ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Can Retrieval Practice Facilitate Verb Learning in Children With Developmental Language Disorder and Their Peers With Typical Language Development?

Laurence B. Leonard,[a] Patricia Deevy,[a] Sabrina Horvath,[b] Sharon L. Christ,[c] Jeffrey Karpicke,[d] and Justin B. Kueser[a]

[a] Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN [b] Division of Speech-Language Pathology, Department of Rehabilitation Sciences, Medical University of South Carolina, Charleston [c] Department of Human Development and Family Science, Purdue University, West Lafayette, IN [d] Department of Psychological Sciences, Purdue University, West Lafayette, IN

ARTICLE INFO

Article History:
Received September 1, 2022
Revision received December 20, 2022
Accepted December 20, 2022

Editor-in-Chief: Stephen M. Camarata
Editor: Mary Alt

https://doi.org/10.1044/2022_JSLHR-22-00509

ABSTRACT

Purpose: Children with developmental language disorder (DLD) have well-documented verb learning difficulties. In this study, we asked whether the inclusion of retrieval practice during the learning period would facilitate these children's verb learning relative to a similar procedure that provided no retrieval opportunities.
Method: Eleven children with DLD (M_{age} = 60.09 months) and 12 children with typical language development (TD; M_{age} = 59.92 months) learned four novel verbs in a repeated spaced retrieval (RSR) condition and four novel verbs in a repeated study (RS) condition. The words in the two conditions were heard an equal number of times, in the context of video-recorded actors performing novel actions.
Results: Recall testing immediately after the learning period and 1 week later revealed greater recall for novel verbs in the RSR condition than for novel verbs in the RS condition. This was true for both groups, and for immediate as well as 1-week testing. The RSR advantage remained when children had to recall the novel verbs while watching new actors perform the novel actions. However, when tested in contexts requiring the children to inflect the novel verbs with *–ing* for the first time, the children with DLD were much less likely to do so than their peers with TD. Even words in the RSR condition were only inconsistently inflected.
Conclusions: Retrieval practice provides benefits to verb learning—an important finding given the challenges that verbs present to children with DLD. However, these benefits do not appear to automatically translate to the process of adding inflections to newly learned verbs but rather appear to be limited to the operations of learning the verbs' phonetic forms and mapping these forms onto associated actions.

Children with developmental language disorder (DLD) represent one of the largest groups of children experiencing difficulties acquiring language (Bishop et al., 2017). These children have significant weaknesses in language that cannot be attributed to biomedical conditions such as neurological damage or disease, hearing impairment, intellectual disability, or autism spectrum disorder (ASD). A diagnosis of DLD can co-occur with weaknesses in areas such as attention or motor coordination, provided that these other weaknesses do not serve as a plausible cause of the language disorder. Although the language symptoms change with maturation, DLD in most cases represents a lifelong condition.

Among the problems with language in the DLD population is a deficit in word learning. This is seen not only in children (e.g., McGregor et al., 2013, 2021; Storkel et al., 2017) but also in adolescents and adults (e.g.,

Correspondence to Laurence B. Leonard: xdxl@purdue.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

McGregor, Arbisi, & Eden, 2017; McGregor et al., 2020; McGregor, Gordon, et al., 2017). In fact, the gap between vocabulary abilities in individuals with DLD and those of same-age peers may grow somewhat larger over time (Rice & Hoffman, 2015). These children's deficits in word learning apply both to real words and experimenter-designed novel words, which are often used to control for familiarity and vocabulary differences between children with DLD and their peers with typical language development (TD). The results of these studies are consistent in showing weaker novel word learning by the children with DLD than by their peers (e.g., Alt, 2011; Gray, 2003; see meta-analysis by Kan & Windsor, 2010).

Many of these studies include both measures of the children's word encoding and the children's longer term retention of the words. Encoding is the process of forming temporary representations of the word. With additional experience, these representations can be refined with greater phonetic precision and richer semantic detail. Longer term retention refers to the ability to recall a word days or weeks after the original learning period (Because longer term retention is often assessed through retrieval tasks, it is more accurate to say we *infer* retention based on the individual's success in retrieving the word well after the learning period. On the other hand, a failure of retrieval in this instance might be due to lack of retention or a failure to retrieve the word in the moment.). Of the two processes of encoding and longer term retention, it is encoding that appears to be the weaker in individuals with DLD (Bishop & Hsu, 2015; Gordon et al., 2020; Gray, 2004; Jackson et al., 2021). When individuals with DLD manage to consistently recall a word shortly after a learning period, they show considerable success in retaining that word over several days or longer (e.g., McGregor et al., 2020; McGregor, Gordon, et al., 2017).

Distinctions are also made in the literature between learning word forms and learning word meanings. Both can be challenging for individuals with DLD. However, learning a word form (e.g., /mærəθɔn/) appears to be more difficult than learning its meaning (e.g., a race of approximately 26.2 miles; e.g., Gray, 2004; McGregor, Arbisi, & Eden, 2017).

In this study, we focus on how well children with DLD learn novel word forms representing verbs (hereafter, simply "novel verbs") and their associated referent actions. We ask whether the children's learning and recall of novel verbs might be improved through our inclusion of "retrieval practice"—that is, providing the children with multiple opportunities to recall the word during the learning period itself. We begin by discussing why verb learning might be especially important for children with DLD, and we follow with how retrieval practice might improve this learning. We then describe a novel verb learning study in which we pit learning through both exposure and frequent opportunities for retrieval against a procedure involving identical exposure without retrieval. Same-age children with TD also participated to determine if retrieval benefits for novel verb learning have broader application.

Why Verbs?

In the literature on TD, there is a long history of research showing that verb learning is more difficult than noun learning (e.g., Gentner, 1978; Gleitman, 1990). This is true across different languages even when input frequency is controlled (e.g., Frank et al., 2021; Imai et al., 2005; Leonard et al., 1981). From a conceptual standpoint, verbs seem more complex than nouns in particular because, whereas many nouns can be individuated, verbs are inherently relational (Gentner, 1978). Perhaps because of this difference, verbs are also less imageable than nouns—a factor known to predict age of acquisition of vocabulary items in toddlerhood (W. Ma et al., 2009).

A critical component of verb learning involves making use of the syntactic contexts in which the verb appears, a process known as syntactic bootstrapping (Gleitman, 1990). Consider, for example, the event of gift-giving. This action may be described from the perspective of the giver, as in, "The girl is giving the boy the gift," or it may be described from the perspective of the receiver, as in, "The boy is receiving the gift from the girl." Absent linguistic context, a child who sees this scene and hears only a novel verb is unable to disambiguate these possible interpretations. This use of linguistic context to infer novel verb meaning has been demonstrated, in principle, in studies in which adults (e.g., Fitch et al., 2021; Gillette et al., 1999) and school-age children (e.g., Piccin & Waxman, 2007) attempt to guess the words spoken by mothers to their young children in video clips with the audio removed. When only visual information is provided, participants are largely unsuccessful at identifying target verbs. However, when linguistic information is added, participants' ability to identify verbs rises sharply. In novel verb learning studies, children as young as 2 years of age appear sensitive to the linguistic context of an unfamiliar verb (Fisher et al., 1994; Naigles, 1990).

As much as verb learning is challenging for children with TD, it is even more difficult for children with DLD (e.g., Fletcher & Peters, 1984; Watkins et al., 1993). This is true also for learning novel verbs. For example, although children with DLD lag behind their same-age peers when learning novel nouns, this learning gap is even greater for novel verbs (see meta-analysis in Kan & Windsor, 2010). Part of this difficulty may rest in the fact

that morphosyntax—an essential part of linguistic context—plays an important role in verb learning and is often an area of special weakness in children with DLD. This is readily seen in studies focusing on English; many 5-year-old English-speaking children with DLD show weaker morphosyntactic skills than children with TD who are 2 years younger, even when the two groups are matched on factors such as sentence length and vocabulary size (see review in Leonard, 2014).

Children with DLD may also differ from peers in how they view details of verb meanings. However, the direction of these differences has not been consistent across studies. For example, Kelly and Rice (1994) found that children with DLD were more likely to interpret a set of novel verbs in terms of their motions, whereas peers interpreted the same novel verbs as reflecting a change of state. Echoing these findings, Ebbels et al. (2012) reported that children with DLD had more difficulty with change-of-state meanings as compared to their typically developing peers. In contrast, Penner et al. (2003) found that German-speaking school-aged children with DLD had a bias toward change-of-state meanings over manner meanings, whereas their typically developing peers showed no such bias. Similarly, in a study of late talkers, Horvath and colleagues (Horvath et al., 2019, 2022) found that the children had a bias toward result meanings over manner meanings in the composition of their verb vocabularies. Still, other works show that children with DLD learn fewer semantic features of novel verbs than children with TD (Alt et al., 2004). Finally, there is also evidence that, relative to typically developing children, verb use and comprehension by children with DLD are not as tied to whether the observed action depicts a completed event (Leonard et al., 2007; Leonard & Deevy, 2010).

In this study, we aimed to constrain children's interpretations by using novel intransitive verbs with simple semantics referring to novel body motions that were performed by familiar grammatical subjects. The sentence contexts used to present the novel verbs were contexts where only verbs could appear. We hoped that, together, these provisions would ensure that children in both groups would be equally likely to learn the simple meanings associated with the verbs and to interpret the novel words as verbs, enabling a fair test of both groups' novel verb learning.

The Benefits of Retrieval Practice

When we attempt to recall information that we just studied, we are doing more than testing what we just learned. These self-tests are also a form of learning. This notion of recall-as-learning is an old one; it first appeared in the literature in the late 1880s and became the subject of rigorous empirical testing beginning with Abbott in 1909. The last 20 years has been a period of especially strong interest in this topic (see reviews in Fazio & Marsh, 2019; Rowland, 2014). Children as well as adults have served as research participants, and the material to be studied and recalled by participants has varied greatly.

Recent efforts have included studies of retrieval during novel word learning by children and adults with DLD (e.g., Chen & Liu, 2014; McGregor, Gordon, et al., 2017; see recent reviews in Gordon, 2020; Leonard & Deevy, 2020). Each study has found clear benefits from retrieval. In a series of such studies, Leonard and colleagues examined the role of retrieval in the novel word learning of 4- and 5-year-old children with DLD and their typically developing same-age peers. These studies made use of two fundamental components of retrieval procedures seen in the memory literature—the inclusion of multiple retrieval trials of each word during the learning period (Roediger & Karpicke, 2006), and the spacing of the retrieval trials so that some retrieval attempts of the word occurred after several other words had intervened (e.g., Karpicke & Roediger, 2007). In each of the studies by Leonard and colleagues, half of the novel words were learned in this type of repeated spaced retrieval (RSR) condition. The other half were learned either without the opportunities for retrieval (the "repeated study" or RS condition; e.g., Leonard, Karpicke, et al., 2019), or were always retrieved in an immediate retrieval condition, with no spacing (e.g., Haebig et al., 2019). The novel words to be learned included those representing nouns (e.g., Leonard, Karpicke, et al., 2019) or adjectives (Leonard, Deevy, et al., 2019). In each study, words learned in the RSR condition showed better recall than those learned in the comparison condition (see Leonard et al., 2021). This was true for the typically developing children as well as the children with DLD. In a subsequent study that represented a more stringent test of RSR, one condition had more spaced retrieval trials but fewer study trials (meaning fewer exposures to the word) and the comparison condition had more study trials (more exposures to the word) but fewer spaced retrieval opportunities. The "more retrieval, less study" condition resulted in greater recall (Leonard et al., 2020).

