ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Predictors of Word Learning With and Without Retrieval Practice in Children With Developmental Language Disorder

Laurence B. Leonard,[a] Sharon L. Christ,[b] Jeffrey D. Karpicke,[c] Patricia Deevy,[a] Justin B. Kueser,[d] Sofía Souto,[e] and Eileen Haebig[f]

[a] Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN [b] Department of Human Development and Family Science and Department of Statistics, Purdue University, West Lafayette, IN [c] Department of Psychological Sciences, Purdue University, West Lafayette, IN [d] Boys Town National Research Hospital, NE [e] Department of Speech, Language and Hearing Sciences, Butler University, Indianapolis, IN [f] Department of Communication Sciences and Disorders, Louisiana State University, Baton Rouge

ARTICLE INFO

Article History:
Received July 25, 2025
Revision received December 6, 2025
Accepted April 17, 2026

Editor-in-Chief: Julie A. Washington
Editor: Raúl Rojas

https://doi.org/10.1044/2026_JSLHR-25-00595

ABSTRACT

Purpose: Both children with developmental language disorder (DLD) and their peers with typical language development (TD) vary in their response to word learning activities, even when a consistent procedure is used. In this study, we examine child factors that might help explain this variation in outcomes and whether any interaction between these factors operates in the same way as a function of whether retrieval practice is included in the word learning procedure.
Method: We examined the relations between a common set of child factors and children's responses in five previously published novel word learning studies with similar procedures. Participants were 4- and 5-year-old English-speaking children with DLD and their same-age peers with TD. Factors included scores on standardized language and nonverbal intelligence tests, nonword repetition, maternal education, age, and group status (TD, DLD). The role of each factor was examined while controlling for the other factors.
Results: With all other factors controlled, age proved to be predictive of children's novel word learning. Nonword repetition was also predictive, though it interacted with age. Findings indicated that for children with DLD, increases in novel word learning for older children were less likely when nonword repetition ability was low. The disadvantage of weak nonword repetition ability was also apparent in learning conditions that lacked opportunities for retrieval. When retrieval practice was included, gains with age were less dependent on nonword repetition ability.
Conclusions: For preschoolers with DLD and especially those with weak nonword repetition skills, including retrieval practice in the experimental procedure provided additional benefit. If future studies reveal that these findings apply as well to the clinical setting, the inclusion of retrieval practice during vocabulary intervention might help children avert a period of especially slow word learning.
Supplemental Material: https://doi.org/10.23641/asha.32914790

Individuals with developmental language disorder (DLD) exhibit a lifelong weakness in their ability to learn and use language (Bishop et al., 2017; Iverson & Williams, 2026; Leonard, 2014; McGregor et al., 2025; Norbury et al., 2016). In childhood, the prevalence of DLD appears to be between 7% and 8% (Norbury et al., 2016; Tomblin et al., 1997). Prevalence calculations are not available for all age groups, though by adulthood, individuals with DLD show a prevalence of between 3% and 4% (Nudel et al., 2023). The language symptoms of individuals with DLD change with age and experience, but word learning remains a consistent challenge for children, adolescents, and adults (e.g., McGregor et al., 2017, 2021; Storkel et al., 2017). For example, from childhood to adolescence,

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.*

 Copyright © 2026 The Authors
This work is licensed under a Creative Commons Attribution 4.0 International License.
Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

individuals with DLD fall further behind their peers with typical language development (TD) in their vocabulary growth (Rice & Hoffman, 2015). These individuals appear to know fewer words than their peers with TD, and the words they "know" are known in less depth (McGregor et al., 2013). These limitations in vocabulary have significance well beyond the short term; limitations in vocabulary knowledge predict later abilities such as reading skill, a well-documented weakness in many children with DLD (Adlof et al., 2017; Catts et al., 2002; Nation et al., 2022; Ziegenfusz et al., 2022).

