learning for the words and definitions. The same scoring criteria were applied.

Data analysis. To address our research questions, a series of mixed-effects models were estimated. In these models, random intercepts were set at the child level and repeated measures were nested within children; additionally, random slopes for time and learning condition were included as appropriate. We included the PPVT-IV scores and maternal education as covariates. Models with interactions are presented when they were statistically significant. Additional models that were estimated can be found in the Supplemental Materials. Lastly, effect sizes are reported using partially standardized beta coefficients (b_{std}).

Results

Word form. Figure 2 provides an illustration of the recall results for word form; Tables 1 and 2 summarize the data analysis. The results revealed a learning condition effect, such that scores were approximately 2.50 points higher in the RRCR/0-2-2 condition than in the IR/0-0-0 condition, both with and without controlling for PPVT-IV scores and maternal education (b_{std} = 0.93, indicating a large effect). Thirteen of the 16 children with DLD had better scores on the 0-2-2 words than on the 0-0-0 words in the five-minute test (two showed the reverse pattern), and 12 of the 16 children showed this pattern on the one-week test (two showing the reverse). One child with DLD failed to recall any words at either time point. Thirteen of the 16 TD children recalled more words in the 0-2-2 condition than in the 0-0-0 condition on both the five-minute and the one-week test (two showed the reverse pattern). Additional analyses revealed that the significant learning condition effect held for each set (Set 1 $t(31) = 4.24$, $p < .001$, $d = 0.93$; Set 2 $t(31) = 3.08$, $p = .004$, $d = 0.62$).

The recall advantage of the words learned in the 0-2-2 over those learned in the 0-0-0 condition is especially noteworthy considering that during the learning period, the children actually produced words in the 0-0-0 schedule (*N* = 1,033) more frequently than words in the 0-2-2 schedule (*N* = 587). This difference occurred because even for the words eventually learned and retained in the 0-2-2 condition, the children were not always successful during the first two or three “2” retrieval trials. Thus, the 0-2-2 schedule led to greater retention *in spite of* these words being produced less frequently during the learning period.

We also found a participant group effect; the scores of the children with DLD were about 1.34 points lower than the scores of the TD group with the covariates included (b_{std} = 0.49). However, this difference should be interpreted within the context of the marginal interaction (p = .056) involving participant group and time (see Models C and D in Table 1). The simple effects (see Table 2) reveal that the DLD group scored approximately 1.60 points lower than the TD group on the five-minute recall test, but did not differ from their TD peers on the one-week test. This was the result of the TD children’s scores dropping by 0.72 points on average between the two testing points while the children with DLD retained the same scores over time.

We also determined whether, for the novel words that were credited to the children, there was a difference in degree of phonetic accuracy. Because the total number of features potentially correct differed according to the length of the novel word, we converted the children’s scores to percentages. We found only a difference for participant group – the TD children (*M* = 89.96 %, *SD* = 13.23 %) were more accurate than the children with DLD (*M* = 81.89 %, *SD* = 13.57 %), *d* = 0.60.

Children’s overall word form accuracy during RRCR/0-2-2 retrieval trials over the course of the learning period also were examined. Descriptively, the children’s productions of the appropriate novel word were much more likely during a “0” trial. For example, during the “0” retrieval trials in the RRCR/0-2-2 training protocol, the TD group had a mean accuracy of 5.43 (*SD* = 1.09) whereas the mean DLD accuracy was 5.13 (*SD* = 1.26; maximum score = 6).

During the first "2" retrieval trial, the TD group had a mean accuracy of 1.13 (*SD* = 1.02) and the DLD group had a mean accuracy of 0.81 (*SD* = 0.98). The TD children were more accurate from the beginning, though the pace of the accuracy gains once retrieval trials were repeated appeared similar in the two groups. Additionally, the two groups were similar in the rate of change from the first to the second day.

Given the participant group differences in word form recall on the five-minute test, it seemed important to determine whether such differences could be attributed to differences between the groups that emerged early in the learning period or because the children with DLD lagged further behind as the learning period proceeded. The IR/0-0-0 condition provided an especially appropriate opportunity to examine this issue because all retrieval trials in this condition were of the "0" type. Accordingly, mixed-effects models were estimated with a random intercept for the child where six repeated trials were nested within child. No random slopes were included for time because these were essentially zero. A linear trajectory and quadratic trajectory were estimated. Participant group differences in the average levels of accuracy, the rate of learning, and the changes in learning rate across time (the quadratic effect) were tested. The best model was a quadratic model where the levels of accuracy differed between the DLD and TD groups, but the rate of change and the quadratic did not differ (see Table 3). An illustration appears in Figure 3. As evidenced by the significant linear slope effect in the model, increases in accuracy averaged about 0.29 words between trials. The rate was lower in early trials and higher in later trials, with increases leveling off around Trial 4. The two participant groups showed no difference in terms of average linear increase or the slowing of the increase across trials. However, across all trials, the children with DLD consistently scored 0.63 words lower than the children in the TD group. A visual inspection of accuracy from Trial 3 to Trial 4 – which occurred on different days – revealed no differences in trajectory for the two groups of children (see Figure 3).

Meaning. We found that recall for meaning was much better than for word form (see

Figure 4). However, a group effect was seen, such that children in the DLD group scored about 1.9 points lower than the children in the TD group, with or without the covariates (b_{std} = 0.79; see Table 4). A learning condition effect was not seen; however, in a model without the condition random slope, the learning condition fixed effect was significant, with children's meaning recall being 0.64 points higher for the RRCR condition (p = .035, b_{std} = 0.26).

To gain insight into the group differences for meaning, we examined the children's retrieval trials during the IR/0-0-0 condition, as was done for word form. Again, mixed-effects models were estimated with a random effect for child and six repeated trials nested within child. We fit the data to a linear model that revealed a significant effect for participant group. Children exhibited a growth rate of 0.11 word meanings between trials on average and this rate did not differ according to group (see Table 5). Children in the DLD group, however, were consistently scoring 0.46 lower across the trials compared to the TD group. Figure 5 provides the trajectories for meaning for the two groups of children.

Form-referent link recognition. An illustration of the form-referent link recognition task results appears in Figure 6. There were significant differences between participant groups, where the children with DLD scored approximately 2.2 points lower than the TD group (b_{std} = 0.94). However, the group effect differed by learning condition. As seen in Tables 6 and 7, the condition effect held for the children with DLD, but not for the TD children, and the group difference favoring the TD group was present for the IR/0-0-0 condition but not for the RRCR/0-2-2 condition. For the children with DLD, scores were 1.06 points higher, on average, for the RRCR/0-2-2 condition than for the IR/0-0-0 condition (b_{std} = 0.45). In contrast, TD children had near-ceiling performance on the IR/0-0-0 and RRCR/0-2-2 items on the form-referent link recognition task.

Discussion

We view the results of Experiment 1 as suggesting that repeated retrieval with contextual reinstatement (the 0-2-2 condition) assists word form learning and retention more than repeatedly retrieving and producing a word with little or no change in context (the 0-0-0 condition). By including two retrieval schedules and documenting differences in learning, we demonstrate that the findings of Leonard et al. (under review) cannot be reduced to a retrieval vs. no-retrieval effect.

Our Experiment 1 findings also provide important information about the role of child productions of target material. We do not claim that production practice is unhelpful. To the contrary, previous studies have documented the importance of repeated productions on learning. For example, Heisler, Goffman, and Younger (2010) found that with repeated production of novel words assigned to novel referents, the motor-articulatory attempts of both typically developing children and children with DLD became more stable. However, our results clearly demonstrate that, if production practice was facilitative, this practice was insufficient to close the gap between the two conditions. This point seems all the more true considering that, during the learning period, productions were actually much more frequent in the 0-0-0 condition than in the 0-2-2 condition.

Furthermore, our Experiment 1 form-referent link recognition findings differ from the Leonard et al. (under review) study. Perhaps due to ceiling effects, Leonard et al. did not observe differences in child form-referent link recognition accuracy according to learning condition, participant group, or time. In the current study, we found a significant difference of group and a significant interaction of group by learning condition. That is, only the DLD group was found to have higher form-referent link recognition accuracy for words that were taught in the RRCR/0-2-2 condition, and group differences favoring the TD children were only observed within the IR/0-0-0 condition. To more fully understand the underlying processes associated with processing recently-taught words that were taught in different learning conditions, we incorporated a multi-level approach by examining neural as well as behavioral data in Experiment 2.

Experiment 2

Experiment 2 aimed to provide a more comprehensive understanding of the effects of retrieval practice on word learning. Therefore, we used event-related brain potentials (ERPs) to compare the neural indices associated with processing words that were taught in the two learning conditions. Our ERP task allows us to examine the real-time neural processes of lexical-semantic processing. Online measures like ERPs often complement and offer new insight into behavioral findings by providing finer-grained information about the information processing that *preceded* the behavioral response.

Event-related brain potentials (ERPs) reflect synchronized neural activity from populations of neurons elicited by a stimulus, such as an auditory tone, or reflect a cognitive process, such as lexical access (Luck, 2014). ERPs have high temporal resolution, which provides valuable information about processing abilities (Luck, 2014). In Experiment 2, we focus on the N400 ERP component (Kutas & Hillyard, 1980). The N400 component has been shown to index lexical-semantic access and the degree of semantic fit of an item within a certain context (Kutas & Federmeier, 2011). As such, the N400 ERP component that is measured in Experiment 2 provides insight into the depth of learning of the newly taught words, relative to the rather superficial measure of learning in the form-referent link recognition task presented in Experiment 1. The magnitude of the N400 elicited from anomalous picture-label pairings indicates how strongly the children learned the association between the newly taught labels and the referent. Thus, differences in the magnitude of the N400 component elicited from words that were taught in the IR and RRCR conditions can provide important information about the depth of learning that results from IR and RRCR learning schedules.

A strength of the N400 component is that it can be elicited in individuals across a wide age range, including young children (e.g., Friedrich & Friederici, 2005; Mills, Coffey-Corina, & Neville, 1993; Silva-Pereyra, Rivera-Gaxiola, & Kuhl, 2005). Following a semantic violation, such as a semantically anomalous word in a sentence (e.g., “Brush your *book*”), individuals typically demonstrate a negative polarity shift that peaks between 200 ms and 600 ms (Kutas &

Federmeier, 2011). In young children, the mean amplitude of the N400 is typically larger and peaks later than the N400 observed in adults (Hahne, Eckstein, & Friederici, 2004; Holcomb, Coffey, & Neville, 1992).

In addition to changes in the N400 that are associated with development, within-individual changes in the N400 can emerge with stimuli repetition. Adult studies have shown that the N400 reduces in amplitude and shortens in duration with repetition of anomalous stimuli (Batterink & Neville, 2011; Besson, Kutas, & Petten, 1992). Nevertheless, the N400 is still elicited in tasks when anomalous stimuli are presented in non-sequential repetitions – for example, presenting a variety of mismatching picture-label pairs instead of sequentially repeating the same picture-label mismatch (Renoult & Debruille, 2009; Renoult, Brodeur, & Debruille, 2010). In fact, our previous work has demonstrated that a robust N400 can be elicited in preschool children with DLD and typically developing preschoolers when semantically anomalous stimuli are repeated throughout a picture-label match-mismatch task (Haebig, Leonard, Usler, Deevy, & Weber, 2018). This detail is important because many word-learning study designs must limit the number of words that are taught to young children to allow for a reasonable degree of successful learning.