The RSR condition used in the previous studies conformed to much of what is assumed in the memory literature. In particular, the "effortful" retrieval that is required to successfully retrieve a word promotes longer term recall. Two other elements of the RSR condition are worthy of note. First, each of the studies eased into spaced retrieval by using one or two immediate retrieval trials before shifting to spaced retrieval trials. Although success on immediate retrieval trials does not directly predict long-term recall, successful immediate retrieval trials do

predict greater success on subsequent spaced retrieval trials during learning, which, in turn, predicts long-term recall (Kueser et al., 2021). We interpreted this finding as suggesting that strategic placement of immediate retrieval trials assists shorter term encoding, which successful spaced retrieval can then stabilize to permit later recall.

A second notable element is that we inserted an additional study trial directly after a retrieval trial in the RSR condition. This study trial served as a type of feedback because it contained the word that the child had just attempted to retrieve. Such feedback has been found to be helpful to learners in retrieval studies (e.g., X. Ma et al., 2020) and especially so when the learner's previous retrieval attempt was incorrect (e.g., Rowland & DeLosh, 2015) or correct but the learner had little confidence in its accuracy (Butler et al., 2008). The early spaced retrieval trials in the RSR condition often involve unsuccessful retrieval attempts and therefore inclusion of feedback in the form of a study trial after a retrieval trial likely assisted the children's learning and later recall.

Given the apparent benefits of RSR, it seemed especially important to determine if these same benefits accrue when children are faced with learning new verbs—an area of special weakness for children with DLD. Accordingly, we adapted our procedures to determine if RSR could lead to more gains than RS when children were asked to learn novel verbs referring to video-recorded actions. To learn more about how RSR affected verb learning in particular, we took two additional steps. First, during subsequent testing, the children had to generalize the novel verbs to new events by recalling the words when the actors in the video-recordings were not the same as those performing the actions during the learning period. This step allowed us to determine if the actions were viewed as separable elements and not tied to particular actors. Second, for novel verbs that were recalled, we also assessed the children's ability to supply the subject noun along with inflecting the novel verb with *–ing* (e.g., "man (is) /faɪbɪŋ/") even though inflections had not been presented with the verbs during the learning period. This step permitted us to assess whether the children were capable of using the novel words in a manner consistent with the structural contexts associated with verbs. These added steps assisted us in determining whether the learned words actually had verb status for the children.

Method

All recruitment and experimental procedures were approved by the authors' institutional review board. Written consent was obtained from the children's families, and verbal assent was provided by the children.

Participants

Selection of Children With DLD

Eleven children (three girls, eight boys) met the selection criteria for exhibiting DLD. The children were recruited through letters and presentations given to speech-language pathologists and preschool directors in the Greater Lafayette, Indiana area. Data from all of the children meeting our selection criteria were included in the study, and all data obtained from these children were used in the analyses. The mean age of the children in the DLD group was 60.09 months (SD = 3.83, range: 54–67 months). All children were White; one child was Hispanic/Latino. All were monolingual speakers of English. These children were enrolled in language intervention or scheduled to begin an intervention program. All scored below a standard score of 87 on the Structured Photographic Expressive Language Test–Preschool 2 (SPELT-P2; Dawson et al., 2005). These scores are below the cutoff yielding good sensitivity and specificity for this test (Greenslade et al., 2009). Each child scored above 75 (on the Kaufman Assessment Battery for Children–Second Edition (KABC-2; Kaufman & Kaufman, 2004), a test of nonverbal intelligence. All scores on the Childhood Autism Rating Scale–Second Edition (CARS-2; Schopler et al., 2010) were in the "minimal-to-no-symptoms" of ASD range. Finally, all children passed a pure-tone hearing screening in both ears at 20 dB at 500, 1000, 2000, and 4000 Hz.

Selection of Children With TD

The remaining 12 children (three girls, nine boys) met the criteria for TD. These children were recruited to be very similar in age to the children with DLD (M = 59.92 months, SD = 3.42, range: 54–65 months). The children were recruited through fliers distributed in preschools in the Greater Lafayette, Indiana area; through the Research Participant Registry in the Department of Speech, Language, and Hearing Sciences at Purdue University; or through parents hearing about the research study from other parents. Ten children were White, and two were identified by their parents as multiracial. Eleven of these children were monolingual speakers of English. The remaining White child was Hispanic/Latino; the father spoke Portuguese in the home, and the mother spoke English. The child spoke English as a native language. No language or other developmental problems were reported for these children, and each child passed the pure-tone hearing screening. All children in this group scored above 87 on the SPELT-P2, and all KABC-2 scores were above 75. Given the children's reported typical developmental history, we did not administer the CARS-2 to this group. Due to illnesses, two of the children in the TD group provided only partial data for the

study. One child participated in only one set of novel verbs, and the other child was absent for the 1-week testing for one of the sets. Our method of data analysis allowed us to include these children despite their missing data (see Data Analysis section below).

Covariates and Other Descriptive Measures

Several additional measures were obtained that were not part of the selection criteria; scores on these measures were free to vary. Two of these were used as covariates and had been consistently used in our previous studies. One of these was maternal education, measured in years of education. The second covariate was the standard score on the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007). This is a test of (static) accumulated receptive vocabulary knowledge, in contrast to our experimental measures of (dynamic) vocabulary learning. In our previous studies, as in the studies of other investigators, PPVT-4 scores have been lower in our DLD groups than in our TD groups. However, these scores do not necessarily predict vocabulary learning in controlled studies. For example, in principle, a child might have a lower PPVT-4 score due to limited access to a rich linguistic environment. Yet, given comparable exposure as in a controlled study, that child might show the same capability to acquire new words as children with higher PPVT-4 scores. Our use of the PPVT-4 scores as a covariate was to determine whether, in our particular study, any group differences in novel verb learning might be accounted for by the children's already-accumulated vocabulary knowledge. A third measure—the Expressive Vocabulary Test–Second Edition (Williams, 2007)—was obtained for purposes of additional clinical description. Scores on this test were not used as a covariate. Table 1 provides a summary of the children's scores on all of the tests administered.

Three other measures were obtained specifically to assist us in interpreting the children's responses on the experimental measures. To determine if the children exhibited any unusual phonological patterns that would complicate scoring of the children's novel word productions, we constructed a production task of actual words that used the segments contained in the novel words in the same word positions. For example, for the novel word /nok/, actual words on the production task included *nose* and *book*. These words were presented in short phrases with the target word in final position (e.g., *"read the book"*) that the child was asked to repeat.

Because one of the experimental tasks required responses of three or more morphemes in length (e.g., "Woman /tɛbɪŋ/"), we computed a mean length of utterance (MLU) in morphemes based on a spontaneous language sample from each child in the DLD group. All children in this group showed an adequate MLU for the task (M = 4.53, SD = 0.94, range: 3.01–5.83). A language sample was not obtained from the TD group as there were no concerns about utterance length for these children.

Table 1. A summary of the children's scores on the tests used as selection criteria and the additional measures whose scores were left to vary.

Measure	DLD	TD
Selection criteria		
SPELT-P2		
M	71.27	115.33
SD	15.8	9.69
Range	39–86	95–128
KABC-2		
M	106.18	112.5
SD	15.39	13.79
Range	81–122	90–132
CARS-2		
M	16.09	(N/A)
SD	0.89	
Range	15.5–18.5	
Additional measures		
Maternal education		
M	16.00	15.83
SD	2.05	2.12
Range	13–19	12–18
PPVT-4		
M	97.9	121.58
SD	16.71	10.72
Range	76–118	104–145
EVT-2		
M	94.6	113.67
SD	7.32	11.3
Range	79–109	97–127

Finally, because one of our experimental measures examined the children's ability to add the inflection *–ing* to the novel verbs being learned, we administered a pre-experimental task of 18 items developed by Fey et al. (2017). Children watched toys and props being manipulated, and while the actions were ongoing, the children were to tell a "shy turtle" (hiding in its shell) what actions were taking place ("Tell the turtle. Right now the ___"). Each of the 18 actions could be described by a different familiar verb. Both groups of children showed very high use of adding the progressive inflection *–ing* to the familiar verb. Mean accuracy for the children with DLD was 97.90%; for the children with TD, the mean was 100%. Accuracy on the auxiliary *is* was not a factor in the scoring, as our interest was in the children's ability to add an inflection to the verb.

Materials

Novel Verbs and Actions

Eight novel words were created to serve as the novel verbs. All were consonant–vowel–consonant (CVC) monosyllables. These were /nok/, /jæd/, /pʌm/, /tɛb/, /meɪp/, /faɪb/, /dik/, and /gɪn/. No two words shared the same initial consonant or vowel. The words were divided into two sets of four words each; within each set, two words were assigned to the RSR condition and two words were assigned to the RS condition. Counterbalancing was used to ensure each word appeared in both conditions across children. Within each set, the words in the two conditions were matched on phonotactic probability (average biphone frequency) and neighborhood density, using the child corpus–based values provided in the supplementary materials of Storkel (2013). The two sets were learned sequentially, with 1 week separating the end of testing for the first set and the beginning of the second set.

Eight novel actions were video-recorded to serve as referents for the novel verbs. Each action was performed, in separate recordings, by four different actors, two adult women and two adult men. Two of the recordings for each action, one of a female and one of a male, were shown during the learning period as well as in the postlearning tests. The remaining two recordings were not shown during the learning period; instead, they served as generalization items on the postlearning tests. Each actor appeared in a video used during the learning phase for one of the verb sets; the same actor appeared (in different color clothing) in a video used for a generalization item for the other set of verbs.

Each recording involved the actor performing novel arm and leg movements. The actions did not resemble any action that could be labeled such as hopping, bowing, clapping, or waving. However, the novel actions were similar to these more common actions in terms of degree of motion and body parts used. Video clips were created, which began with the actor standing in place before performing the action for 5–6 s and then returning to the standing position. Actors were recorded with a metronome to ensure consistency in timing across actors and actions.

As described under Procedure section (see below and Table 2), we presented a sequence of study trials and retrieval trials. For study trials, an 18- to 20-s iMovie was created using the 5- to 6-s clips and three audio-recorded sentences that were presented as the action was taking place. All actors were referred to as either "this man" or "this woman" or were referred to by their corresponding pronouns. For example, for one item, the sentences were as follows: "This woman likes to /tɛb/. She really likes to /tɛb/. This woman can really /tɛb/." The first and third sentences provided the subject in noun form, and the second sentence used the pronoun form. Items using male actors were described with "this man" and "he." As seen from the example, the novel verb always appeared in a bare-stem form, in a sentence-final position. In the sentence frame used, the novel verb appeared in a position that only a verb can fill. Audio-recordings of these sentences accompanied the video clips to make up each study trial movie. Sound and video were presented in the following pattern: First, the video clip was shown without sound; as soon as it ended, the first sentence was heard; after the audio stopped, the video clip began to play again, this time accompanied by the second sentence; once this video/audio ended, the third sentence was heard and then the video clip was played one more time (At the time of the study, little was known about video–audio timing relationships in studies of DLD. Therefore, to ensure we would not exclude a more effective timing relationship, we used each of the three timing possibilities.).

For retrieval trials, the same video-recordings were shown and the audio-recording was changed to, for example, "Tell me about the woman. The woman likes to ___," with the child requested to complete the sentence with the novel verb. The video clips were played on loop until the child responded.