Weaknesses in word learning are also seen in studies that make use of novel words or real but unfamiliar words to examine the children's learning ability. These studies are important because they shed light on the process of word learning in children. In contrast, standardized tests of vocabulary ability usually focus on static word knowledge—what the child has already accumulated through experience. Results from the word learning studies show that children with DLD require more exposure to the novel/unfamiliar words to reach the same learning criterion as their same-age peers (see reviews in Jackson et al., 2016; Kan & Windsor, 2010).

Although, at the group level, children with DLD acquire fewer of these words than their peers, considerable variability in learning success is seen in both groups, though especially for the children with DLD. For example, in a study by Gray (2004), preschoolers with DLD and same-age peers with TD were asked to learn four unfamiliar words to a specific accuracy criterion. The group with TD met the accuracy criterion for a mean of 3.45 words with an *SD* of 0.76, whereas the group with DLD showed a mean of 2.05 and an *SD* of 1.23. In a study asking children to learn nine novel words, Gordon et al. (2021) found that after six learning sessions, preschoolers with TD produced a mean of 4.20 novel words (*SD* = 1.10) with phonological precision with relatively little variability, while the preschoolers with DLD produced a mean of 3.25 novel words (*SD* = 2.43) with much greater variability.

Further analyses of these studies have led to the view that children with DLD struggle the most with the process of *encoding*—forming initial representations of the word that, with further experience, are refined with greater phonological precision and semantic detail. Relative to their encoding weaknesses, children with DLD appear to have better longer term retention of words (e.g., Bishop & Hsu, 2015; Calabrese et al., 2025; Gordon et al., 2021; Gray, 2004). Another distinction that has proven important is the distinction between the phonetic form of the word (hereafter, *word form*; e.g., "cherish" /ʧɛrɪʃ/) and *word meaning* (e.g., to care deeply for someone or something). Novel word studies have shown that for both children with DLD and those with TD, learning word forms (e.g., /taɪmɪk/) is more difficult than learning their arbitrarily assigned "meanings" (e.g., "likes birds;" Haebig et al., 2019). However, most studies show larger differences between these two groups (favoring the group with TD) for word forms than for meanings (e.g., Gray, 2004; McGregor et al., 2017).

What might account for this variability in word learning? Certainly, the heterogeneity of DLD is likely to be a contributing factor to variability. As McGregor et al. (2025) remind us, "No two individuals with DLD will present in exactly the same way" (p. 6072). The wide range of potential genetic sources of DLD is likely part of the heterogeneity. van Wijngaarden et al. (2024) reported as many as 45 genes that might play a role in DLD. Although each genetic variant may be insufficient to significantly alter language ability, combinations of these variants may have the cumulative effect of producing more pronounced symptoms of this disorder, with many individual differences (see Niemi et al., 2018). Children with DLD can also differ in the extent to which their language problems are accompanied by weakness in nonverbal cognitive ability (e.g., Gallinat & Spaulding, 2014), attention (e.g., Finneran et al., 2009), and motor function (e.g., Sack et al., 2022). As noted by Accardo and Shapiro (2005) in reference to neurodevelopmental disorders more generally, intervention programs associated with poor outcomes can be due to the children's co-occurring weaknesses not being taken into account. Certainly, this can apply to word learning procedures in particular. Finally, differences in degree of word learning across studies could be due to the particular procedures used and which measures (e.g., recognition, cued recall, free recall) are used to assess outcomes.

The Current Study

The field is not yet in a position to identify particular DLD subtype profiles that might predict novel word learning outcomes. However, there are existing measures that might have merit as predictors. The overall goal of the present study is to determine whether some of these predictors, singly or in interaction, might explain some of the within-group variability in novel word learning seen in 4- and 5-year-old English-speaking children with DLD and their same-age peers with TD. Our data come from five previously published studies (Leonard et al., 2023, 2024; Leonard, Deevy, et al., 2019; Leonard, Karpicke, et al., 2019; Souto et al., 2025). The five studies followed a highly similar learning procedure and the same participant selection criteria. The children learned the novel words in two learning sessions held on consecutive days, and their recall was tested 5 min after the second learning session and 1 week later. Despite these shared characteristics in design and method, in each study, there were considerable within-group differences in the children's learning and recall of the novel words, even though, within each study, all children were provided with the same number of exposures