Therefore, in Experiment 2, we incorporate ERP data to address our overarching research questions, which are: Does repeated retrieval with contextual reinstatement enhance novel word learning to a greater degree than immediate retrieval practice? Do preschool children with DLD resemble typically developing preschoolers in their learning patterns across the RRCR and IR learning conditions? We hypothesized that data from our match-mismatch ERP task would reveal differences in the underlying processing of recently-taught words according to the learning conditions in which the words were taught. Specifically, when pictures of RRCR items are displayed, we anticipated that a child would likely retrieve the correct label and therefore a matching label would result in a non-detectable N400; however, we expected a mismatching label to elicit a large N400. Because words in the IR condition (that involved no contextual reinstatement) were expected to be learned less well, we expected mismatching labels to elicit a smaller or non-detectable N400. When directly comparing the N400 components between the two learning conditions (using difference waves, explained below), we predicted that the N400 resulting from mismatch trials for words in the RRCR condition would be larger than mismatches for words in the IR condition. In addition, when comparing the match-mismatch behavioral judgments that the children provided during the task, we anticipated higher judgment accuracy for words that were taught in the RRCR condition. Lastly, we hypothesized that children with DLD would have lower accuracy in match-mismatch judgments and a smaller N400 component relative to their TD peers. Our Experiment 2 method allowed us to determine whether potential group differences are limited to the online neural measure.

Methods

Participants. Of the 32 participants from Experiment 1, 27 also participated in Experiment 2 during the one-week test visits (DLD n = 14, TD n = 13). This subset of children was matched on chronological age (M_{DLD} = 58.86 mo, SD_{DLD} = 5.23; M_{TD} = 60.46 mo, SD_{TD} = 4.91; $t(25) = 0.83$, $p = .414$) and gender ($\chi^2 = 0.02$, $p = .88$). This study was approved by the Purdue Institutional Review Board. All participants provided verbal assent and a parent or legal guardian provided informed written consent.

Match-mismatch task to elicit ERPs. During the one-week test visits, the children also completed a match-mismatch task, in which novel word processing was assessed while online electroencephalographic (EEG) data were collected. Given that the children learned two separate sets of words (with six words taught within each set), they also completed two separate match-mismatch tasks; this held the time between teaching and test experiments constant across all words.

The novel word processing tasks followed a match-mismatch paradigm. During match trials, a picture of one of the novel objects was displayed on a screen and an auditory recording of the correct label for the picture was played (e.g., picture: /daɪbo/, label: “/daɪbo/”) via soundfield. In mismatch trials, the label did not match the picture (e.g., picture: /daɪbo/, label: "/nɛp/"). At the end of each trial, children were prompted to judge whether or not the picture and label matched.

Within each match-mismatch task, each of the 6 labels and pictures were presented 20 times, 10 in the match condition and 10 in the mismatch condition. Therefore, there was a total of 120 test trials (30 match IR, 30 mismatch IR, 30 match RRCR, 30 mismatch RRCR). During the mismatch conditions, each label was paired with each of the remaining 5 incorrect pictures twice; therefore, mismatch trials occurred across both the IR condition and the RRCR condition. Match and mismatch trials were pseudorandomized so that labels and pictures repeated no more than twice consecutively, there were no more than three consecutive match or mismatch trials, and there were no more than three consecutive IR or RRCR labels presented.

Visual task stimuli consisted of 2-D pictures that depicted each image used in the word-learning task (McGregor, 2014). The images were approximately 12 cm wide and 9 cm tall and were presented on a 47.5 cm monitor that was 164 cm in front of the child. Auditory stimuli were naturally spoken novel words produced by a female adult with a Midwestern American English dialect. Each word was produced in isolation. Sound stimuli were normalized to have an amplitude of approximately 65 dB using PRAAT software (Boersma & Weenink, 2006). The sound stimuli ranged in duration between 576 ms and 1092 ms.

At the beginning of the match-mismatch task, the examiner explained to the child that he/she would see a picture and hear a name and that the child should tell the examiner if the name matched the picture (i.e., "yes/no"). The children first completed 4 practice trials during which different familiar pictures (e.g., moose, rose) appeared on the screen and a matching or mismatching label was presented via soundfield. The examiner provided feedback. Following the practice trials, the children completed 120 test trials across both learning conditions with no feedback about the child's matching judgments.

A depiction of the match-mismatch task is provided in Figure 7. At the beginning of the trial, one of the six images appeared in the center of the screen at a height visual angle of 3.14º and width visual angle of 4.19º. The picture remained on the screen in silence for 650 ms before the label was presented via a speaker that was mounted above the display screen. Following the completion of the audio file, the picture remained on the screen for an additional 1000 ms (the total time of picture on display ranged from 2226 ms to 2742 ms). Afterwards, a question mark “?” appeared in the center of the screen to prompt the child to judge whether or not the picture and label matched. Once the child made a verbal judgment, the examiner recorded the child’s response by pressing one of two buttons on a response pad. After the child’s response was recorded, the question mark was removed from the screen and a picture of a smiling child appeared in the center of the screen until the examiner advanced the task to the next trial. Fourteen breaks were presented, each after 8 to 12 trials. Breaks alternated between short video clips of nature scenes with music and engaging pictures. During the picture breaks, the child added a sticker to a visual schedule that displayed the child’s progress in the task

Electroencephalographic recordings. Children completed the novel word picture-auditory task while their electroencephalography (EEG) was recorded. We recorded electrical activity at the scalp using a 32 electrode array that was secured in an elastic cap (ActiveTwo head cap, Cortech Solutions). Before the match-mismatch task, children sat and watched a child-friendly movie of choice or played a video game while an examiner measured the child’s head circumference and placed an appropriately-sized elastic electrode cap on the child. A second examiner sat with the child and talked with him/her about the movie or video game while the examiner who was preparing the cap applied gel to each electrode location and subsequently attached the corresponding electrodes to the cap. The electrodes were positioned over homologous hemisphere locations according to the International 10-10 system (Jurcak, Tsuzuki, & Dan, 2007). Locations were: lateral sites F7/F8, FC5/FC6, T7/T8, CP5/CP6, P7/P8; mid-lateral sites FP1/FP2, AF3/AF4, FC1/FC2, F3/F4, CP1/CP2, P3/P4, PO3/PO4, O1/O2; and midline sites FZ, CZ, PZ, OZ. Additional electrodes were placed over the left and right outer canthi for bipolar recordings of horizontal eye movement. Bipolar recordings from electrodes placed over the left inferior and superior orbital ridge (FP1) were used to monitor vertical eye movement. The continuous electroencephalogram data were recorded using the Biosemi ActiveTwo® system with a band-pass filter between .1 and 100 Hz.

ERP measures. The neural data were processed using EEGLAB and ERPLAB (Lopez-Calderon & Luck, 2010), which are Matlab© toolboxes (MathWorks, Natick, MA, USA). During the data processing procedures, the electrical recordings were referenced to the average of the electrodes on the left and right mastoids and the EEG signals were down-sampled at a rate of 256 Hz. Additionally, a band-pass filter from .1 to 30 Hz with a 12-dB roll-off was applied to remove high frequency noise and to minimize offsets and drift. An independent component analysis (ICA; EEGLAB) was completed to identify and remove eye artifact. Briefly, ICA identifies independent sources of EEG signals and yields components that represent patterns from the EEG signal. Two independent trained research assistants identified ICA components that reflected artifact, for example blinks and horizontal eye movements. When there were discrepancies, the coders agreed upon a consensus list of artifact components that were then extracted from the EEG data. Afterwards, the data were epoched from 200 ms prior to the onset of the label to 2000 ms post-stimulus in order to average ERP component measures within each condition for each child's waveforms. During this procedure, epochs were baseline-corrected from -200 ms to the onset of the auditory label (0 ms). The EEG channels underwent automatic voltage-dependent thresholds to remove any trials that still contained artifact.

Within each set, a minimum of 15 artifact-free trials within each condition were required for a child's ERP data to be included in the analyses. We included trials in which children provided accurate or inaccurate judgments. One child with DLD did not complete the set 1 match-mismatch task and the data of another child with DLD did not contain enough usable trials for the set 1 match-mismatch task. Additionally, data from two children were removed from the set 2 ERP dataset, including data from one child with DLD and one typically developing child, because there were not enough usable trials within each condition. Within the IR condition, the average number of artifact-free trials within the match condition was 26.04 for the TD group and 23.08 for the DLD group and the average number of artifact-free trials for the mismatch condition for the TD group was 25.08 and 23.80 for the DLD group. Within the RRCR condition, the average number of artifact-free trials within the match condition was 25.24 for the TD group and 24.00 for the DLD group and the average number of artifact-free trials for the mismatch condition for the TD group was 26.24 and 23.64 for the DLD group. Finally, the artifact-free EEG epochs were averaged within task conditions for each individual.

To capture the temporal aspects of the N400, we selected an early and late N400 analysis window (respectively, 300 to 500 ms and 500 to 700 ms post-onset of the novel label). After examination of the grand averaged waveforms, we chose to only examine the 500 to 700 ms window when analyzing the difference waves, as they captured the greatest differences between the match and mismatch trials. The ERP difference waves were formed by subtracting the match from the mismatch ERPs to isolate the N400 component while removing other trial characteristics that are represented in the waveforms. The selected time windows are centered around the regions of maximal activity, which aligns with windows that have been used in previous studies examining language processing in children (e.g., Neville, Coffey, Holcomb, & Tallal, 1993; Sabisch, Hahne, Glass, von Suchodoletz, & Friederici, 2006; Usler & Weber-Fox, 2015). Notably, our analysis windows align with the two previous studies that have examined the N400 component in typically developing preschoolers and preschool children with DLD (Haebig et al., 2018; Pijnacker et al., 2017). As a second step in the window selecting procedures, we examined each individual's waveforms to ensure the windows captured the N400 for each child. We measured the N400 from a specified region of interest (P3, Pz, P4, PO3, PO4, O1, Oz, O2), which aligns with regions of interest used in previous studies examining word processing in preschool children with DLD (Haebig et al., 2018).

Analysis procedures. Behavioral judgments were converted into A' scores to control for response bias (Grier, 1971; Rice et al., 1999). A' scores are derived from the proportion of correct responses in a two-alternative forced-choice task. Therefore, the A' value consists of scores from a control condition and an experimental condition (e.g., match trials and mismatch trials). To calculate A' scores, we used the following formula: A' = 0.5 + (y – x) (1 + y – x) / 4y (1 – x), where y represents correct identifications (hits) and x represents incorrect identifications (false alarms; Linebarger, Schwartz, & Saffran, 1983). An A' value of 1.00 represents perfect discrimination of correct and incorrect picture-label pairings. An A' value of .50 indicates chance performance, such as a “yes” response to 50% of the match trials and to 50% of the mismatch trials. A mixed-effect random intercept model was used to test for differences in A' scores according to group, learning condition, and set (using lme4; Bates, Maechler, Bolker, & Walker, 2015). Given that not all of the children completed both sets of the match-mismatch task, we chose to use mixed-effect models because they tolerate missing data. This approach allowed us to retain all participants and to minimize the analyses that were run (i.e., one mixed-effect model versus an ANOVA for set 1 data and an ANOVA for set 2 data).