To familiarize the children with the task prior to the appearance of the novel verbs and actions, a practice audio- and video-recording of a woman jumping was used. This practice study trial used the sentences, "This woman likes to jump. She really likes to jump. This woman can really jump." The practice retrieval trial was of the form: "Tell me about the woman. The woman likes to ___." Then, to introduce the children to the novel verbs and actions, we presented, for each novel verb, two 5- to 6-s video-recordings of the actors (one actor in each video) to be used during the learning period. The two recordings were presented side-by-side simultaneously, and the children heard, for example, "Wow! They really like to /tɛb/."

Postlearning Tests

Three tests were created to assess the children's learning and recall. The first, referred to here as the "Verb Recall Test," used the same audio-recordings as the retrieval trials used during the learning period (e.g., "Tell me about the woman. The woman likes to ___"). Four video clips were used to test each verb. Two of these (presented as the first and third items for each verb) were the same recordings as during the learning period, with one female and one male actor. These items are referred to as "learned" items. The remaining two items for each verb were video-recordings of a different female and male performing the action. These recordings followed the same timing format as the other recordings. They were presented as the second and fourth items for each verb and are

Table 2. The procedure used during the learning phase and postlearning tests for one of the two sets of four novel verbs.

Set 1, Day 1							
I.	**Familiarization**						
	Familiarization trial: study and retrieval of familiar verb (*jump*)						
	Introduction of novel verbs /nok/, /jæd/, /pʌm/, /tɛb/ and corresponding actions						
II.	**Learning phase**						
	Subject	**Verb**	**Condition**	**Exposure**			
	Block 1						
	$woman_1$	/nok/	RSR	Study–retrieval–study			
	man_1	/jæd/	RS	Study–study			
	man_2	/pʌm/	RSR	Study–retrieval–study			
	$woman_2$	/tɛb/	RS	Study–study			
	man_3	/nok/	RSR	Study–retrieval–study			
	$woman_3$	/jæd/	RS	Study–study			
	$woman_4$	/pʌm/	RSR	Study–retrieval–study			
	man_4	/tɛb/	RS	Study–study			
	$woman_1$	/nok/	RSR	Retrieval–study			
	man_1	/jæd/	RS	Study			
	$woman_2$	/pʌm/	RSR	Retrieval–study			
	man_2	/tɛb/	RS	Study			
	Block 2						
	man_3	/nok/	RSR	Retrieval–study			
	$woman_3$	/jæd/	RS	Study			
	$woman_4$	/pʌm/	RSR	Retrieval–study			
	man_4	/tɛb/	RS	Study			
	$woman_1$	/nok/	RSR	Retrieval–study			
	man_1	/jæd/	RS	Study			
	$woman_2$	/pʌm/	RSR	Retrieval–study			
	man_2	/tɛb/	RS	Study			
	man_3	/nok/	RSR	Retrieval–study			
	$woman_3$	/jæd/	RS	Study			
	$woman_4$	/pʌm/	RSR	Retrieval–study			
	man_4	/tɛb/	RS	Study			
Set 1, Day 2							
I.	**Familiarization: /tɛb/, /pʌm/, /jæd/, /nok/**						
	(Re-)Familiarization trial: Study and retrieval of familiar verb (*jump*)						
	(Re-)Introduction of novel verbs /nok/, /jæd/, /pʌm/, /tɛb/, and corresponding actions						
II.	**Learning phase**						
	Subject	**Verb**	**Condition**	**Exposure**			
	Block 3						
	$woman_2$	/tɛb/	RS	Study			
	man_2	/pʌm/	RSR	Retrieval–study			
	man_1	/jæd/	RS	Study			
	$woman_1$	/nok/	RSR	Retrieval–study			
	man_4	/tɛb/	RS	Study			
	$woman_4$	/pʌm/	RSR	Retrieval–study			
	$woman_3$	/jæd/	RS	Study			
	man_3	/nok/	RSR	Retrieval–study			
	$woman_2$	/tɛb/	RS	Study			
	man_2	/pʌm/	RSR	Retrieval–study			
	man_1	/jæd/	RS	Study			
	$woman_1$	/nok/	RSR	Retrieval–study			

(*table continues*)

Table 2. *(Continued).*

	Block 4						
	man$_4$	/tɛb/	RS	Study			
	woman$_4$	/pʌm/	RSR	Retrieval–study			
	woman$_3$	/jæd/	RS	Study			
	man$_3$	/nok/	RSR	Retrieval–study			
	woman$_2$	/tɛb/	RS	Study			
	man$_2$	/pʌm/	RSR	Retrieval–study			
	man$_1$	/jæd/	RS	Study			
	woman$_1$	/nok/	RSR	Retrieval–study			
	man$_4$	/tɛb/	RS	Study			
	woman$_4$	/pʌm/	RSR	Retrieval–study			
	woman$_3$	/jæd/	RS	Study			
	man$_3$	/nok/	RSR	Retrieval–study			
III.	**5-min break**						
IV.	**Verb Recall Test**						
	Subject	**Verb**	**Condition**	**Item type**			
	man$_2$	/pʌm/	RSR	Learned			
	woman$_2$	/tɛb/	RS	Learned			
	woman$_1$	/nok/	RSR	Learned			
	man$_1$	/jæd/	RS	Learned			
	woman$_5$	/pʌm/	RSR	Generalized			
	man$_5$	/tɛb/	RS	Generalized			
	man$_6$	/nok/	RSR	Generalized			
	woman$_6$	/jæd/	RS	Generalized			
	woman$_4$	/pʌm/	RSR	Learned			
	man$_4$	/tɛb/	RS	Learned			
	man$_3$	/nok/	RSR	Learned			
	woman$_3$	/jæd/	RS	Learned			
	man$_7$	/pʌm/	RSR	Generalized			
	woman$_7$	/tɛb/	RS	Generalized			
	woman$_8$	/nok/	RSR	Generalized			
	man$_8$	/jæd/	RS	Generalized			
Set 1, 1 week later							
I.	**Repeat Verb Recall Test (see Day 2, IV)**						
II.	**Progressive Test**						
	Practice (4 items): woman clapping, waving, jumping, spinning.						
	Progressive Test:						
	Subject	**Verb**	**Condition**	**Item type**			
	man$_1$	/jæd/	RS	Learned			
	woman$_1$	/nok/	RSR	Learned			
	woman$_2$	/tɛb/	RS	Learned			
	man$_2$	/pʌm/	RSR	Learned			
	woman$_6$	/jæd/	RS	Generalized			
	man$_6$	/nok/	RSR	Generalized			
	man$_5$	/tɛb/	RS	Generalized			
	woman$_5$	/pʌm/	RSR	Generalized			
	woman$_3$	/jæd/	RS	Learned			
	man$_3$	/nok/	RSR	Learned			
	man$_4$	/tɛb/	RS	Learned			
	woman$_4$	/pʌm/	RSR	Learned			
	man$_8$	/jæd/	RS	Generalized			
	woman$_8$	/nok/	RSR	Generalized			
	woman$_7$	/tɛb/	RS	Generalized			
	man$_7$	/pʌm/	RSR	Generalized			

(*table continues*)

Table 2. *(Continued).*

III.	Recognition Test					
					Condition	Learned
	woman$_6$ /jæd/	woman$_2$ /tɛb/	**woman$_4$** /pʌm/	woman$_8$ /nok/	RSR	Learned
	woman$_7$ /tɛb/	woman$_3$ /jæd/	woman$_1$ /nok/	woman$_5$ /pʌm/	RS	Generalized
	man$_2$ /pʌm/	**man$_6$** /nok/	man$_1$ /jæd/	man$_5$ /tɛb/	RSR	Generalized
	woman$_8$ /nok/	woman$_7$ /tɛb/	woman$_4$ /pʌm/	**woman$_3$** /jæd/	RS	Learned
	man$_3$ /nok/	man$_7$ /pʌm/	**man$_4$** /tɛb/	man$_8$ /jæd/	RS	Learned
	woman$_5$ /pʌm/	woman$_1$ /nok/	woman$_6$ /jæd/	woman$_2$ /tɛb/	RSR	Generalized
	man$_2$ /pʌm/	**man$_8$** /jæd/	man$_6$ /nok/	man$_4$ /tɛb/	RS	Generalized
	man$_1$ /jæd/	man$_5$ /tɛb/	man$_7$ /pʌm/	**man$_3$** /nok/	RSR	Learned
	man$_2$ /pʌm/	man$_1$ /jæd/	man$_6$ /nok/	**man$_5$** /tɛb/	RS	Generalized
	woman$_7$ /tɛb/	woman$_3$ /jæd/	**woman$_8$** /nok/	woman$_4$ /pʌm/	RSR	Generalized
	man$_1$ /jæd/	man$_7$ /pʌm/	man$_5$ /tɛb/	man$_3$ /nok/	RS	Learned
	man$_4$ /tɛb/	man$_3$ /nok/	man$_8$ /jæd/	**man$_7$** /pʌm/	RSR	Generalized
	woman$_1$ /nok/	**woman$_2$** /tɛb/	woman$_5$ /pʌm/	woman$_6$ /jæd/	RS	Learned
	man$_8$ /jæd/	**man$_2$** /pʌm/	man$_4$ /tɛb/	man$_6$ /nok/	RSR	Learned
	woman$_2$ /tɛb/	woman$_8$ /nok/	**woman$_6$** /jæd/	woman$_4$ /pʌm/	RS	Generalized
	woman$_1$ /nok/	woman$_5$ /pʌm/	woman$_7$ /tɛb/	woman$_3$ /jæd/	RSR	Learned

Note. Distinct male and female actors are designated by distinct subscripts. The target for each recognition task item is shown in bold font; the remaining entries are the foils. RSR = repeated spaced retrieval condition; RS = repeated study condition.

referred to as "generalization" items. Sixteen items were used for this test (4 video clips × 4 verbs).

The second test, referred to here as the "Progressive Test," served both as an additional assessment of the children's novel verb recall and as a measure of the children's ability to inflect the novel verbs with progressive *–ing*. To introduce the children to this test, we video-recorded a woman performing four common actions—jumping, clapping, waving, and spinning—each in a separate video clip. Accompanying each video was the auditory promp, "What's she doing? Tell me about the woman. Right now, the ___." This context obligated responses such as, "Woman (is) jumping." Following these introductory videos, the remaining items were the same video clips as in the Verb Recall Test. However, the audio-recordings were changed to the same prompts used in the introductory items. A response such as, "woman (is) /tɛbɪŋ/," was expected in this context. Again, 16 items comprised this test.

This test provided us with two kinds of information. First, because the video-recordings were identical to those of the Verb Recall Test, we could determine if the children could recall the novel verbs even if they failed to include the inflection, thus producing only the verb stem, as in, "woman (is) /tɛb/." We refer to this measure as the "Progressive Test: Stem Recall." The second kind of information obtained was the degree to which the children added the inflection *–ing* to those verb stems they could correctly recall. We refer to this measure as the "Progressive Test: Inflection Use."

The third test, referred to as the "Recognition Test," required the children to point to the video-recording (from an array of four) that corresponded to the novel action heard in the audio-recording prompt. For this test, 16 items were created. Each item used video clips of four different novel actions, with each novel action serving as the target for four items. Two of the items for each verb showed a novel verb–actor pairing from the learning phase, and two used a novel verb–actor pairing from the generalization test. The three novel verb–actor pairings serving as foils in each item included one from the same phase as the target (e.g., learned) and two from the other phase (e.g., generalization); all actors in the foils were of the same gender as that used in the target video. The target and foil videos played simultaneously for 4 s, providing children an opportunity to preview them. After the preview, children were prompted to find one of the novel verbs as the video continued to play. The prompt was of the form, "Which one shows 'She likes to /tɛb/'?" Before beginning, one familiarization item was provided in which the four "real verb" videos were presented (e.g., "Which one shows 'She likes to clap'?").