The ERP data were analyzed using a series of mixed-effect models estimated with random intercepts at the child level and repeated measures nested within child. First, we analyzed ERP data within each learning condition with the following fixed effects: set (set 1 vs. set 2), trial type (match vs. mismatch), group (TD vs. DLD), electrode site (electrodes within the ROI), and an interaction between trial type and group. In these within-condition analyses, our primary variables of interest were effects of trial type and group. To directly compare ERP differences between the IR and RRCR learning conditions, we also compared across the learning conditions by analyzing the mean amplitude of the difference waves (dependent variable) in the 500 to 700 ms analysis window. In this model, the fixed effect independent variables included: set (set 1 vs. set 2), learning condition (IR vs. RRCR), group (TD vs. DLD), electrode site (electrodes within the ROI), and an interaction between learning condition and group. In our difference wave analysis, our primary variables of interest were effects involving learning condition and group. We do not report significant differences between electrodes because we were interested in the entire ROI; however, the full model output that includes each electrode within our ROI can be found in the Supplemental Materials. Effect sizes are reported using partially standardized beta coefficients (b_{std}). Lastly, we directly examined the Experiment 2 behavioral and ERP data by conducting a bivariate correlation with the A' scores and the mean amplitude of the difference waves within our ROI.

Results

Behavioral performance. First, we examined the children's behavioral judgments. Descriptively, the TD group had a mean A' score of 0.67 (*SD* = 0.13) for the IR condition and 0.69 (*SD* = 0.13) for the RRCR condition. The DLD group mean A' score for the IR condition was 0.59 (*SD* = 0.13) and 0.63 (*SD* = 0.15) for the RRCR condition. Our mixed-effect model revealed a significant effect of learning condition, with higher accuracy for the RRCR condition (b = 0.03; 95% CI: 0.01 0.05). A' scores increased .205 standard deviations when children judged picture-word pairs for words that were taught in the RRCR learning condition relative to the IR condition (b_{std} = .20, indicating a small effect). There was no significant difference between the TD and DLD groups and scores did not differ according to set. Also, there were no significant interactions across group, learning condition, and set (see Table 8). We also analyzed child accuracy by using the percentage of correct judgments as the dependent variable; this analysis resulted in the same pattern of findings.

N400 mean amplitude. Second, we examined the ERPs to better understand the neural indices associated with processing newly taught words. Figure 8 depicts the waveforms for IR and RRCR learning conditions for the children with typical language development and the children with DLD. Table 9 provides the model output for each analysis window and learning condition. The first set of mixed-effect models examined the words taught in the IR learning condition. We were primarily interested in whether there was a significant difference in mean amplitude (N400 mean amplitude) between match and mismatch trials and whether there was an interaction between group and trial type (match vs. mismatch). As seen in Figure 8, the waveforms between match and mismatch trials overlap indicating that, when testing words that were taught in the IR condition, mismatch trials did not elicit an increased N400 amplitude relative to match trials. Our analyses for the 300 to 500 ms window of interest and the 500 to 700 ms window of interest confirmed this observation. In addition, there were no significant differences between the TD children and the children with DLD, nor was there an interaction between trial type and group.

Our next analyses examined the ERPs for words that were taught in the RRCR learning condition. As in the previous analyses, we separately examined an early and late N400 window (300 - 500 ms, 500 - 700 ms). There was a significant effect of trial type (Table 9). As can be seen in Figure 8, mismatch trials elicited a larger N400 mean amplitude compared to the match trials. The significant difference between match and mismatch trials was apparent in both the early and late analysis windows. Within the 300 – 500 ms window, mismatch trials were .20 standard deviation units more negative than match trials ($b_{std} = .20$). Within the 500-700 ms window, mismatch trials were .44 standard deviation units more negative than match trials ($b_{std} = .44$). There was no significant difference between groups and there was no interaction between group and trial type.

In addition, we directly compared the IR and RRCR learning conditions by analyzing the difference waves within a temporal window of 500 and 700 ms post stimulus onset, where the maximal differences between learning conditions were seen. Our mixed-effect model (Table 10) revealed a significant effect of learning condition, indicating the difference between match and mismatch trials was .53 standard deviation units greater for words that were taught in the RRCR learning condition relative to the IR learning condition ($b_{std} = .53$; see Figure 9). There was no significant effect of group nor was there an interaction between group and learning condition.

Lastly, our bivariate correlation analyses that examined whether there is an association between the behavioral judgments (A' scores) and the ERP data (mean amplitude of the difference waves for the IR and RRCR conditions) yielded a nonsignificant correlation (*p*s > .25).

Discussion

Experiment 2 used a receptive match-mismatch task to examine how recently-taught words that were taught using two different repeated retrieval schedules are processed. We found that at both the behavioral and neural levels, word processing differed according to the learning condition in which the novel words were taught. Preschool children with typical language development and DLD were more accurate in judging matching and mismatching label-referent pairings when the stimuli had been taught in the RRCR learning condition. Moreover, Experiment 2 demonstrated that the underlying neural patterns differed when children were presented with label-referent pairings from the RRCR condition. Presentation of a label-referent mismatch elicited a larger N400 component for words that were taught in the RRCR condition; unlike the Experiment 1 form-referent link recognition task, a learning condition effect was present for both groups in our Experiment 2 match-mismatch task. The ERP data also revealed that typically developing preschool children and preschool children with DLD processed the recently-taught words similarly.

It is important to note that, although both the behavioral and ERP findings identified a difference between the IR and the RRCR learning condition, the difference in behavioral judgments between RRCR and IR trials was small. This was rather unsurprising given the near-ceiling performance in our TD group from Experiment 1 and the lack of a learning condition effect in the Leonard et al. (under review) findings on the form-referent link recognition task. Although our match-mismatch task differed from the form-referent link recognition task, it was still less demanding than our word form recall task. In contrast, the effect sizes were larger in our ERP findings. This is particularly interesting given that our ERP measurements were derived from all of the trials, regardless of the trial-level accuracy of the child's behavioral judgment. These findings suggest that, relative to the behavioral data, the ERP data were more sensitive to learning differences between words that were taught in the RRCR and IR learning conditions.

In addition, the robustness of the N400 and other ERP components can be influenced by multiple factors. Most relevant to this study, stimuli repetition can dampen the amplitude of the N400 and shorten the duration (Batterink & Neville, 2011; Besson et al., 1992). Importantly, studies have confirmed that the N400 is still elicited when anomalous stimuli repetitions are presented in non-sequential order (Renoult & Debruille, 2009; Renoult, Brodeur, & Debruille, 2010). Given this, the Haebig et al. (2018) familiar word ERP study served as a precedent for the current experiment because it demonstrated that despite stimuli repetition, picture-label mismatch trials elicited a robust N400 in preschool children with typical language development and developmental language disorder. In Experiment 2, we found a very clear N400 for words in the RRCR condition, even though a limited number of words were used and considerable repetition occurred. In contrast to words in the RRCR condition, words in the IR condition showed no indication of an N400, suggesting that, for these words, children developed more superficial word label representations that were not strongly primed by the picture.

Furthermore, as with the Haebig et al. (2018) study, the N400 that was elicited by mismatching picture-label pairs did not differ between preschool children with DLD and those with typical development. It is important to note that some previous studies have identified N400 differences in children with DLD relative to their peers (Kornilov, Magnuson, Rakhlin, Landi, & Grigorenko, 2015; Pijnacker et al., 2017). These studies sometimes used sentence-level stimuli or examined words across different classes (adjectives, verbs, nouns). In contrast, Haebig and colleagues only examined word processing using early-acquired nouns that were presented at the single-word level, which did not appear to tax processing abilities in the preschool children with DLD. Similarly, there were no significant group differences in the amplitude of the N400 that was elicited by words that were taught in the RRCR learning condition in Experiment 2. These findings suggest that neural processing for lexical retrieval and integration for words taught using RRCR is similar for children with typical language development and children with DLD. This provides additional support to the Experiment 1 behavioral findings, that RRCR enhances longer-term retention of recently-taught words and the learning benefit associated with RRCR is similar for typically developing children and children with DLD.

General Discussion

The current study directly compared the effectiveness of two repeated retrieval schedules on word learning using a multilevel approach. This work served as the next logical extension of Leonard et al. (under review), who demonstrated that repeated retrieval practice that engages contextual reinstatement greatly enhances word learning and retention relative to a more common repeated study learning protocol. In order to more comprehensively understand how retrieval practice benefits learning, it was necessary to determine that retrieval practice alone was not the sole source of learning enhancement. In addition to serving as the first study to directly compare repeated retrieval schedules in preschool children, to our knowledge, this study is the first to examine the neural correlates associated with processing recently-taught words in children with DLD. In both the behavioral and ERP data, we observed that preschool children experienced enhanced learning when novel words were taught in the RRCR learning condition, relative to the IR condition. These findings underscore the importance of retrieval practice that requires contextual reinstatement during the learning process. Importantly, our findings reveal that RRCR enhances learning in children with typical language development as well as children with DLD.

Word Form

In Experiment 1, we found that on average, children accurately recalled 2.5 more novel labels for words that were taught in the RRCR/0-2-2 learning condition relative to the IR/0-0-0 learning condition. The strong effect of the RRCR learning condition is thought to come from the enhanced representation that children create when retrieving the novel words in slightly changing contexts – which in the current study was created by inserting different novel labels between the encoding (study) context and retrieval opportunities.

Although the RRCR learning condition enhanced learning of word form for both groups, word learning overall was not equivalent. An examination of the early learning stages of the IR/0-0-0 retrieval trials on the first day, revealed that the children with DLD were less accurate in recalling the words even though the recall trials occurred immediately after a study trial. Despite this early reduction in performance, the TD children and children with DLD followed a similar trajectory of learning throughout the IR/0-0-0 trials. This aligns with other work that suggests that children with DLD have encoding deficits that significantly impact the early stages of word learning (Alt & Plante, 2006; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017).

In addition to facilitating word form recall, the RRCR condition also appeared to facilitate form-referent link recognition. Scores were much higher for recognition than for recall, but the condition effects were nevertheless quite clear for the children with DLD. Given the ceiling effects apparent in the TD group for the form-referent link recognition task, the group by learning condition interaction was expected. The recognition task served as a rather superficial assessment that only required children to be able to hold a shallow representation of each novel word form. Therefore, the Experiment 2 ERP data provided more sensitive information and did indeed reveal a strong effect for learning condition, across both groups.

ERP findings not only complement our behavioral data but inform our understanding of online processing in preschool children. The literature reflects many examples of the valuable information that measures of online processing offer to our understanding of word processing (Ellis, Borovsky, Elman, & Evans, 2015; Haebig et al., 2018; McMurray, Horst, & Samuelson, 2012). In the current study, we found that a robust N400 was elicited during mismatch trials when the words being tested were taught using repeated retrieval with contextual reinstatement. This, along with the current match-mismatch judgments and the behavioral findings from Experiment 1, indicates that the RRCR condition enabled children to reinstate the contextual representation of words and update them with new features, thereby increasing their effectiveness during retrieval. As a result, the labels of words that were learned in the RRCR condition were likely retrieved more automatically during the Experiment 2 match-mismatch task and were more effectively primed by the picture. The automaticity of picture-label (or referent-label) pairings influences the N400 component (Juottonen, Revonsuo, & Lang, 1996; Kutas & Federmeier, 2011). Therefore, match trials resulted in a very small N400 or no N400. Furthermore, because pictures of words learned in the RRCR condition would more effectively prime the label, a stronger semantic anomaly effect occurred during trials with mismatching picture-label pairs, which resulted in a robust N400.