Procedure

Table 2 provides the procedure used during the learning phase and postlearning tests for one of the two sets four novel verbs. The study trials and retrieval trials of the learning phase began after both familiarization of

the task with an actual verb ("jump") and the introduction of the novel verbs and associated actions. Within each set, novel verbs from the two conditions appeared in alternating order, with the condition of the first verb (RSR or RS) being counterbalanced across children in each group. Half of the presentations of each novel verb showed a female actor performing the action (e.g., "woman$_1$ /nok/"), and the remaining presentations showed a male actor (e.g., "man$_3$ /nok/"; We use different subscript numbers to indicate different actors performing the actions; see Table 2). All sessions were audio-recorded.

The learning phase took place in two sessions held on consecutive days. Each session was approximately 20 min in duration. On each day, the items were divided into two blocks. Dividing the learning period into separate blocks enabled us to vary the specific order of the particular actors performing the actions (compare Block 1 and Block 2 in Table 2) and the order of the two conditions (RSR first or RS first; compare Block 1 and Block 3 in Table 2).

The first two retrieval trials of each word in the RSR condition involved immediate retrieval because they occurred immediately after a preceding study trial with no intervening words. Retrieval trials, in turn, were always followed by another study trial, which served as a type of feedback (though the child was not told whether the retrieval was correct). All subsequent retrieval trials were spaced retrieval trials because three other words intervened between the retrieval trial and the last time the child heard the word in a study trial. On the second day, the first two retrieval trials of each word in the RSR condition were again immediate retrieval trials. Subsequent retrieval trials were spaced retrieval trials.

During the learning period, words in the RS condition appeared in study trials only. As seen in Table 2, these words received the same number of study trials—and hence exposures—as the words in the RSR condition. All novel verbs were heard a total of 50 times—2 times during familiarization and 48 times in study trials (specifically, 3 times in each of 16 study trials). For novel verbs in the RSR condition, there were 12 retrieval trials for each verb.

Five minutes after the second learning session, we administered the Verb Recall Test, which used the same requests and videos that were used for retrieval trials during the learning period (e.g., "Tell me about the woman. The woman likes to ___"). Each verb was tested 4 times on this test, with each item of the same verb separated by items of the other three verbs (see Table 2), again in a different order from the learning phase. For the first and third items for each verb, the items were learned items (e.g., "woman$_1$ /nok/," "man$_3$ /nok/"); for the second and fourth items for each verb, the items were generalization items (e.g., "man$_6$ /nok/," "woman$_8$ /nok/").

One week later, the Verb Recall Test was re-administered. After a short break, we introduced the Progressive Test that required the children to both recall the novel verb and inflect it with *–ing*. This test began with familiarization items using the common actions of jumping, clapping, waving, and spinning and prompts of the form, "What's she doing? Tell me about the woman. Right now the ___." These familiarization items were repeated if the children did not initially provide a response that included *–ing* attached to the novel verb. The children were then shown the same video-recordings of the novel actions used in the Verb Recall Test, but with the prompt changed to the progressive prompt.

Finally, the children participated in the Recognition Test. Here, the children were asked to point to the video-recording (from an array of four recordings) that matched the novel verb. Sixteen recognition test items were created, targeting each of the four novel verbs 4 times: Two of these used the actors in the learned items, and two used the actors in the generalization items.

Scoring and Reliability

For the Verb Recall Test and the Progressive Test: Stem Recall, we used a multistep process to score the children's recall attempts. Initially, productions were scored as incorrect if they could be interpreted as actual words (e.g., "dance") for the actions. Second, if, subjectively, we regarded a production as a potential attempt at the novel verb, we flagged the production for further inspection. This involved the option of consulting the transcriptions of the child's pronunciations on the real-word production probe administered at the beginning of the study. As noted earlier, the real words on this pre-experiment probe (e.g., *nose*, *book*) contained consonants and vowels used in the novel verbs (e.g., /nok/). If a child had unusual substitutions for particular segments, this was taken into consideration. We then scored the candidate productions of the novel verbs using as our basis the scoring system of Edwards et al. (2004)—the system used in our prior novel noun and novel adjective studies. In this system, each consonant is credited with 1 point each for correct place, manner, and articulation. For vowels, 1 point is awarded for each of length, height, and backness. An additional point is given for correct syllable shape (CVC). For fully accurately pronounced novel verbs, the production was scored as correct. Because all novel verbs had the syllable shape CVC, such productions earned 10 points. For other candidate productions to be scored as correct, the production had to have a higher point total than the total that would be credited if the child had instead been trying to produce one of the other novel words. For example, if a child

produced the novel word /gɪn/ as /gin/, the production would be given 9 points (3 + 2 + 3 + 1). An alternative assumption that the child's production of /gin/ was actually an attempt at /dik/ would lead to a score of 6 points (2 + 3 + 0 + 1). Given the lower score, this alternative interpretation (and others like it) would be excluded and the production would be scored as correct recall of /gɪn/. With this method of scoring, we could make judgments of correct/incorrect while still allowing for some imprecision in the children's productions.

Alternative forms of scoring had certain disadvantages. For example, requiring 100% accuracy for a novel word would mean calling a consistent production of /gɪm/ for /gɪn/ incorrect. Yet, this would ignore the fact that the child had consistently mapped a word form onto an appropriate referent action. An alternative scoring that simply used the full range of points (0–10) for each production without a correct/incorrect judgment would not provide a basis for excluding productions that were haphazard responses or even closer approximations to an incorrect novel verb than to the correct novel verb. However, as a supplement to our main scoring method, we also report in the Results section the phonetic points credited for those responses that were judged to be correct according to our correct/incorrect binary system.

To score the Progressive Test: Inflection Use, we identified instances in which the children recalled the correct novel verb for the item, independent of whether the inflection *–ing* was included in their production. We then computed the percentage of correctly recalled verbs that were inflected. Productions with the subject in noun (e.g., "woman") or pronoun (e.g., "she") form were accepted, and production of auxiliary *is* was not required. Finally, for the recognition test, we computed the number of correct pointing responses by the children.

Interjudge scoring reliability was assessed for the scores on the Verb Recall Test and the Progressive Test: Stem Recall. The data from four children from each group were randomly selected. Two judges independently scored the Verb Recall Tests for each child, broken down according to the conjunction of learning condition (RSR, RS), item type (learned item, generalization item), and time (5 min, 1 week). Likewise, scoring of the Progressive Test: Stem Recall for each child was broken down according to the conjunction of learning condition and item type (note that the Progressive Test: Stem Recall was administered only at the 1-week point). For the Verb Recall Test, this yielded 16 scores to compare for each child; for the Progressive Test: Stem Recall, eight scores were compared for each child. Note that this was a stringent system for determining agreement. For example, on the Verb Recall Test, if one judge scored a child as having four correct generalization items in the RSR condition during 1-week testing and the other judge scored the child as having three correct generalization items in the RSR condition during 1-week testing, we scored this as a disagreement for generalization items in the RSR condition during 1-week testing for that child. Yet, the two judges actually disagreed on only one of the four items. Interjudge reliability proved acceptable for both the Verb Recall Test and the Progressive Test: Stem Recall. For the Verb Recall Test, overall agreement was 100% for the children with DLD and 98% for the children in the TD group. For the Progressive Test: Stem Recall, overall agreement was 91% for both the DLD and TD groups.

Data Analysis

Children's responses on the Verb Recall Test, the Progressive Test: Stem Recall, and the Recognition Test were evaluated using a series of mixed-effects models, with and without the covariates of PPVT-4 standard score and maternal education in years. The outcome was number of correctly recalled words in a set of four words. Diagnostic group (DLD, TD) was a between-participants variable; within-participant variables were learning condition (RSR, RS), item type (learned, generalization), and time (5-min, 1-week, for the Verb Recall Test only). Random slopes for learning condition, time, and item type were included in the models when they were not close to zero. As a result, the random slope for the learning condition variable was included in all of the models.

As noted earlier, two children in the TD group were tested on only one of the two sets at 5 min and/or 1 week (Recall that each set contained an equal number of items from the RSR and RS conditions. Therefore, the included data had the same number of items from the two conditions.). We analyzed data at the level of repeated trials for each set. Therefore, only the completed sets from these children were included in the analysis.

Main effects models and full factorial models that included all possible two-way, three-way, and four-way interactions were tested hierarchically. We present the main effects models with no interactions to provide baseline, pooled effects of each model variable. The Verb Recall Test had a two-way interaction that did not involve the learning condition, and the Recognition Test showed a statistically significant three-way interaction, but upon further inspection, there were no pertinent simple effects. We include these particular interactions in our presentation of the results.

Effect sizes are reported as partially standardized beta coefficients (b_{std}), which are comparable to a Cohen's *d*

except they represent conditional standardized mean differences, conditioned on other variables in the model. Restricted maximum likelihood estimation was used. Bootstrapped standard errors with 1,000 replicates were used to account for nonnormal error terms. Stata Version 17.0 was used for mixed-effects model analyses (StataCorp, 2019).

Results

Verb Recall Test

As seen in Table 3, when covariates were applied, a learning condition effect was seen with scores for the RSR condition approximately 0.69 points higher (on a 0–4 scale) than scores for the RS condition (with a medium effect size, $b_{std} = 0.51$). Small effect sizes were seen with higher scores for learned items than for generalization items ($b_{std} = 0.09$), and higher scores at 5-min testing than at 1-week testing ($b_{std} = -0.09$). The numerical advantage of the TD group over the DLD group was diminished when the covariates were applied, and although still not small ($b_{std} = -0.39$), the estimate was too variable to be statistically significant. The only interaction observed was an interaction of Item Type × Time ($p = .028$), with simple effects indicating higher scores for learned items than for generalization items at 5-min testing (small effect size $b_{std} = 0.11$). There were no interactions involving learning condition, indicating that the learning condition effect (RSR > RS) applied across groups, item type, and time. An illustration of the results with scores collapsed across time is provided in Figure 1.

As described earlier, our scoring system allowed for some phonetic imprecision in the children's recall responses. For example, children were credited with correct recall even with productions such as /gin/ for /gɪn/ and /mok/ for /nok/. To determine if there was also a trace of group or learning condition effects within the recall responses regarded as correct, we did a further examination of the phonetic details of these responses. All novel verbs were CVC monosyllables whose scores on the Edwards et al. (2004) system would be 10 for a completely accurate production. We calculated for each child the mean phonetic score for these novel verb productions, separated according to learning condition and testing time. Because more novel verbs from the RSR condition were originally regarded as correctly recalled than novel verbs from the RS condition, the mean phonetic scores for the RSR condition are based on a larger number of responses. For this reason, the results should be viewed as only suggestive. For the children with DLD, phonetic accuracy was higher for novel verbs in the RSR condition at both 5 min ($M = 9.50$, $SD = 0.50$) and 1 week ($M = 9.45$, $SD = 0.63$) than for novel verbs in the RS condition (5 min: $M = 8.40$, $SD = 1.19$; 1 week; $M = 8.67$, $SD = 1.18$). For the children with TD, however, phonetic accuracy differences across learning conditions were not apparent. Phonetic accuracy for novel verbs in the RSR condition at 5 min ($M = 9.30$, $SD = 1.17$) and 1 week ($M = 9.41$, $SD = 1.15$) was no higher (and possibly even lower) than for novel verbs in the RS condition (5 min: $M = 9.57$, $SD = 0.74$; 1 week: $M = 9.71$, $SD = 0.42$). The DLD group's seemingly lower phonetic accuracy than the TD group in the RS condition stands in contrast to the two groups' similar accuracy levels in the RSR condition.

Progressive Test: Stem Recall

As noted earlier, the children's responses on the Progressive Test was examined in two ways. In this subsection, we report on the children's recall of the verbs themselves, independent of whether the children included the *–ing* inflection in their response. As seen in Table 4, the learning condition effect indicated that scores for the RSR condition were approximately 0.65 points higher (on a 0–4 scale) than scores for the RS condition (medium effect size, $b_{std} = 0.48$). The numerical advantage of the TD group over the DLD group was not statistically reliable when the covariates were applied. PPVT-4 scores had a main effect, but with a very small effect size ($b_{std} = 0.02$). No interactions were observed. Figure 2 provides an illustration of the results.

Progressive Test: Inflection Use

The previous analysis revealed how well the children could recall the novel verbs whether these verbs were inflected with *–ing* or not. However, an essential part of the task was an assessment of the children's ability to add the *–ing* inflection to the novel verb stem, given the progressive context provided by the prompt. An examination of the children's tendency to include the inflection revealed some striking differences between the two groups of children. Of the novel verbs correctly recalled, a much smaller percentage were inflected by the children with DLD than by the children in the TD group, as shown in Table 5. For example, in the RSR condition, the children with DLD correctly recalled a total of 13 novel verb stems constituting learned items. Only four of these (31%) were produced with *–ing*. In contrast, 44 novel verb stems constituting learned items were recalled by the children with TD in the RSR condition. Almost all (43 of 44, 98%) were inflected with *–ing*. For the children in the TD group, the tendency to inflect the novel verb stems was not related to whether the test items were learned items or generalization items, or whether the novel verbs were learned in the RSR condition

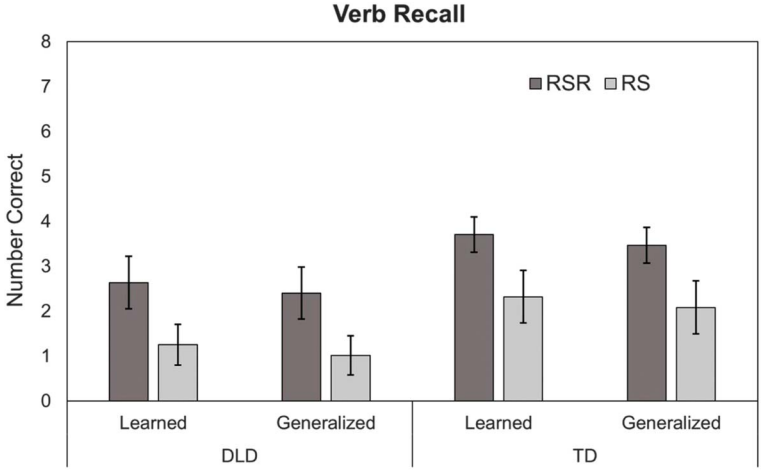
Table 3. Verb Recall Test main effects model results (n = 23, o = 356).[a]

	Main effects: no covariates					Main effects: with covariates				
Fixed effects	***b***	**95% CI**		**b_{std}**	***p* value**	***b***	**95% CI**		**b_{std}**	***p* value**
Group (DLD vs. TD)	–0.93	–1.44	–0.41	–0.68	.000	–0.53	–1.35	0.28	–0.39	.198
Condition (RSR vs. RS)	0.69	0.29	1.09	0.51	.001	0.69	0.29	1.09	0.51	.001
Item (Learn vs. Gen)	0.12	0.04	0.19	0.09	.002	0.12	0.05	0.19	0.09	.001
Time (1 wk vs. 5 min)	–0.13	–0.24	–0.01	–0.09	.030	–0.13	–0.24	–0.01	–0.09	.028
Covariates										
PPVT						0.02	0.00	0.04	0.01	.112
Mother's education						0.07	–0.07	0.22	0.05	.314
Intercept	1.30	0.80	1.80		.000	–1.91	–5.19	1.38		.256
Random effects	**σ^2**					**σ^2**				
Condition	1.12	0.57	2.20			1.01	0.51	1.97		
Intercept	0.57	0.31	1.04			0.43	0.21	0.88		
Residual	0.83	0.54	1.27			0.83	0.54	1.26		

Note. CI = confidence interval; DLD = developmental language disorder; TD = typical language development; RSR = repeated spaced retrieval condition; RS = repeated study condition; Learn = items using the same video-recorded actions as during the learning period; Gen = items using video-recorded actions with actors that had not been seen during the learning period; wk = week; PPVT = Peabody Picture Vocabulary Test.

[a]Bootstrapped standard errors.

Figure 1. The conditional means and standard errors reflecting the number of novel verb items recalled correctly on the Verb Recall Test. RSR = repeated spaced retrieval condition; RS = repeated study condition; Learned = items using the same video-recorded actions as during the learning period; Generalized = items using video-recorded actions with actors that had not been seen during the learning period; DLD = children with developmental language disorder; TD = children with typical language development.



or in the RS condition. For the children with DLD, too few novel verbs were recalled in the RS condition to draw a conclusion. More verbs were recalled in the RSR condition, and, for these verbs, *–ing* was included at much lower rates than was the case for the TD group.

The responses lacking the inflection were varied among the children with DLD. In most instances, only the novel verb in bare-stem form was produced (e.g., /nok/). In other instances, the subject and bare stem (with or without the auxiliary) were produced, as in, "Man is /nok/." There were also instances in which the children produced both the prompt and the novel verb in the way they appeared in study trials during the learning period, as in, "The man likes to /nok/."

Recognition Test

The results for the recognition test did not follow the pattern seen for the Verb Recall Test or Progressive Test: Stem Recall (see Table 6 and Figure 3). There was a main effect for group, with the children with TD recognizing approximately 1.25 more items than the children with DLD (large effect size $b_{std} = -0.86$). There were no other main effects or any two-way interactions. There was a Group × Learning Condition × Item Type interaction ($p = .029$). However, inspection of the simple effects revealed relatively little that had any bearing on the main effect for group. The only comparisons with $p < .05$ were the four comparisons showing greater recognition by the TD group than by the DLD group and the DLD group's higher scores on RSR learned items than on RS learned items ($p = .018$).

Discussion

Learning Condition Effects

Recall scores on both the Verb Recall Test and Progressive Test: Stem Recall were higher for verbs learned in the RSR condition than for verbs learned in the RS condition. These were main effects holding across group (DLD, TD), item type (learned, generalized), and time (5 min, 1 week). The RSR advantage for item type was informative because the generalization items required the children to recall the novel verb while viewing a new actor performing the action. The finding of the continuing RSR advantage 1 week after the learning period suggests that RSR benefits for verb learning are not transitory phenomena. The fact that the RSR advantage was seen across the two groups of children makes the case that RSR may be helpful to children more generally and should not be limited to remedial goals.

A more ideal result would have been gains by both groups but with the children with DLD narrowing the gap between themselves and the TD group in the RSR condition in particular. This could have been seen in a Group × Learning Condition interaction and is, in fact, what we found in a multistudy analysis of aggregated data from our novel noun and adjective studies (Leonard et al., 2021).

It is possible that the absence of such an interaction in this study is related to the size of the difference of the RSR versus RS main effect. That is, although recall was stronger for the RSR words than for the RS words, the effect size for learning condition ($b_{std} = 0.51$) was not as

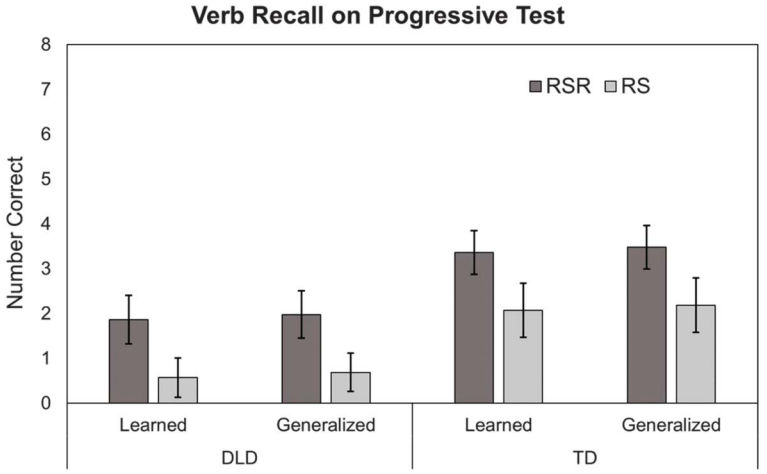
Table 4. Verb recall on progressive test main effects model results (n = 23, o = 176).[a]

	Main effects: no covariates					Main effects: with covariates				
Fixed effects	***b***	**95% CI**		**b_{std}**	***p* value**	***b***	**95% CI**		**b_{std}**	***p* value**
Group (DLD vs. TD)	–1.25	–1.78	–0.73	–0.93	.000	–0.75	–1.56	0.06	–0.56	.070
Condition (RSR vs. RS)	0.65	0.29	1.01	0.48	.000	0.65	0.28	1.01	0.48	.000
Item (Learn vs. Gen)	–0.06	–0.12	0.01	–0.04	.074	–0.06	–0.12	0.01	–0.04	.076
Covariates										
PPVT						0.02	0.00	0.04	0.02	.038
Mother's education						0.08	–0.06	0.21	0.06	.271
Intercept	1.34	0.81	1.88		.000	–2.42	–5.64	0.80		.141
Random effects	**σ^2**					**σ^2**				
Condition	0.74	0.36	1.51			0.71	0.35	1.43		
Intercept	0.50	0.05	4.73			0.44	0.20	0.97		
Residual	0.67	0.39	1.17			0.68	0.39	1.19		

Note. CI = confidence interval; DLD = developmental language disorder; TD = typical language development; RSR = repeated spaced retrieval condition; RS = repeated study condition; Learn = items using the same video-recorded actions as during the learning period; Gen = items using video-recorded actions with actors that had not been seen during the learning period; PPVT = Peabody Picture Vocabulary Test.

[a]Bootstrapped standard errors.

Figure 2. The conditional means and standard errors reflecting the number of novel verb items recalled correctly on the Progressive Test: Stem Recall regardless of whether the inflection *–ing* was included in the child's response. RSR = repeated spaced retrieval condition; RS = repeated study condition; Learned = items using the same video-recorded actions as during the learning period; Generalized = items using video-recorded actions with actors that had not been seen during the learning period; DLD = children with developmental language disorder; TD = children with typical language development.



large as in our earlier novel noun study (b_{std} = 1.09) that had the same number of words to be learned, the same number of exposures and retrieval opportunities, the same measures serving as covariates, a similar though not identical retrieval schedule for the RSR words, and highly similar CVC monosyllables as the novel words (Leonard, Karpicke, et al., 2019). The comparison between studies is not a pure one because, in this study, test items were divided into learned and generalization items, whereas, in the noun study, only learned items were used on the recall tests. However, as a possible offsetting factor, in the noun study, the children had to learn two details rather than only one in this study. In the noun study, testing included the word form, and also what the referent "liked." This study was limited to the children learning the word forms.