In contrast, words learned in the IR condition may have been retrieved less successfully because, during retrieval practice, the children were less likely to reinstate and update the stored contextual representation. Therefore, its value as a retrieval cue was not increased as in the RCCR condition. Consequently, the priming between picture and label may have been weaker, and any mismatch between the two less impactful. The ERP literature has found that the strength of an association between a prime and target stimulus, and the frankness of a violation, impacts the amplitude of the N400 (Federmeier & Kutas, 2001; Kutas & Federmeier, 2011). In Experiment 2, there were no differences between the N400 amplitude that was elicited by match and mismatch trials for words learned in the IR condition.

Another notable finding in Experiment 2 was that the mismatch trials elicited a significant N400 effect in the early N400 window (300 – 500 ms), which is most often associated with the N400, despite the young age of our participants (Kutas & Federmeier, 2011). Examining both an ‘early’ (300-500 ms) and ‘later’ (500-700 ms) window allowed us to investigate possible timing differences or relative delays of the N400 between the groups. Our results indicate that the timing of processing for children with DLD when distinguishing matches or mismatches of novel image-label pairings is comparable to their typically developing peers when provided with the opportunity for repeated retrieval with contextual reinstatement.

Meaning

Although we anticipated meaning recall to exceed word form recall, we predicted that TD children and children with DLD would demonstrate a pattern similar to the pattern found for the word form data. Interestingly, in Experiment 1, we found an effect for learning condition only when applying a model without the condition random slope. Further, unlike the Leonard et al. (under review) meaning recall findings, TD children recalled more word meanings than the children with DLD.

When examining the early learning stages for meaning in the IR/0-0-0 retrieval trials on the first day, the children with DLD were less accurate. The importance of this finding lies in the fact that, whereas word form encoding is assumed to engage the procedural system, our learning task for meaning seems to involve primarily the declarative system. That is, when learning new word forms, the children had to encode sequences of consonants and vowels. However, when learning the meanings, the words constituting the “definitions” (e.g., “clouds,” “rain”) were already known to the children; their task was to associate each definition with the appropriate picture. According to Ullman and Pierpont (2005), even in the latter case, the act of retrieving the definition might involve the procedural system. However, we kept retrieval demands to a minimum in the 0-0-0 condition, yet group differences were seen from the outset of learning. This finding suggests that encoding weaknesses in children with DLD can extend to cases in which sequence learning is not involved.

Clinical Implications and Future Studies

The current findings lend support to the belief that retrieval practice influences the nature of learning (Karpicke & Roediger, 2008) and “the nature of storage and retrieval from human memory” (Bjork, 1988, p. 396), with benefits extending to young children with DLD. This is particularly relevant given the storage-elaboration hypothesis, which proposes that children with DLD have deficits in encoding, leading to insufficient storage of the details of words (Kail &

Leonard, 1986; Kail, Hale, Leonard, & Nippold, 1984; Leonard, Nippold, Kail, & Hale, 1983). As such, the storage-elaboration hypothesis predicts that children with DLD will have superficial lexical-semantic representations, impaired word processing, and reduced word recall abilities (Leonard, 2014). Given these deficits, it is necessary to develop evidence-based techniques that will support word learning. The current study contributes such evidence; despite encoding deficits, repeated retrieval with contextual reinstatement can nonetheless improve the longer-term retention of newly taught words in children with DLD and does so as effectively as in typically developing children.

Although the current study provides an important first step in understanding the importance of RRCR, there still is much to be learned. For example, given that learning characteristics may change with the course of development, it will be important for future studies to explore the role of RRCR on learning in individuals with DLD at different points in development. Our current findings align with recent work that examined the role of retrieval practice in word learning in young adults with DLD (McGregor et al., 2017); despite this, additional studies are needed to determine whether there are nuanced effects of RRCR on learning at different developmental stages. This is especially important given that the gap in word knowledge in preschool children with DLD only widens as they age into adulthood (Rice & Hoffman, 2015). Future work also should explore whether similar learning methods enhance learning of different types of words (e.g., verbs, adjectives). A meta-analysis of word learning in children with DLD indicates that some word classes, such as verbs, are more difficult to learn (Kan & Windsor, 2010). Additional studies examining both the behavioral and neural correlates associated with processing recently-taught verbs or adjectives that have been taught using RRCR would be valuable.

In conclusion, we observed, at both the behavioral and neural level, enhanced word learning for words that were taught in the RRCR learning condition. Our findings promote not only the importance of word exposure, but the crucial role of retrieval practice that engages contextual reinstatement. This work serves as an important first step in laying the groundwork of evidence for retrieval-practice in addressing weaknesses in word knowledge in young children with DLD. Given the importance of word knowledge on child academic and social outcomes, it will be important to conduct further research in this promising area.

Acknowledgments

We thank the families who participated in this study. Also, we thank Megan Miller, Janell Blunt, Sarah Barnes, Anna Redmaster, Julia Bergman, Kelsey Delacroix, Rachel Willing, Joseph Gardner, Taylor Jagiella, and Kaitlyn Brickey for their help with data collection for Experiment 1. Additionally, we thank Katie Gerwin, Jen Schumaker, and Gina Catania for their help with data collection and processing for the Experiment 2 data.

References

Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders*, *52*, 207–220.

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping language impairment. *Journal of Speech Language and Hearing Research*, *49*, 941–955.

American Speech-Language-Hearing Association. (1997). Guidelines for audiologic screening. Retrieved from Available from www.asha.org/policy.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience*, *23*, 3181–3196.

Besson, M., Kutas, M., & Petten, C. Van. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience*, *4*, 132–149.

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. *Practical Aspects of Memory: Current Research and Issues, Vol. 1: Memory in Everyday Life*, *1*, 396–401.

Boersma, M., & Weenink, D. (2006). PRAAT: Doing phonetics by computer. Amsterdam, the Netherlands: University of Amsterdam. Retrieved from http://www.praat.org

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language,* *and Hearing Research*, *45*, 1142–1157.

Cepeda, N. J., Pashler, H., Vul, E., Rohrer, D., & Wixted, J. (2006). Distributed practice in verbal recall tasks: A review and quantitative syntesis. *Psychological Bulletin*, *132*, 354–380.

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

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Bloomington, MN: NCS Pearson.

Edwards, J., Beckman, M. E., & Munson, B. (2004). Vocabulary Size and Phonotactic Production Accuracy and Fluency in Nonword Repetition. *Journal of Speech, Language, and Hearing Research*, *47*, 421–436.

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders*, *58*, 143–157.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 202–224.

Friedrich, M., & Friederici, A. D. (2005). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *Neuroreport*, *16*, 653–656.

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

Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research*, *37*, 913–923.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R. A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory and Cognition*, *3*, 177–182.

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

Greenslade, K. J., 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.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*, 424–429.

Haebig, E., Leonard, L. B., Usler, E., Deevy, P., & Weber, C. (2018). An initial investigation of the neural correlates of word processing in preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *61*, 729–739.

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, *58*, 1251–1263.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, *16*, 1302–1318.

Hall, W. S., Nagy, W. E., & Linn, R. L. (1984). *Spoken words, effects of situation and social* *group on oral word usage and frequency*. Hillsdale, NJ: Lawrence Erlbaum.

Heisler, L., Goffman, L., & Younger, B. (2010). Lexical and articulatory interactions in children's language production. *Developmental Science*, *13*, 722–730.

Holcomb, P. J., Coffey, S. a., & Neville, H. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*, 203–241.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*, 99–107.

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, *34*, 1600–1611.

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

Kail, R., & Leonard, L. B. (1986). Word-finding abilities in language-impaired children. *ASHA Monographs*, *25*, 1–39.

Kail, R. V, Hale, C. A., Leonard, L. B., & Nippold, M. A. (1984). Lexical storage and retrieval in language-impaired children. *Applied Psycholinguistics*, *5*, 37–49.

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*, 739–757.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *37*, 1250–1257.

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

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, *7*, 1–9.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). *Retrieval-based learning: An episodic context account*. (B. Ross, Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 60). San Diego, CA: Elsevier Academic Press. doi:10.1016/B978-0-12-800283-4.00007-1

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

Kaufmann, A., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children*. Circle Pines, MN: AGS.

Kornilov, S. A., Magnuson, J. S., Rakhlin, N., Landi, N., & Grigorenko, E. L. (2015). Lexical processing deficits in children with developmental language disorder: An event-related potentials study. *Development and Psychopathology*, *27*, 459–476.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading sensless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In & R.

N. S. M. Gruneberg, P. E. Morris (Ed.), *Practical Aspects of Memory* (pp. 625–632). London, England: Academic Press.

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

Lehman, M., Smith, M. A., & Karpicke, J. D. (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. B. (2014). *Children with Specific Language Impairment* (2nd ed.). Cambridge, MA: MIT Press.

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

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361–392.

Lopez-Calderon, J., & Luck, S. J. (2010). ERPLAB Toolbox (1.1.0). Retrieved from http://erpinfo.org/erplab

Lucas, R., & Norbury, C. F. (2015). Making inferences from text: It's vocabulary that matters. *Journal of Speech, Language, and Hearing Research*, *58*, 1224–1232.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second). Boston, MA: MIT Press.

McGregor, K. K. (2014). What a difference a day makes: Change in memory for newly learned word forms over twenty-four hours. *Journal of Speech, Language, and Hearing Research*, *57*, 1842–1850.

McGregor, K. 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. 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.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*, 831–877.

Mills, D. L., Coffey-Corina, S. a., & Neville, H. J. (1993). Language Acquisition and Cerebral Specialization in 20-Month-Old Infants. *Journal of Cognitive Neuroscience*, *5*, 317–334.

Neville, H., Coffey, S. A., Holcomb, P. J., & Tallal, P. (1993). The neurobiology of sensory and language processing in language-impaired children. *Journal of Cognitive Neuroscience*, *5*, 235–253.

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

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*, 554–566.

Pijnacker, J., Davids, N., Van Weerdenburg, M., Verhoeven, L., Knoors, H., & Van Alphen, P. (2017). Semantic processing of sentences in preschoolers with specific language impairment: Evidence from the N400 effect. *Journal of Speech, Language, and Hearing Research*, *60*, 627–639.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental Relations Between Vocabulary Knowledge and Reading Comprehension: A Latent Change Score Modeling Study. *Child Development*, *86*, 159–175.

Renoult, L., Brodeur, M. B., & Debruille, J. B. (2010). Semantic processing of highly repeated concepts presented in single-word trials: Electrophysiological and behavioral correlates. *Biological Psychology*, *84*, 206–220.

Renoult, L., & Debruille, J. B. (2009). N400-like potentials and reaction times index semantic relations between highly repeated individual words. *Cognitive Neuroscience, Journal Of*, *23*, 905–922.

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

Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality judments of extended optional infinitive grammar: Evidence from english-speaking children with specific langauge impairment. *Journal of Speech, Language, and Hearing Research*, *42*, 943–961.

Roark, B., & Demuth, K. (2000). Prosodic constraints and the learner’s environment: A corpus study. *Proceedings of the 24th Annual Boston University Conference on Language Development*, 597–608.