Approximate comparisons can also be drawn between this study on novel verbs and our earlier study on novel adjectives (Leonard, Deevy, et al., 2019). Each study used eight novel words. The adjective study was similar to, though not identical, this study in its retrieval schedule for the RSR condition, the number of exposures and retrieval opportunities, and covariates. Importantly, like this study, the adjective study used both learned items and generalization items. As in this study, the advantage of learned items over generalization items represented a small effect size (verbs: b_{std} = 0.09; adjectives: b_{std} = 0.12). Yet, the effect size for the RSR versus RS comparison was clearly larger in the adjective study (b_{std} = 0.91) than for the current verb study (b_{std} = 0.51). Comparisons with the earlier novel noun and adjective studies, then, suggest that the advantages of RSR for novel verbs may be reduced relative to the advantages seen for learning novel nouns and adjectives.

Unlike the two tests involving recall, the recognition tests showed no learning condition effect. This test was exceedingly difficult for the children with DLD, as their performance (average 3/8 correct) was not clearly above the chance level of 25%. For this reason, it may have been difficult to detect a difference in accuracy between the RSR and RS items.

Table 5. The total number and percentage of correctly recalled novel verb stems that were produced with *–ing*.

R	RSR		RS	
Group	**Learned**	**Generalization**	**Learned**	**Generalization**
DLD	4/13	8/16	0/1	0/2
	31%	50%	0%	0%
TD	43/44	42/45	29/30	29/30
	98%	93%	97%	97%

Note. RSR = repeated spaced retrieval condition; RS = repeated study condition.

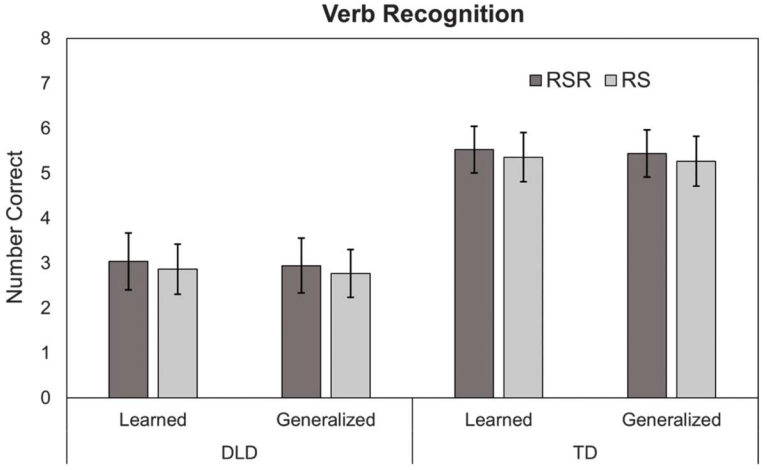
Table 6. Recognition test main effects model results (n = 23, o = 176).[a]

	Main effects: no covariates					Main effects: with covariates				
Fixed effects	***b***	**95% CI**		**b_{std}**	***p* value**	***b***	**95% CI**		**b_{std}**	***p* value**
Group (DLD vs. TD)	–1.69	–2.21	–1.17	–1.17	.000	–1.25	–2.14	–0.35	–0.86	.006
Condition (RSR vs. RS)	0.09	–0.16	0.33	0.06	.488	0.09	–0.15	0.33	0.06	.482
Item (Learn vs. Gen)	0.05	–0.18	0.27	0.03	.692	0.05	–0.19	0.28	0.03	.708
Covariates										
PPVT						0.02	0.00	0.04	0.01	.075
Mother's education						0.08	–0.08	0.23	0.05	.350
Intercept	2.85	2.40	3.31		.000	–0.60	–4.51	3.31		.762
Random effects	**σ^2**					**σ^2**				
Condition	0.15	0.00	2639.87			0.17	0.00	2177.44		
Intercept	0.49	0.22	1.07			0.44	0.21	0.96		
Residual	0.85	0.58	1.24			0.85	0.57	1.24		

Note. CI = confidence interval; DLD = developmental language disorder; TD = typical language development; RSR = repeated spaced retrieval condition; RS = repeated study condition; Learn = items using the same video-recorded actions as during the learning period; Gen = items using video-recorded actions with actors that had not been seen during the learning period; PPVT = Peabody Picture Vocabulary Test.

[a]Bootstrapped standard errors.

Figure 3. The conditional means and standard errors reflecting the number of novel verb items selected correctly on the Recognition Test. RSR = repeated spaced retrieval condition; RS = repeated study condition; Learned = items using the same video-recorded actions as during the learning period; Generalized = items using video-recorded actions with actors that had not been seen during the learning period; DLD = children with developmental language disorder; TD = children with typical language development.



However, for the TD group, the reason for the lack of a learning condition difference is unclear. These children performed above the level of chance yet were well below ceiling levels. In our earlier studies with novel nouns and adjectives, this group characteristically scored at ceiling levels, obviating any learning condition differences (We speculate below why this task might have been especially difficult in this study).

Time Effects

Recall across 1 week was not as robust as we found in the noun and adjective studies, though the effect size (b_{std} = −0.09) was considerably smaller than for the RSR versus RS comparison (b_{std} = 0.51). Notably, the decline from 5 min to 1 week was no greater for the children with DLD than for the children with TD. There was no sign that the children with DLD were more likely than their peers to forget words.

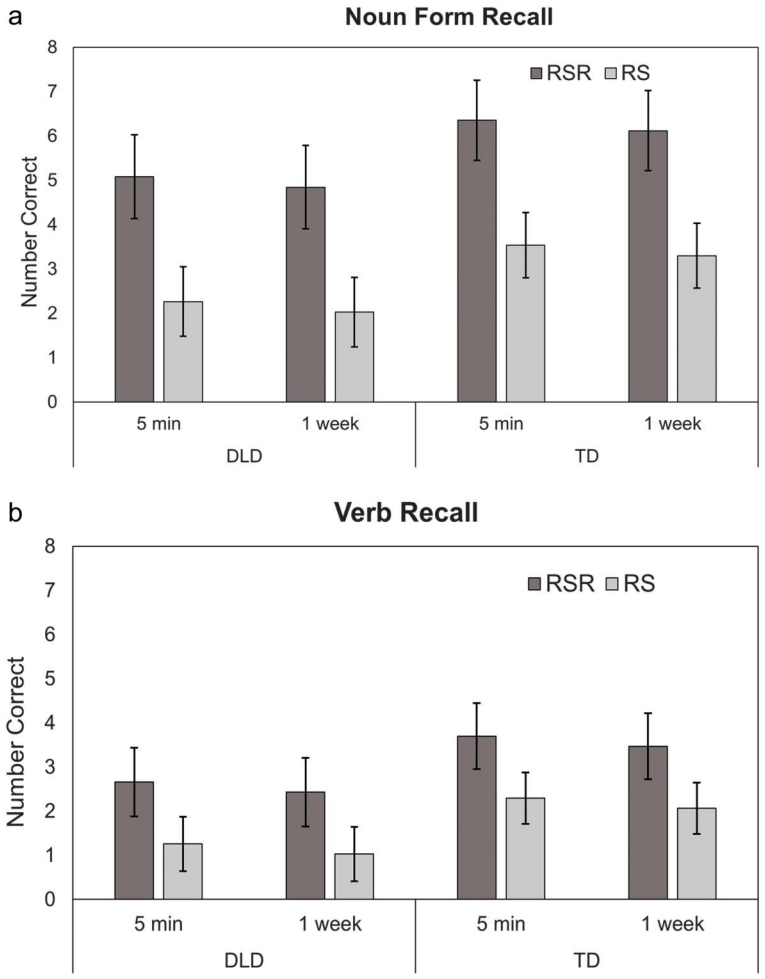
Group Effects

For the two recall tests, what appear to be differences between the DLD and TD groups were neutralized by the covariates. In our earlier studies, we found no group differences for the recall of novel adjectives and only select differences in the recall of novel nouns. In our multistudy aggregated analysis of these data (Leonard et al., 2021), there was a group effect favoring the TD group, but, as noted earlier, an interaction indicated that the difference was most attributable to group differences in the comparison condition, not the RSR condition. Importantly, the covariates played no role in the results.

The usual way we would interpret the covariate influence in this study is to say that the TD group recalled more novel verbs than the DLD group but the difference can be accounted for by the factors measured by the covariates. Because the two groups were very similar in the factor of maternal education, we suspect the more relevant factor was the PPVT-4 score. Two interpretations seem plausible. First, the two groups differed in their PPVT-4 scores to a greater extent than in our other studies. This could have allowed these scores to play a larger role in this study. More likely is the interpretation that children's word knowledge as estimated by the PPVT-4 is more closely associated with children's verb learning than of their noun or adjective learning. In this study, PPVT-4 scores were correlated with Verb Recall Test scores, r = .431, whereas the correlations in the earlier noun and adjective studies were r = .01 and r = .07, respectively.

We can only speculate on why PPVT-4 scores would be more closely associated with verb learning than with noun or adjective learning. The words to be learned and recalled in our previous studies were concrete nouns whose referents could be readily individuated and adjectives referring to visually salient attributes. At the ages of the children in our study, sets of items on the PPVT-4 include an increasing proportion of not only verbs but more abstract words including superordinate-level nouns and adjectives including participial adjectives derived from verbs (e.g., *puzzled, inflated*). Therefore, it may have been less about the PPVT-4 tapping into an ability uniquely tied to novel verb learning than about the content of the PPVT-4 going well beyond the concrete and visually salient referents that we used in our previous studies.

Figure 4. (a) The conditional means and standard errors reflecting the number of novel noun items recalled correctly on the word form recall test in the study of Leonard, Karpicke, et al. (2019). (b) The corresponding values for the Verb Recall Test in the current novel verb study. RSR = repeated spaced retrieval condition; RS = repeated study condition; 5 min = the number correct when tested 5 min after the learning period; 1 week = the number correct when tested 1 week after the learning period; DLD = children with developmental language disorder; TD = children with typical language development.



Indeed, as these words appeared in sets on the test, children with DLD were more likely to hit ceiling than children with TD.

Why Were Novel Verbs So Difficult?

Verb learning is generally more difficult than noun learning, and this applies to the learning of novel words representing verbs versus nouns. A similar interpretation might apply when we compare the results of this study with those of the novel noun study of Leonard, Karpicke, et al. (2019)—a study sharing many of the same characteristics. Eight novel words were learned in each study. Both studies provided 48 exposures of each word during study trials and 12 retrieval opportunities for the words in the RSR condition. Similar CVC monosyllabic words were used in the two studies, and the same covariates were used. Yet scores were considerably higher in the novel noun study. Figures 4a and 4b provide a visual comparison of the noun word form recall and the novel verb recall at 5 min and 1 week. The values shown in both 4a and 4b are conditional means and standard errors for the models that controlled for the covariates. As can be seen, for each group, learning condition, and time point, recall was much higher for nouns.

We must acknowledge that the differences reflected in these figures are probably affected not only by the difficulty of verbs but also by the relative ease of learning concrete nouns. Recent studies that have looked in greater detail at the types of nouns taught in vocabulary learning studies have reported that for children at the age level of our own participants, learning concrete imageable nouns

is much more successful than learning highly abstract nouns (Hadley et al., 2021). It is likely that factors such as imageability operate across form classes such as "noun" and "verb" and, if controlled, might result in a better test of the difference between noun and verb learning.