Sabisch, B., Hahne, A., Glass, E., von Suchodoletz, W., & Friederici, A. D. (2006). Lexical-semantic processes in children with specific language impairment. *NeuroReport*, *17*, 1511–1514.

Schopler, E., Van Bourgondien, M., Wellman, G., & Love, S. (2010). *Childhood Autism Rating Scale-Second Edition* (2nd ed.). Los Angeles: CA: Western Psychological Services.

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Cognitive Brain Research*, *23*, 247–258.

Storkel, H. L., & Hoover, J. R. (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. L., Komesidou, R., Fleming, K., & Swinburne 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. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K., & Swinburne 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.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*, 1245–1260.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*, 399–433.

Usler, E., & Weber-Fox, C. (2015). Neurodevelopment for syntactic processing distinguishes childhood stuttering recovery versus persistence. *Journal of Neurodevelopmental*

Disorders, *7*, 1–21.

Supplemental Materials

Supplemental Table 1 presents additional model output results for the word form analyses. These models extend the models presented in Table 1 in the main text by including two- and three-way interactions of group, time, and learning condition. Because these interactions across group, time, and learning condition are not statistically significant in the additional models, the group effect found Table 1, Model A in the main paper does not differ across time or learning condition and the learning condition effect does not differ across time or group.

Supplemental Table 2 provides the additional model output results for the meaning outcome analyses. These models contain additional fixed effects relative to the models presented in Table 4; these fixed effects include two- and three-way interactions of group, time, and learning condition. As noted in the main text, Model B does not contain a random slope for learning condition; this model yields a significant effect of learning condition that does not differ across time or group, given that the interactions learning condition, group, and time are not statistically significant in Models C and D.

Supplemental Table 3 provides the additional model results for the meaning outcome analyses. These models extend the results presented in Table 4 in the main text and include additional interactions of group, time, and learning condition. The more parsimonious models are presented in Table 4 in the main text.

Supplemental Table 4 provides the full within-condition ERP results from the mixed-effect models that correspond to Table 9. This supplemental table incorporates each electrode from the region of interest (ROI). Clear trial type effects are observed for the RRCR/022 condition.

Supplemental Table 5 provides the full between-condition ERP output that includes each electrode in the ROI. This table corresponds to the in-text Table 10. The model output documents the larger N400 component for the words that were learned in the RRCR/022 condition.

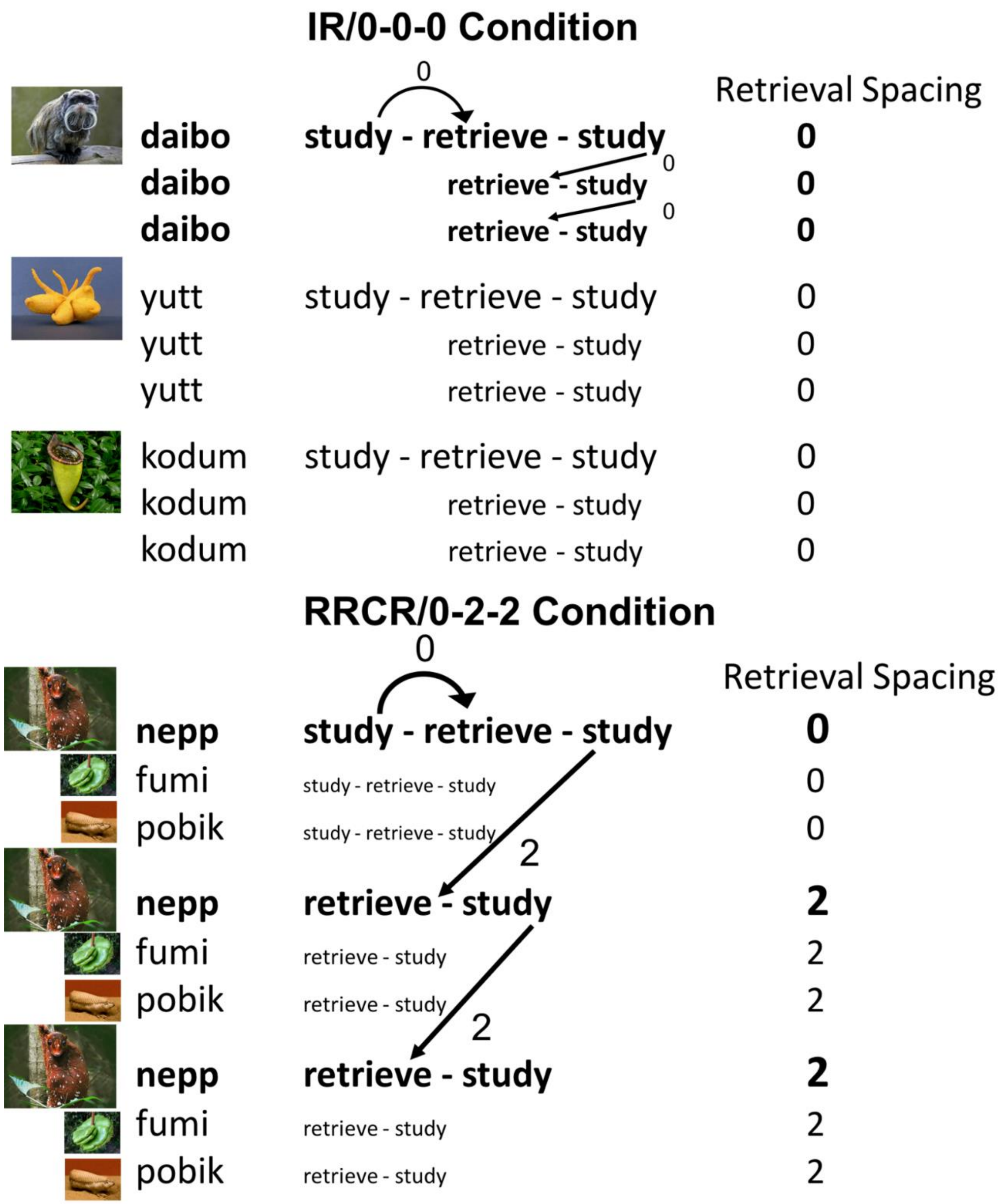

Figure 1. Word learning task design for the immediate retrieval (IR/0-0-0) learning condition and the repeated retrieval with contextual reinstatement (RRCR/0-2-2) learning condition.

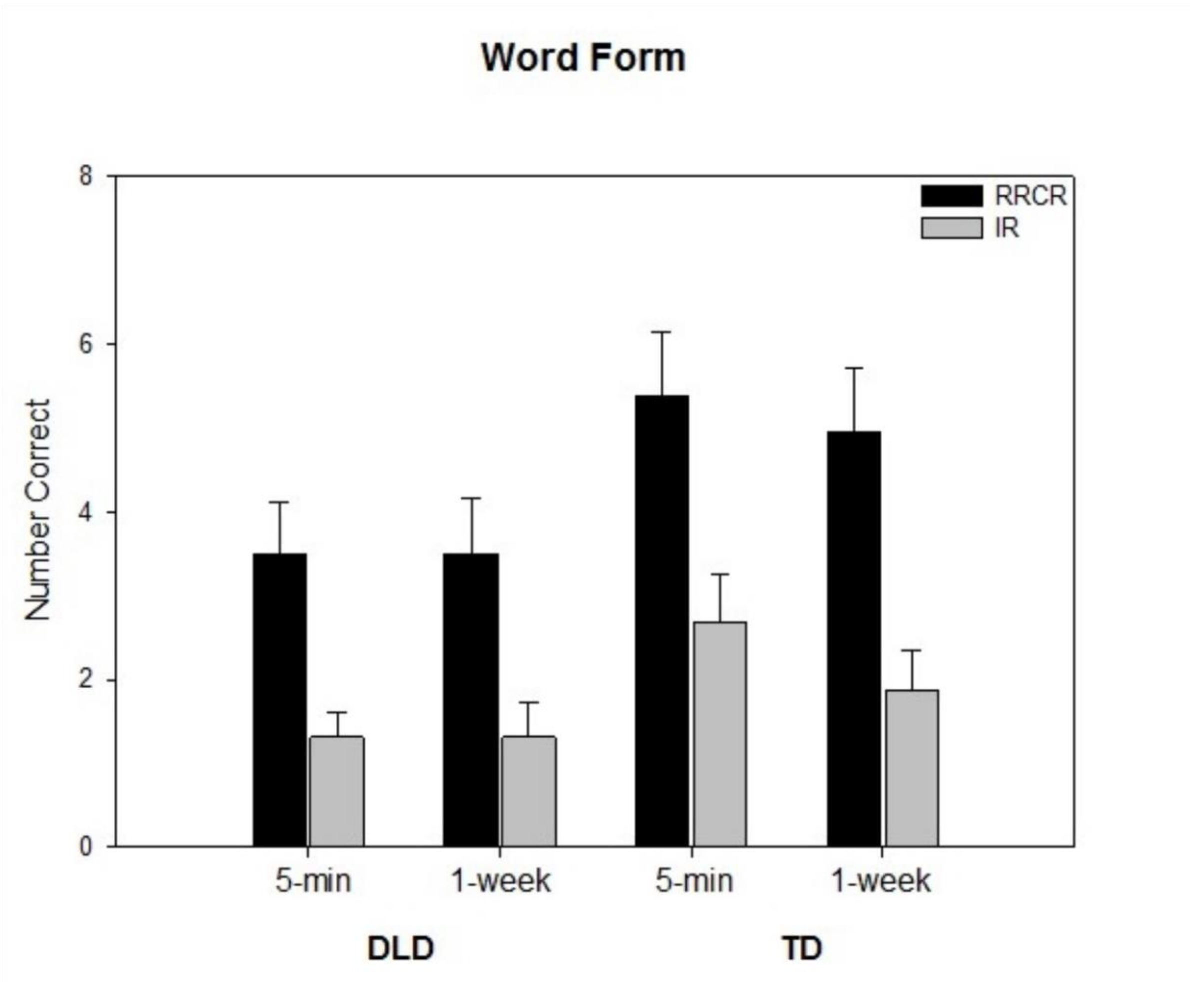

Figure 2. The mean number of items correct on the recall test of Experiment 1 at five minutes and one week for novel words in the repeated retrieval with contextual reinstatement (RRCR) condition and the immediate retrieval (IR) condition by the children with developmental language disorder (DLD) and the children with typical language development (TD). Error bars mark standard errors.

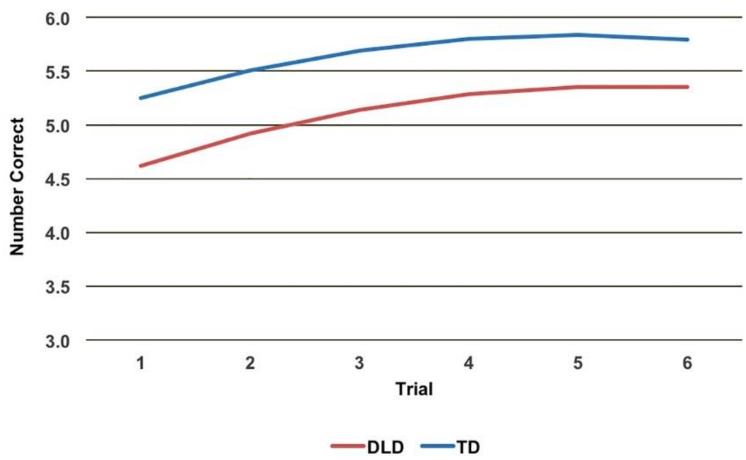

Figure 3. Average trajectories for novel word form retrieval across the learning period for novel words in the immediate retrieval (IR) condition for the children with developmental language disorder (DLD) and children with typical language development (TD).

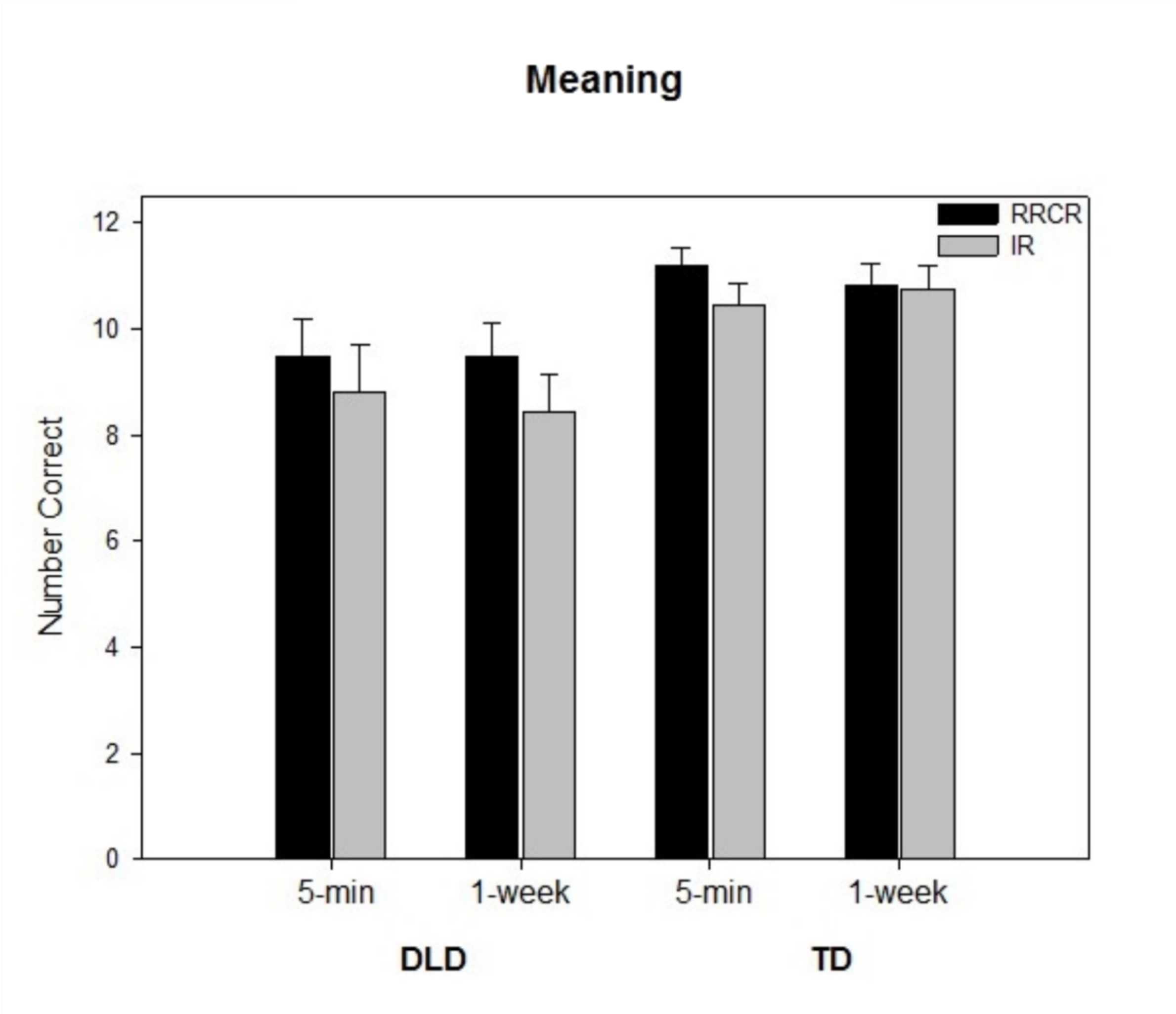

Figure 4. The mean number of items correct on the meaning recall test of Experiment 1 at five minutes and one week for novel words in the repeated retrieval with contextual reinstatement (RRCR) condition and the immediate retrieval (IR) condition by the children with developmental language disorder (DLD) and the children with typical language development (TD). Error bars mark standard errors.

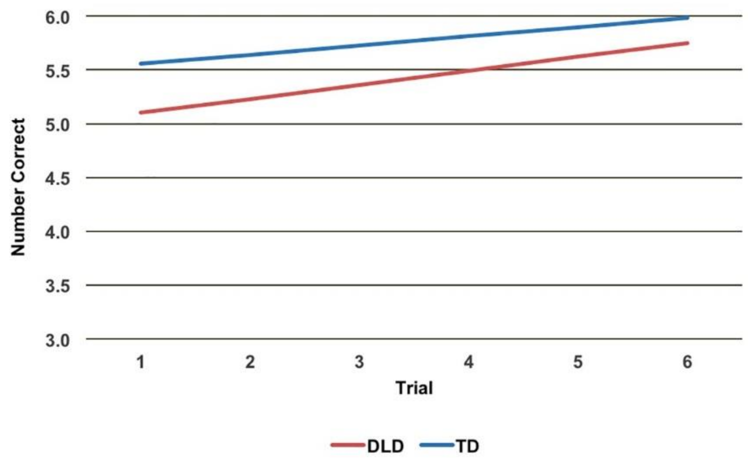

Figure 5. Average trajectories for the retrieval of meaning across the learning period for novel words in the immediate retrieval (IR) condition for the children with developmental language disorder (DLD) and children with typical language development (TD).

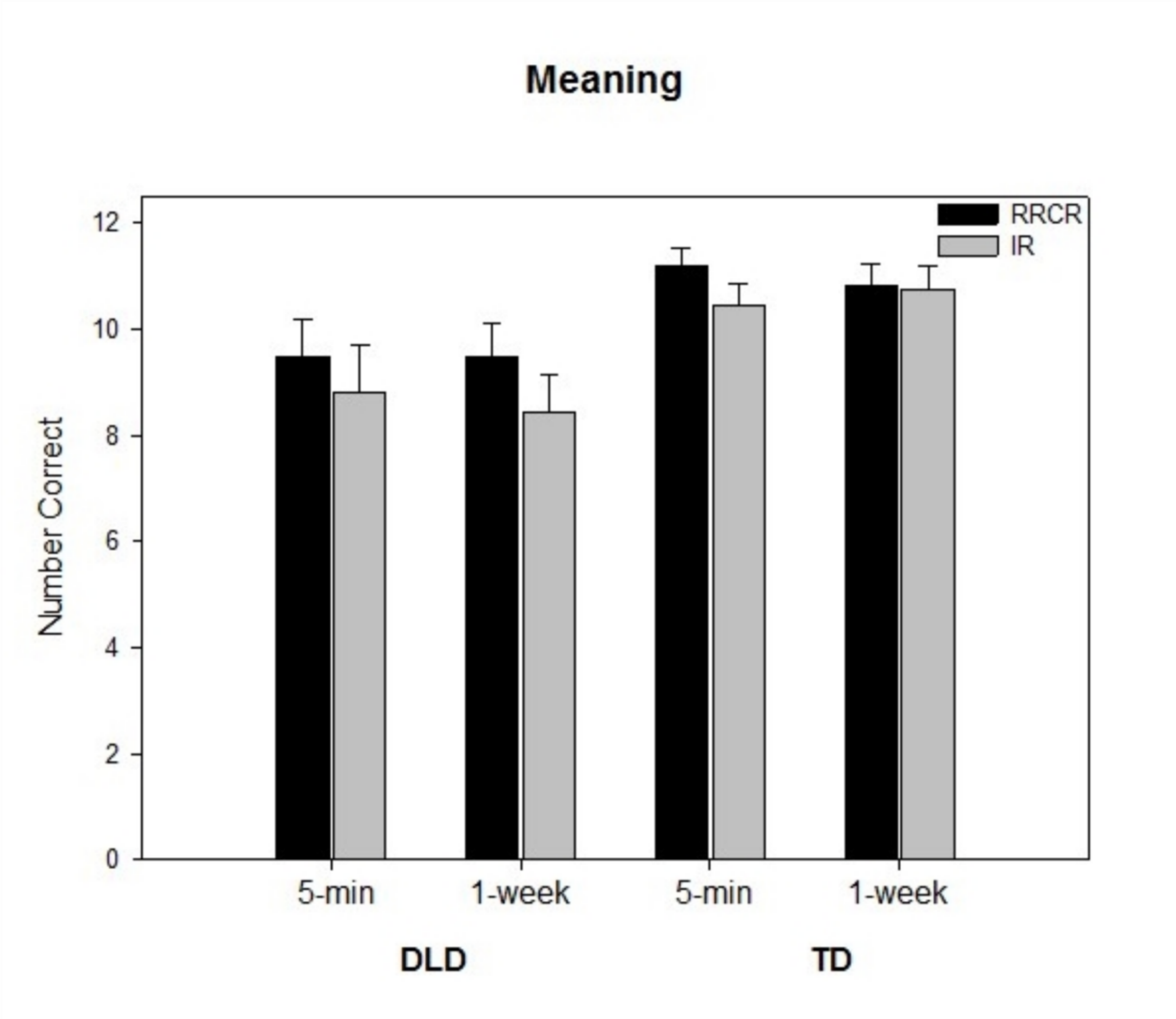

Figure 6. The mean number of items correctly identified on the form-referent link recognition test of Experiment 1 at the one-week test for novel words taught in the repeated retrieval with contextual reinstatement (RRCR) condition and the immediate retrieval (IR) condition by the children with developmental language disorder (DLD) and the children with typical language development (TD). Error bars mark standard errors.