In this study, we simplified the actions and sentence frames to reduce the likelihood of confusion on the part of the children. One of two familiar actors performed a body motion referred to by an intransitive verb, and the verb was presented in a bare-stem form in a sentence-final position in three consecutive sentences of similar structure. Within the sentence structures used, the verbs appeared in positions appropriate only for verbs. Of course, in daily life, verbs are heard in a much wider variety of sentence structures, and this variety enables children to home in on the precise meaning of the verbs. However, because morphosyntax is a problem area for children with DLD, we chose to simplify the input, while making it clear that the novel word corresponded to the novel action.

However, even within these boundaries, decisions were made that might not have been ideal. For example, our study trials presented the verb in three consecutive sentences with the middle sentence employing a subject in pronoun form, as in, "This woman likes to /tɛb/. She really likes to /tɛb/. This woman can really /tɛb/." Varying from noun to pronoun forms seemed more natural than using a noun as the subject in all three sentences. However, in a recent novel verb learning study with young typically developing children, Horvath and Arunachalam (2021) found that varying between noun and pronoun subjects was not as helpful to the children as consistently using the same noun subject.

The timing between the presentation of the visual information and the audio information may also have been a factor. For the types of verbs and actions used in the study, data on timing are limited, though there is some speculation that verbs presented during the action might be preferable (Ambalu et al., 1997; see Horvath & Arunachalam, 2019, for a discussion). Given the lack of definitive data, and the fact that the novel verb appeared in three sentences in each study trial, we incorporated all three timing possibilities into these trials. As noted earlier, the first sentence was preceded by the video clip, the second occurred along with the video clip, and the third preceded the video clip. Although this balanced arrangement seemed very appropriate, it is possible that a different timing pattern might have produced better learning by the children. To this point, Horvath (in preparation) has conducted the first-ever study on verb learning and timing with children with DLD; results indicate that participants are more successful at both learning and generalizing a novel verb when they see the referent action before hearing the novel verb's phonological form.

Finally, in retrospect, our selection of actions with movements of the arms and legs may not have been ideal because they had no function or goal and, consequently, were not especially engaging. In our effort to make the actions simple, we perhaps made them too homogenous. This probably contributed to the difficulty the children had with the recognition test. The array of four video-recordings with body movements shown simultaneously could have challenged the children's attention.

Choosing similar intransitive physical actions also made the names of the actions less amenable to the process of syntactic bootstrapping. By hearing a new verb in a variety of syntactic contexts, the learner can gradually narrow the verb's meaning down to a close approximation of an adultlike interpretation of the meaning. However, our choice of actions made it difficult to select syntactic contexts that could be helpful to the children's learning. The sentence contexts we provided made it clear that the new words were words referring to the actions. However, the contribution of the sentence contexts went no further. The children had to rely almost exclusively on the topography of the actors' movements—the visual information—to learn the meaning of each novel verb.

One positive detail in the findings was the ability of the DLD group to recall the novel verbs even in the context of observing new actors performing the actions. If the children recalled the learned items for a particular novel verb, they were very likely to recall the novel verb when presented in generalization items.

We recognize that we assessed only a limited form of generalization. For example, although the actors were different in the generalization items, these items were similar to learned items in several ways. First, the plain background in the video clips was the same for both learned and generalization items. Second, the details of the new actors' actions were designed to closely resemble those of the original actors. This was not the same as, for example, testing "hop" while observing a new actor hopping on one foot rather than on two feet or testing "clap" while observing clapping with hands over the head rather than directly in front of the body. Third, although the children were asked to provide both the subject and the novel verb in the Progressive Test, the actors could be referred to as "man" and "woman" for the generalization items just as for the learned items. There was no requirement to use a new name in combination with the novel verb.

Verb Inflection Differences

The above factors might have added challenges to the DLD group's ability to learn, but they were clearly not the only problem these children had with the novel verbs. These children were surprisingly poor at adding the

progressive *–ing* inflection to those novel verbs they could recall. We chose *–ing* as the inflection to be added to the novel verbs because it is one of the earliest grammatical morphemes acquired by children with DLD and TD alike (see review in Leonard, 2014). This decision seemed to be strengthened because, as noted earlier, on a pre-experiment task requiring the children to use this inflection with familiar verbs, the children with DLD were very accurate with *–ing*. We had an additional opportunity to observe the children's use of *–ing* by including practice items on the progressive use test that required the children to produce *–ing* with real verbs. These practice items used the same prompts as for the novel verb items, and the video-recordings were in the same format. The children were successful with these items, though some children required extra practice.

As seen in Table 5, the TD group had no difficulty adding *–ing* to the novel verbs they could recall. Although the learning condition influenced the number of novel verbs the TD children recalled, it did not affect these children's ability to use the inflection with the words that were remembered.

There are several possible reasons why the children with DLD had such difficulty inflecting the novel verbs. First, unlike their TD peers, these children may have required more time and experience with new verbs before they could consistently inflect them. There is a basis for entertaining this possibility: Leonard et al. (1999) found that preschoolers with DLD were less proficient with verb-related morphemes than younger TD children even when the two groups' verb inventories were taken into account.

Another possibility relates to the longer response required for the progressive use test items. The prompt for these items was of the type, "What's she doing? Tell me about the woman. Right now the ___." The children with DLD might have found it difficult to produce the verb inflection when they also had to include the subject in their responses—a type of trade-off effect. This could apply to responses such as, "man /nok/," and "man is /nok/."

There is a third possibility that may apply to only some of the children in the DLD group. These particular children may have developed a response set during the retrieval trials in the learning period and Verb Recall Test, all of which involved the prompt of the type, "This woman likes to ____," presented along with the video-recorded novel action. With no variation in the prompt, some of the children with DLD may have begun focusing almost exclusively on the visual action, producing the novel verb more as a bare-stem gerund. Then, when the prompt changed to test for progressive *–ing* use—with no change in the video-recorded actions—the children continued to respond in the same way, with little regard to the prompt that, in this case, required the verb to be produced in progressive form. Note that focusing on the visual information would not have to affect the children's success in providing the correct label for generalization items. In fact, during the Verb Recall Test, the prompts for the learned and generalization items were the same. For other children, the sentence frame for study trials was still kept firmly in mind, leading to responses such as, "The woman likes to /tɛb/," even when the prompt had become, "Right now the ____."

There is a literature that aligns well with the observation that the children with DLD were poor at altering the form of the novel verb when the context required it. Like much younger typically developing children (e.g., Theakston et al., 2003), preschool-age children with DLD are strongly influenced by the sentence frames in which they hear new (including novel) verbs. For example, children with DLD are more likely to omit present third-person singular *–s* in a frame such as, "Every day the cat ___," if they consistently hear a novel verb in a frame such as, "Let's watch the man /krit/," or "Does the man /krit/?" than if they hear it in frames such as, "All day long, the man /krits/" or "Do you think the man /krits/?" (Leonard et al., 2015). Conversely, they are more likely to produce the inflection in an inappropriate context (e.g., completing a prompt like, "We wanna watch the man ___," with "/krits/") if they consistently hear the novel word in sentences such as, "Every day the cat /krits/."

This exaggerated influence of the input has been attributed in part to morphosyntactic comprehension weaknesses in children with DLD. For example, these children's productions such as, "Mommy open the present," could have their origins in hearing structures such as, "Let's watch Mommy open the present," or "Did Mommy open the present?" The assumption is that the children might not understand the specific constraints that the sentence contexts place on the use of the inflection, especially when a verb is new to the child. It is not too speculative, therefore, to suggest that hearing /tɛb/ and /nok/ in a single context with no inflection might lead the children to produce these novel verbs in the same way even in a context requiring an inflection.

There could be another reason why new words are not inflected properly when the context calls for it. Plante et al. (2014) found in an intervention study that children with DLD did not show increased consistency in using inflections until they had systematic exposure to the inflection attached to 24 different verbs. Although we found that the children with DLD in our study were able to use *–ing* during our pre-experiment testing, the verbs were highly familiar and had no doubt been heard frequently with *–ing*. For these children, additional experience with a wider range of verbs inflected with *–ing* might have better equipped them to inflect the novel verbs they were asked to learn in our study.

We can also turn the question around and ask: If the novel verbs had originally been presented with a variety of inflections (e.g., /noks/, /nokt/, /nokŋ/), would the children's recall of the verbs have been enhanced? In addition, would RSR continue to be facilitative? This last question seems important to pursue, for it addresses whether the advantages of RSR hold when grammatical morphology is inserted into the verb learning process. If answered in the affirmative, RSR might have dual benefits—assisting children in word learning while also providing the children with experience hearing grammatical morphemes applied to new words.

Conclusions

Given the well-documented verb learning difficulties experienced by children with DLD, we asked whether the inclusion of retrieval practice during the learning period would be of significant benefit. This appeared to be true for the children with DLD in this study as well as for a group of same-age children with TD. Recall of novel verbs proved to be greater for verbs that included RSR trials during learning than for verbs with equal exposure that provided no opportunities for retrieval. In several additional respects, the findings were encouraging. The retrieval advantage held when the children were tested 1 week later, and the advantage was still present when the novel verbs had to be recalled with new actors performing the actions.

The children with DLD did not recall as many novel verbs as their peers, though differences in the two groups' preexperiment vocabulary test scores could account for the differences in novel verb recall. In other respects, there were more similarities than differences between the two groups. For those novel verbs that were recalled, retention over 1 week was comparable, as was the ability to extend the novel verbs to new actors performing the actions.

Although retrieval practice was clearly beneficial, the medium effect sizes seen for this advantage were not as large as the effect sizes reported in similar studies of novel noun and novel adjective learning. Part of this difficulty could have a straightforward explanation—verbs are simply harder to learn, and even helpful modifications such as providing retrieval practice may have limits. However, some of the details surrounding our choice of actions, sentence frames, and video–audio timing relationships might also have been contributing factors.

Even in a study primarily of word learning such as this, the morphosyntactic weaknesses of the children with DLD became apparent. Specifically, unlike their typically developing peers, when the children with DLD were faced with contexts requiring them to add the progressive *–ing* inflection to those novel verbs they could recall, they were surprisingly poor. This was true even though the children showed evidence of using this inflection with familiar words. The advantages of retrieval practice for learning the verbs themselves did not have any impact on the children's tendency to use the inflection with the verbs recalled. We conclude that important boosts to verb learning and recall can occur through procedures such as retrieval practice. However, these benefits may apply only to basic verb-to-action mapping and 1-week recall. For children with DLD, the work of using these new verbs flexibly in different morphosyntactic contexts will have only just begun.

Data Availability Statement

The data sets used for this study are available from the corresponding author on reasonable request.

Acknowledgments

This research was supported by funding from National Institute on Deafness and Other Communication Disorders Grants R01 DC014708 awarded to Laurence Leonard, F31 DC018435 awarded to Justin Kueser, and T32 DC00030 support provided to Sabrina Horvath (Elizabeth Strickland, Project Director). We thank the children and their families for participating in this research and the research assistants in the Speech, Language, and Hearing Sciences Child Language Lab for their assistance.