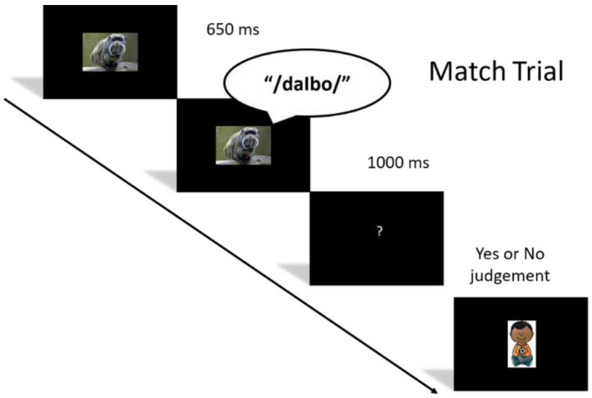

Figure 7. Match-Mismatch Task Procedure

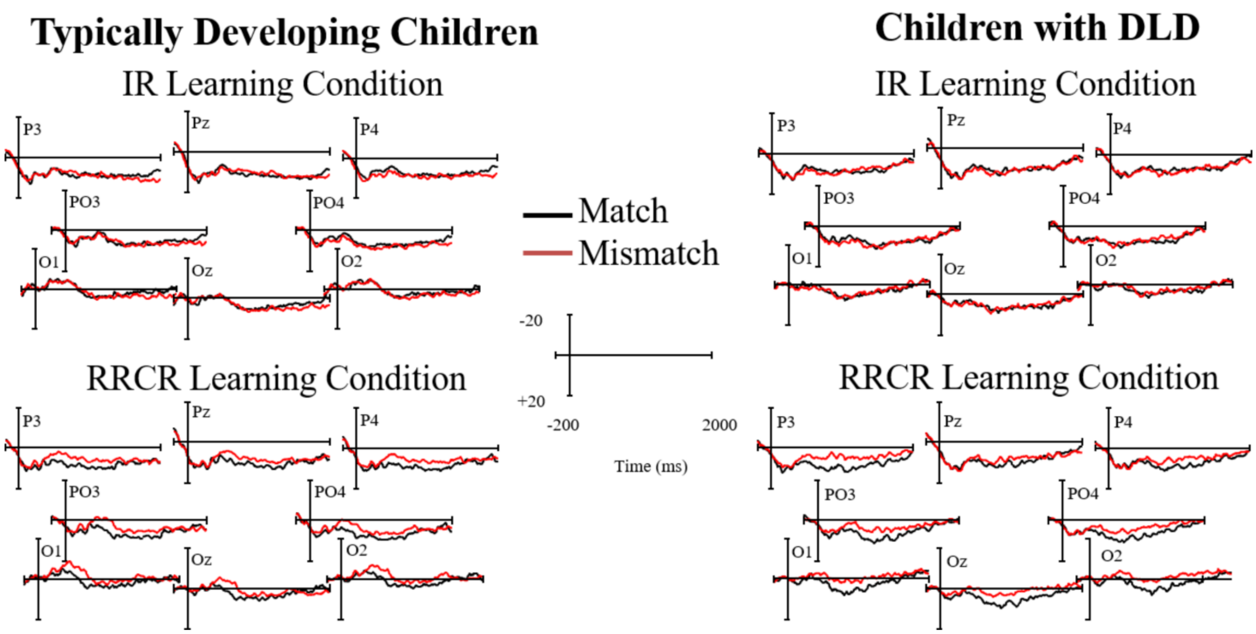

Figure 8. Match-Mismatch task waveform averages for the typically developing children and the children with developmental language disorder (DLD) for the immediate retrieval learning condition and the repeated retrieval with contextual reinstatement (RRCR) learning condition.

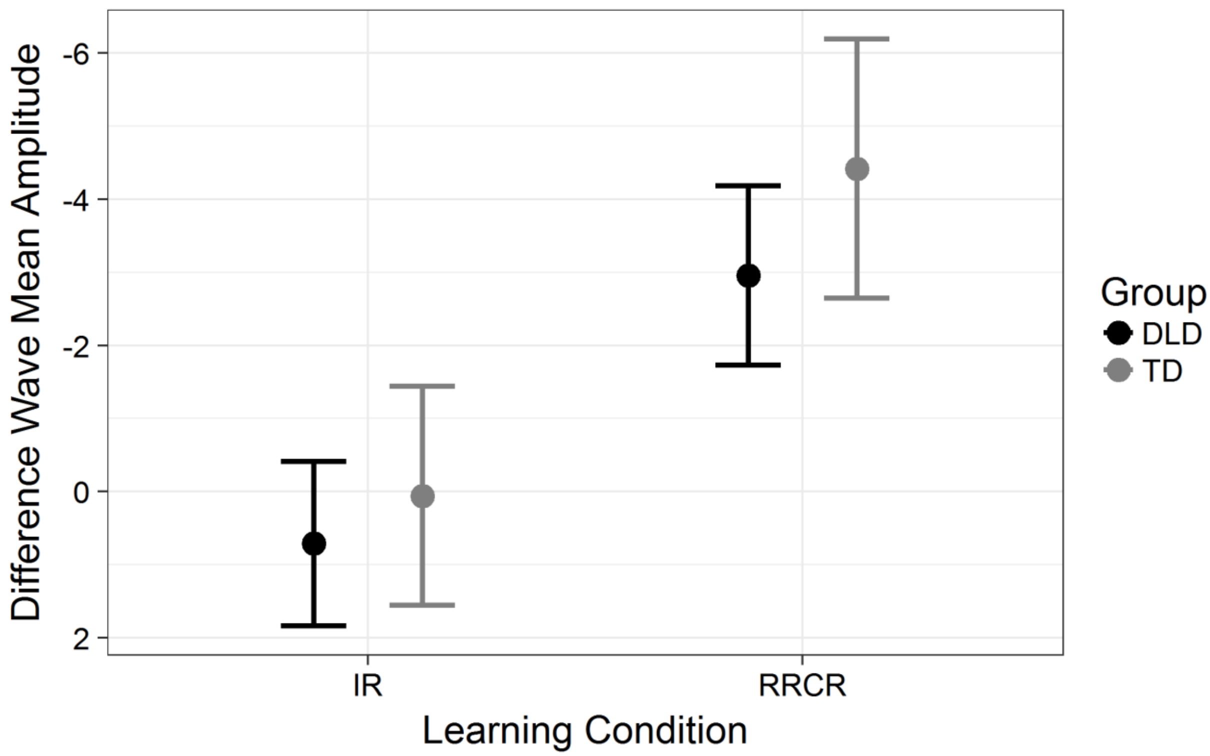

Figure 9. Difference Wave Mean Amplitudes according to Learning Condition and Group

Note. Circles represent the mean of the difference wave (mismatch trials – match trials) within the 500 ms to 700 ms analysis window for the region of interest (ROI). Error bars represent standard errors.

Table 1. Model Results for the Word Form Recall Outcome in Experiment 1

	Model A			**Model B**			**Model C**			**Model D**		
fixed effects	*b*	95% CI		*b*	95% CI		*b*	95% CI		*b*	95% CI	
group (DLD vs. TD)	-1.02	-2.14	0.1	-1.34	-2.8	0.13	-1.28	-2.45	-0.11	-1.6	-3.1	-0.1
condition (022 vs 000)	2.53	1.7	3.37	2.53	1.69	3.37	2.78	1.55	4.01	2.78	1.55	4.02
time (1wk vs. 5min)	-0.31	-0.64	0.01	-0.31	-0.64	0.01	-0.72	-1.28	-0.16	-0.72	-1.27	-0.16
group X time							0.63	-0.02	1.27	0.63	-0.02	1.27
condition X time							0.19	-0.46	0.83	0.19	-0.45	0.83
group X condition							-0.69	-2.37	0.99	-0.69	-2.37	1
PPVT				-0.02	-0.08	0.03				-0.02	-0.08	0.03
mother's education				0.06	-0.31	0.43				0.06	-0.31	0.43
intercept	2.46	1.66	3.27	4.09	-3.46	11.63	2.64	1.8	3.48	4.27	-3.28	11.82
random effects	σ^2			σ^2			σ^2			σ^2		
condition	4.94	2.76	8.83	4.96	2.77	8.88	5.02	2.79	9.04	5.05	2.81	9.09
intercept	2.19	1.2	4	2.31	1.24	4.29	2.2	1.2	4	2.32	1.25	4.29

Note. *N* = 32, observations = 128. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 2. Simple Effects Table for Word Form Recall for Model D in Table 1

	b	95% CI		b_{std}
1 week versus 5 min. for DLD group	-0.09	-0.65	0.46	-0.03
1 week versus 5 min. for TD group	-0.72	-1.27	-0.16	-0.26
DLD versus TD for 5 minutes	-1.60	-3.10	-0.10	-0.59
DLD versus TD for 1 week	-0.98	-2.48	0.53	-0.36

Note. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 3. Estimated Effects for the Word Form Learning Trajectories of the IR/0-0-0 Condition in Experiment 1

	Quadratic			**Quadratic w/groups**		
	b	95% CI		*b*	95% CI	
fixed effects						
intercept	4.94	4.66	5.22	5.25	4.88	5.63
slope	0.32	0.13	0.51	0.29	0.03	0.56
quadratic	-0.04	-0.07	0	-0.04	-0.09	0.01
group (DLD vs. TD)				-0.63	-1.15	-0.11
group X slope				0.04	-0.21	0.29
group X quadratic				0	-0.01	0.01
random effects	σ^2			σ^2		
intercept	0.32	0.17	0.58	0.25	0.13	0.48

Note. *N* = 32, observations = 192. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 4. Model Results for the Meaning Recall Outcome in Experiment 1

	Model A			**Model B**		
fixed effects	*b*	95% CI		*b*	95% CI	
group (DLD vs. TD)	-1.89	-3.38	-0.41	-1.96	-3.93	0.02
condition (022 vs 000)	0.64	-0.13	1.41	0.64	-0.14	1.42
time (1wk vs. 5min)	-0.11	-0.55	0.33	-0.11	-0.55	0.33
PPVT				-0.01	-0.08	0.063
mother's education				0.09	-0.41	0.587
intercept	10.61	9.53	11.69	10.25	0.07	20.42
random effects	σ^2			σ^2		
condition	3.32	1.50	7.33	3.43	1.58	7.45
intercept	3.92	2.09	7.34	4.26	2.25	8.04

Note. *N* = 32, observations = 128. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 5. Estimated Effects for the Word Meaning Learning Trajectories of the IR/0-0-0 Condition in Experiment 1

	Linear				**Linear w/ groups**			
	b	95% CI		Cohen's *d*	*b*	95% CI		Cohen's *d*
fixed effects								
intercept	5.33	5.15	5.50	-0.40	5.56	5.32	5.79	-0.06
slope	0.11	0.06	0.16	0.16	0.09	0.02	0.15	0.13
group (DLD vs. TD)					-0.46	-0.79	-0.12	-0.68
group X slope					0.04	-0.05	0.14	0.07
random effects	σ^2				σ^2			
intercept	0.07	0.03	0.18		0.05	0.02	0.15	

Note. *N* = 32, observations = 192. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 6. Model Results for the Form-Referent Link Recognition Outcome in Experiment 1

	Model A			**Model B**			**Model C**			**Model D**		
fixed effects	*b*	95% CI		*b*	95% CI		*b*	95% CI		*b*	95% CI	
group (DLD vs. TD)	-2.24	-3.64	-0.84	-2.01	-3.84	-0.18	-2.5	-3.92	-1.08	-2.31	-4.16	-0.46
condition (022 vs 000)	0.47	-0.07	1.01	0.47	-0.07	1.01	-0.13	-0.85	0.6	-0.13	-0.85	0.6
group X condition							1.19	0.17	2.21	1.19	0.17	2.21
PPVT				0.02	-0.04	0.09				0.02	-0.04	0.09
mother's education				-0.19	-0.65	0.27				-0.19	-0.65	0.27
intercept	11.25	10.25	12.24	11.63	2.19	21.06	11.38	10.37	12.38	11.78	2.34	21.22
random effects	σ^2			σ^2			σ^2			σ^2		
condition	1.38	0.22	8.5	1.2	0.14	10.27	1.22	0.2	7.49	1.06	0.12	9.05
intercept	3.66	2.02	6.65	3.78	2.05	6.98	3.71	2.08	6.64	3.84	2.11	6.98

Note. *N* = 32, observations = 64. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 7. Simple Effects Table for Form-Referent Link Recognition for Model D in Table 6

	b	95% CI		b_{std}
022 versus 000 for DLD group	1.06	0.34	1.78	0.45
022 versus 000 for TD group	-0.13	-0.85	0.60	-0.05
DLD versus TD for 000 condition	-2.31	-4.16	-0.46	-0.97
DLD versus TD for 022 condition	-1.12	-3.10	0.86	-0.47

Note. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.0.5

Table 8. Model Results for the Match-Mismatch Task Behavioral Judgments

fixed effects	*b*	95% CI	
group (DLD vs. TD)	0.07	-0.02	0.16
condition (IR vs RRCR)	0.03	0.01	0.05
set (set 1 vs. set 2)	0.01	-0.01	0.04
condition X group	-0.02	-0.07	0.02
condition X set	0.01	-0.04	0.06
group X set	0.00	-0.05	0.05
group X condition X set	0.06	-0.03	0.15
intercept	0.64	0.60	0.69
random effects	σ^2		
intercept	0.01	0.01	0.02
error	0.00	0.00	0.01

Note. *N* = 27, observations = 100. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 9. ERP Within-Condition Mixed-Effect Models

	IR 300-500 ms			**IR 500-700 ms**			**RRCR 300-500 ms**			**RRCR 500-700 ms**		
fixed effects	b	95% CI		b	95% CI		b	95% CI		b	95% CI	
set[1]	-2.18	-3.02	-1.34	-2.26	-3.16	-1.37	-1.27	-2.14	-0.40	-2.37	-3.32	-1.41
group[2]	3.23	-1.48	7.93	3.27	-1.40	7.94	2.62	-1.33	6.58	3.41	-0.24	7.06
trial type[3]	-0.78	-1.92	0.37	-0.17	-1.38	1.05	2.09	0.90	3.27	4.42	3.12	5.72
group X trial type	-0.15	-1.77	1.47	-0.62	-2.34	1.09	-0.62	-2.30	1.05	-0.96	-2.80	0.88
intercept	8.19	4.62	11.77	6.93	3.35	10.51	4.87	1.79	7.47	3.11	0.18	6.04
random effects	σ^2			σ^2			σ^2			σ^2		
intercept	36.46	20.60	64.53	35.61	20.07	63.17	24.85	13.90	44.44	20.35	11.20	36.97
error	34.13	30.88	37.74	38.27	34.62	42.31	36.49	33.01	40.34	43.99	39.79	48.64

Note. [1] set 1 vs. set 2, [2] DLD vs. TD, [3] match vs. mismatch, *N* = 27, observations = 800. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Table 10. ERP Between-Condition Mixed-Effect Model

	Difference Waves 500-700 ms		
fixed effects	b	95% CI	
set[1]	3.58	2.56	4.59
group[2]	1.36	-1.70	4.41
learning condition[3]	4.59	3.20	5.98
group X learning condition	-0.34	-2.30	1.63
intercept	-6.35	-8.95	-3.75
random effects	σ^2		
intercept	12.87	6.85	24.17
error	50.11	45.32	55.40

Note. [1] set 1 vs. set 2, [2] DLD vs. TD, [3] RRCR vs. IR, *N* = 27, observations = 800. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha = 0.05.

Supplemental Materials

Supplemental Table 1.

Additional Model Results for the Word Form Outcome

	Model E			Model F		
fixed effects	b	95% CI		b	95% CI	
group (DLD vs. TD)	-1.38	-2.59	-0.16	-1.69	-3.23	-0.16
condition (022 vs 000)	2.69	1.41	3.96	2.69	1.41	3.96
time (1wk vs. 5min)	-0.81	-1.46	-0.17	-0.81	-1.46	-0.17
group X time	0.81	-0.10	1.73	0.81	-0.10	1.72
cond X time	0.38	-0.54	1.29	0.38	-0.54	1.29
group X cond	-0.50	-2.30	1.30	-0.50	-2.30	1.30
group X cond X time	-0.38	-1.67	0.92	-0.38	-1.66	0.91
PPVT				-0.02	-0.08	0.03
mother's education				0.06	-0.31	0.43
intercept	2.69	1.83	3.54	4.31	-3.24	11.87
random effects	σ^2			σ^2		
condition	5.01	2.78	9.03	5.04	2.80	9.08
intercept	2.19	1.20	4.00	2.31	1.25	4.29

*Note. N = 32, observations = 128. E*ffects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. The current models extend the models presented in Table 1 in the main text. Notably, the non-significant interactions across group, learning condition, and time indicate that the group effect in Model A in Table 1 does not differ across time and learning condition, and the learning condition effect does not differ across group or time.

Supplemental Table 2.

Additional Model Results for Meaning Outcome

	Model B			**Model C**			**Model D**		
fixed effects	b	95% CI		b	95% CI		b	95% CI	
group (DLD vs. TD)	-1.52	-3.17	0.14	-1.89	-3.46	-0.32	-1.95	-3.98	0.08
condition (022 vs 000)	0.64	0.05	1.24	0.48	-0.70	1.67	0.48	-0.71	1.68
time (1wk vs. 5min)	-0.11	-0.70	0.49	0.05	-0.73	0.83	0.05	-0.73	0.82
group X time				-0.16	-1.06	0.74	-0.16	-1.05	0.74
cond X time				-0.16	-1.06	0.74	-0.16	-1.05	0.74
group X cond				0.47	-1.08	2.02	0.47	-1.10	2.04
group X cond X time									
PPVT	0.01	-0.05	0.07				-0.01	-0.08	0.06
mother's education	0.06	-0.36	0.48				0.09	-0.41	0.59
intercept	8.49	-0.04	17.02	10.57	9.44	11.70	10.19	0.03	20.35
random effects	σ^2			σ^2			σ^2		
condition				3.33	1.47	7.54	3.45	1.55	7.67
intercept	2.72	1.39	5.32	3.88	2.06	7.31	4.22	2.22	8.01

Note. *N* = 32, observations = 128. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Parameters are equivalent in some cases due to identical outcome mean values for different combinations of independent variable values. Model B does not include the random effect for condition; in this model, there is a significant effect of learning condition, with higher performance in the 022 condition. Given that the interactions across group, time, and learning condition are not significant, the more parsimonious models that identify a significant fixed effect of group are presented in Table 4 within the main text.

Supplemental Table 3.

Additional Model Results for the Meaning Outcome

	Model E			Model F		
fixed effects	b	95% CI		b	95% CI	
group (DLD vs. TD)	-1.63	-3.26	0.01	-1.69	-3.77	0.40
condition (022 vs 000)	0.75	-0.52	2.02	0.75	-0.53	2.03
time (1wk vs. 5min)	0.31	-0.58	1.21	0.31	-0.58	1.20
group X time	-0.69	-1.96	0.58	-0.69	-1.95	0.57
cond X time	-0.69	-1.96	0.58	-0.69	-1.95	0.57
group X cond	-0.06	-1.86	1.73	-0.06	-1.87	1.74
group X cond X time	1.06	-0.73	2.86	1.06	-0.72	2.84
PPVT				-0.01	-0.08	0.06
mother's education				0.09	-0.41	0.59
intercept	10.44	9.28	11.59	10.07	-0.11	20.24
random effects	σ^2			σ^2		
condition	3.35	1.49	7.56	3.47	1.57	7.69
intercept	3.89	2.06	7.33	4.23	2.23	8.03

Note. *N* = 32, observations = 128. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Parameters are equivalent in some cases due to identical outcome mean values for different combinations of independent variable values. Given that the interactions of group, time, and learning condition are not significant, the more parsimonious models that identify a significant fixed effect of group are presented in Table 4 within the main text.

Supplemental Table 4.
Full ERP Within-Condition Mixed-Effect Models with with ROI Electrodes

	IR 300-500 ms			**IR 500-700 ms**			**RRCR 300-500 ms**			**RRCR 500-700 ms**		
fixed effects	b	95% CI		b	95% CI		b	95% CI		b	95% CI	
set[1]	-2.18	-3.02	-1.34	-2.26	-3.16	-1.37	-1.27	-2.14	-0.40	-2.37	-3.32	-1.41
group[2]	3.23	-1.48	7.93	3.27	-1.40	7.94	2.62	-1.33	6.58	3.41	-0.24	7.06
trial type[3]	-0.78	-1.92	0.37	-0.17	-1.38	1.05	2.09	0.90	3.27	4.42	3.12	5.72
group x trial type	-0.15	-1.77	1.47	-0.62	-2.34	1.09	-0.62	-2.30	1.05	-0.96	-2.80	0.88
electrode site[4]												
PO3	-3.48	-5.10	-1.87	-1.69	-3.40	0.03	-3.83	-5.51	-2.16	-1.97	-3.80	-0.13
O1	-9.59	-11.21	-7.97	-5.76	-7.48	-4.05	-9.75	-11.43	-8.08	-6.23	-8.07	-4.40
Pz	2.04	0.43	3.66	2.45	0.74	4.17	2.23	0.56	3.91	2.40	0.56	4.24
Oz	-6.79	-8.41	-5.17	-3.53	-5.25	-1.82	-6.96	-8.63	-5.29	-4.06	-5.90	-2.22
P4	-0.47	-2.09	1.15	0.32	-1.40	2.03	0.48	-1.20	2.15	1.03	-0.81	2.87
PO4	-2.91	-4.53	-1.29	-0.79	-2.51	0.92	-2.65	-4.32	-0.97	-0.79	-2.63	1.05
O2	-9.83	-11.45	-8.22	-6.04	-7.75	-4.32	-9.15	-10.83	-7.95	-6.12	-7.96	-4.28
intercept	8.19	4.62	11.77	6.93	3.35	10.51	4.87	1.79	7.47	3.11	0.18	6.04
random effects	σ2			σ2			σ2			σ2		
intercept	36.46	20.60	64.53	35.61	20.07	63.17	24.85	13.90	44.44	20.35	11.20	36.97
error	34.13	30.88	37.74	38.27	34.62	42.31	36.49	33.01	40.34	43.99	39.79	48.64

Note. [1] set 1 vs. set 2, [2] DLD vs. TD, [3] match vs. mismatch, [4] reference electrode (site 1) is P3, *N* = 27, observations = 800. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Significant differences are observed between word learning sets and, within the RRCR/022 condition, trial type fixed effects across the ROI.

Supplemental Table 5.

ERP Between-Condition Mixed-Effect Model with ROI Electrodes

	Difference Waves 500-700 ms		
fixed effects	b	95% CI	
set[1]	3.58	2.56	4.59
group[2]	1.36	-1.70	4.41
learning condition[3]	4.59	3.20	5.98
group x learning condition	-0.34	-2.30	1.63
electrode site[4]			
PO3	-0.08	-2.05	1.88
O1	-0.65	-2.61	1.31
Pz	0.77	-1.19	2.73
Oz	-0.39	-2.35	1.57
P4	0.93	-1.03	2.89
PO4	0.29	-1.67	2.25
O2	-0.60	-2.56	1.36
intercept	-6.35	-8.95	-3.75
random effects	σ^2		
intercept	12.87	6.85	24.17
error	50.11	45.32	55.40

Note. [1] set 1 vs. set 2, [2] DLD vs. TD, [3] RRCR vs. IR, [4] reference electrode (site 1) is P3, *N* = 27, observations = 800. Effects with 95% confidence intervals that do not include 0 are statistically significant at alpha=0.05. Across the ROI, there is a significant fixed effect of learning condition, with a larger N400 elicited in the RRCR/022 condition relative to the IR/000 condition, and a significant fixed effect of set.