References

Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *Psychological Monographs: General and Applied, 11*(1), 159–177. https://doi.org/10.1037/h0093018

Alt, M. (2011). Phonological working memory impairments in children with specific language impairment: Where does the problem lie? *Journal of Communication Disorders, 44*(2), 173–185. https://doi.org/10.1016/j.jcomdis.2010.09.003

Alt, M., Plante, E., & Creusere, M. (2004). Semantic features in fast-mapping: Performance of preschoolers with specific language impairment versus preschoolers with normal language. *Journal of Speech, Language, and Hearing Research, 47*(2), 407–420. https://doi.org/10.1044/1092-4388(2004/033)

Ambalu, D., Chiat, S., & Pring, T. (1997). When is it best to hear a verb? The effects of the timing and focus of verb models on children's learning of verbs. *Journal of Child Language, 24*(1), 25–34. https://doi.org/10.1017/S0305000996002978

Bishop, D. V. M., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology, 3*(1), 3. https://doi.org/10.1186/s40359-015-0062-7

Bishop, D. V. M., Snowling, M., Thompson, P., Greenhalgh, T., & CATALISE Consortium. (2017). CATALISE: A multinational and multidisciplinary Delphi consensus study of

problems with language development. Phase 2: Terminology. *The Journal of Child Psychology and Psychiatry, 58,* 1068–1080. https://doi.org/10.1111/jcpp.12721

Butler, A., Karpicke, J., & Roediger, H. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Chen, Y., & Liu, H.-M. (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities, 35*(1), 10–20. https://doi.org/10.1016/j.ridd.2013.10.010

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

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

Ebbels, S., Dockrell, J., & van der Lely, H. (2012). Production of change-of-state, change-of-location and alternating verbs: A comparison of children with specific language impairment and typically developing children. *Language and Cognitive Processes, 27*(9), 1312–1333. https://doi.org/10.1080/01690965.2011.605598

Edwards, J., Beckham, M., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

Fazio, L., & Marsh, E. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science, 28*(2), 111–116. https://doi.org/10.1177/0963721418806673

Fey, M., Leonard, L., Bredin-Oja, S., & Deevy, P. (2017). A clinical evaluation of the competing sources of input hypothesis. *Journal of Speech, Language, and Hearing Research, 60*(1), 104–120. https://doi.org/10.1044/2016_JSLHR-L-15-0448

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua, 92,* 333–375. https://doi.org/10.1016/0024-3841(94)90346-8

Fitch, A., Arunachalam, S., & Lieberman, A. (2021). Mapping word to world in ASL: Evidence from a human simulation paradigm. *Cognitive Science, 45*(12), e13061. https://doi.org/10.1111/cogs.13061

Fletcher, P., & Peters, J. (1984). Characterizing language impairment in children. *Language Testing, 1*(1), 33–49. https://doi.org/10.1177/026553228400100104

Frank, M., Braginsky, M., Yurovsky, D., & Marchman, V. (2021). *Variability and consistency in early language learning: The Wordbank project*. MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development, 49*(4), 988–998. https://doi.org/10.2307/1128738

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gordon, K. (2020). The advantages of retrieval-based and spaced practice: Implications for word learning in clinical and educational contexts. *Language, Speech, and Hearing Services in Schools, 51*(4), 955–965. https://doi.org/10.1044/2020_LSHSS-19i-00001

Gordon, K., McGregor, K., & Arbisi-Kelm, T. (2020). Optimising word learning in post-secondary students with developmental language disorder: The roles of retrieval difficulty and retrieval success during training. *International Journal of Speech-Language Pathology, 23*(4), 405–418. https://doi.org/10.1080/17549507.2020.1812719

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

Gray, S. (2004). Word learning by preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*(5), 1117–1132. https://doi.org/10.1044/1092-4388(2004/083)

Greenslade, K., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the Structured Photographic Expressive Language Test–Preschool: Second Edition. *Language, Speech, and Hearing Services in Schools, 40*(2), 150–160. https://doi.org/10.1044/0161-1461(2008/07-0049)

Hadley, E., Dedrick, R., Dickinson, D., Kim, E., Hirsh-Pasek, K., & Golinkoff, R. (2021). Exploring the relations between child and word characteristics and preschoolers' word-learning. *Journal of Applied Developmental Psychology, 77,* 101332. https://doi.org/10.1016/j.appdev.2021.101332

Haebig, E., Leonard, L., Deevy, P., Karpicke, J., Christ, S., Usler, E., Kueser, J., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder II: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research, 62,* 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0071

Horvath, S. (in preparation). *After is easier: How verb timing impacts referential mapping in children with DLD*.

Horvath, S., & Arunachalam, S. (2019). Optimal contexts for verb learning. *Perspectives of the ASHA Special Interest Groups, 4*(6), 1239–1249. https://doi.org/10.1044/2019_PERSP-19-00088

Horvath, S., & Arunachalam, S. (2021). Repetition versus variability in verb learning: Sometimes less is more. *Journal of Speech, Language, and Hearing Research, 64*(11), 4235–4249. https://doi.org/10.1044/2021_JSLHR-21-00091

Horvath, S., Kueser, J., Kelly, J., & Borovsky, A. (2022). Difference or delay? Syntax, semantics, and verb vocabulary development in typically developing and late-talking toddlers. *Language Learning and Development, 18*(3), 352–376. https://doi.org/10.1080/15475441.2021.1977645

Horvath, S., Rescorla, L., & Arunachalam, S. (2019). The syntactic and semantic features of two-year-olds' verb vocabularies: A comparison of typically developing children and late talkers. *Journal of Child Language, 46*(3), 409–432. https://doi.org/10.1017/S0305000918000508

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development, 76*(2), 340–355. https://doi.org/10.1111/j.1467-8624.2005.00849_a.x

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2021). Word learning and verbal working memory in children with developmental language disorder. *Autism and Developmental Language Impairments, 6,* 239694152110041–239694152110020. https://doi.org/10.1177/23969415211004109

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*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Karpicke, J. D., & Roediger, H. III (2007). Expanding retrieval practice promotes short-term retention, but equally spaced

retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704–719. https://doi.org/10.1037/0278-7393.33.4.704

Kaufman, A., & Kaufman, N. (2004). *Kaufman Assessment Battery for Children–Second Edition*. AGS.

Kelly, D. J., & Rice, M. L. (1994). Preferences for verb interpretation in children with specific language impairment. *Journal of Speech and Hearing Research, 37*(1), 182–192. https://doi.org/10.1044/jshr.3701.182

Kueser, J. B., Leonard, L. B., Deevy, P., Haebig, E., & Karpicke, J. D. (2021). Word-learning trajectories influence long-term recall in children with developmental language disorder and typical development. *Journal of Communication Disorders, 94,* 106160. https://doi.org/10.1016/j.jcomdis.2021.106160

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

Leonard, L., Christ, S., Deevy, P., Karpicke, J., Weber, C., Haebig, E., Kueser, J., Souto, S., & Krok, W. (2021). A multistudy examination of the role of repeated spaced retrieval in the word learning of children with developmental language disorder. *Journal of Neurodevelopmental Disorders, 13*(1), 20. https://doi.org/10.1186/s11689-021-09368-z

Leonard, L., & Deevy, P. (2010). Tense and aspect in sentence interpretation by children with specific language impairment. *Journal of Child Language, 37*(2), 395–418. https://doi.org/10.1017/S0305000909990018

Leonard, L., & Deevy, P. (2020). Retrieval practice and word learning in children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research, 63*(10), 3252–3262. https://doi.org/10.1044/2020_JSLHR-20-00006

Leonard, L., Deevy, P., Karpicke, J., Christ, S., & Kueser, J. (2020). After initial retrieval practice, more retrieval produces better retention than more study in the word learning of children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(8), 2763–2776. https://doi.org/10.1044/2020_JSLHR-20-00105

Leonard, L., Deevy, P., Karpicke, J., Christ, S., Weber, C., Kueser, J., & Haebig, E. (2019). Adjective learning in young typically developing children and children with developmental language disorder: A retrieval-based approach. *Journal of Speech, Language, and Hearing Research, 62*(12), 4433–4449. https://doi.org/10.1044/2019_JSLHR-L-19-0221

Leonard, L., Deevy, P., Kurtz, R., Chorev, L. V., Owen, A., Polite, E., & Finneran, D. (2007). Lexical aspect and the use of verb morphology by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 50*(3), 759–777. https://doi.org/10.1044/1092-4388(2007/053)

Leonard, L., Fey, M., Deevy, P., & Bredin-Oja, S. (2015). Input sources of third person singular -s inconsistency in children with and without specific language impairment. *Journal of Child Language, 42*(4), 786–820. https://doi.org/10.1017/S0305000914000397

Leonard, L., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., Souto, S., Kueser, J., & Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder I: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research, 62*(4), 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0070

Leonard, L., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*(3), 678–689. https://doi.org/10.1044/jslhr.4203.678

Leonard, L., Schwartz, R., Morris, B., & Chapman, K. (1981). Factors influencing early lexical acquisition: Lexical orientation and phonological composition. *Child Development, 52*(3), 882–887. https://doi.org/10.2307/1129090

Ma, W., Golinkoff, R., Hirsh-Pasek, K., McDonough, C., & Tardiff, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language, 36*(2), 405–423. https://doi.org/10.1017/S0305000908009008

Ma, X., Li, T., Duzi, K., Li, Z., Ma, X., Li, Y., & Zhou, A. (2020). Retrieval practice promotes pictorial learning in children aged six to seven years. *Psychological Reports, 123*(6), 2085–2100. https://doi.org/10.1177/0033294119856553

McGregor, K., Arbisi, T., & Eden, N. (2017). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology, 19*(1), 43–57. https://doi.org/10.3109/17549507.2016.1159337

McGregor, K., Arbisi-Kelm, T., Eden, N., & Oleson, J. (2020). The word learning profile of adults with developmental language disorder. *Autism and Developmental Language Impairments, 5,* 1–19. https://doi.org/10.1177/2396941519899311

McGregor, K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research, 60*(10), 2891–2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031

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

McGregor, K., Owen Van Horne, A., Curran, M., Wagner Cook, S., & Cole, R. (2021). The challenge of rich vocabulary instruction for children with developmental language disorder. *Language, Speech, and Hearing Services in Schools, 52*(2), 467–484. https://doi.org/10.1044/2020_LSHSS-20-00110

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*(2), 357–374. https://doi.org/10.1017/S0305000900013817

Penner, Z., Shulz, P., & Wymann, K. (2003). Learning the meaning of verbs: What distinguishes language-impaired from normally developing children? *Linguistics, 41*(2), 289–319. https://doi.org/10.1515/ling.2003.010

Piccin, T. B., & Waxman, S. R. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development, 3*(4), 295–323. https://doi.org/10.1080/15475440701377535

Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyer, C., Lieser, A. M., & Burton, R. (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology, 23*(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58*(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

Roediger, H. L. III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory, 23*(3), 403–419. https://doi.org/10.1080/09658211.2014.889710

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

StataCorp. (2019). *Stata statistical software: Release 16*. StataCorp LLC.

Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech, and Hearing Services in Schools, 48*(1), 16–30. https://doi.org/10.1044/2016_LSHSS-16-0014

Storkel, H. L. (2013). A corpus of consonant–vowel–consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods, 45*(4), 1159–1167. https://doi.org/10.3758/s13428-012-0309-7

Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2003). The role of the input in the acquisition of third person singular verbs in English. *Journal of Speech, Language, and Hearing Research, 46*(4), 863–877. https://doi.org/10.1044/1092-4388(2003/067)

Watkins, R. V., Rice, M. L., & Moltz, C. C. (1993). Verb use by language-impaired and normally developing children. *First Language, 13*(37), 133–143. https://doi.org/10.1177/014272379301303708

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