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

(dose). This was especially true for learning the phonetic forms of novel words (e.g., /paɪb/, /wæd, /gɪs/). Across the five studies, each required the children to learn a total of eight novel word forms. For the group with DLD, the maximum number of word forms learned within a single study averaged 6.4 out of eight, while the minimum number averaged 1.2 out of eight. For the group with TD, the means for the maximum number and minimum number of word forms were eight and 2.2, respectively. Arbitrary meanings assigned to these word forms (e.g., "likes rain") were less difficult to learn for both groups and showed less variability. For this reason, we focus on the children's learning and recall of novel word forms in this investigation.

An important component of the five studies was the inclusion of a condition in which the children engaged in retrieval practice. The study of retrieval practice has a long history in the literature on memory (see Karpicke, 2025; Rowland, 2014, for a review) and has been shown to facilitate learning in both children with TD (see Fazio & Marsh, 2019) and those with DLD (see Gordon, 2020). The chief insight from this work is that the act of repeatedly attempting to recall (retrieve) information during the learning session creates new learning. Longer term retention proves better with repeated retrieval than if the same amount of exposure occurs without including retrieval attempts. In the five studies in the present investigation, half of the novel words in each study had both study trials (exposure to the novel word forms) and opportunities to retrieve these novel words during learning, whereas the other half of the novel words had the same number of study trials (exposure) but no opportunities for retrieval. For both groups in all five studies, better recall directly after learning and 1 week later was seen for the novel words in the retrieval condition than for the novel words in the study-only condition.

The beneficial effects of retrieval practice seen in these five studies might relate to the just-noted encoding weaknesses experienced by children with DLD. Retrieval practice is considered an active form of encoding (e.g., Jonsson et al., 2021; Karpicke & Grimaldi, 2012). When retrieving words, participants obtain additional encoding opportunities, and when this retrieval involves the right balance of effort with minimal forgetting, the strength of the encoding will exceed that involved in simply studying the material without any attempts at retrieval. This strengthened encoding through retrieval could be beneficial for those children who are not able to quickly encode words with the amount of exposure that is sufficient for most children. Given the expected benefits from encoding with retrieval practice, a major question in the present investigation is whether the factors that help to predict novel word form learning differ according to whether children engage in retrieval practice during the learning process. Are the predictors of word form learning outcomes different for novel words involved in retrieval practice and novel words that are studied to the same degree but without retrieval?

Some of the predictors chosen for the present study have been employed in earlier studies, though they have not produced consistent findings. These include standardized language and nonverbal intelligence test scores, nonword repetition, maternal education, and age. These are described in more detail in the Method section. One advantage of our use of five different studies with similar procedures and selection criteria is that we can determine whether some of the inconsistency seen in earlier studies might be due to slight differences in the procedure that had been followed by the previous investigators or due to the manner in which the children had been selected. In the present investigation, each of the five studies is entered separately in the analysis, permitting us to determine if any predictor operates across all studies or fails to apply in one or more of the studies included.

Discovery of factors that influence whether a child is among the better or poorer word learners has important clinical implications. For example, if factors such as low scores on a particular standardized language test or weak nonword repetition ability are associated with very slow progress in learning new word forms, clinicians should plan for additional sessions or other adjustments in dose. Caregivers should be informed about the possibility of a longer treatment period. Furthermore, if particular factors are indicative of poor learning outcomes only for word forms learned without retrieval practice, then the inclusion of retrieval practice might be especially warranted during word learning activities.

In summary, in this investigation, we pursue answers to three experimental questions:

1. Are there particular measures, singly or in interaction, that account for significant amounts of variance in the novel word form recall of 4- and 5-year-old children with DLD, as measured shortly after the learning period and 1 week later?
2. Are the predictors that help account for the novel word form learning of the children with DLD different from the predictors found for their same-age peers with TD?
3. Are the predictors different if retrieval practice is included in the learning phase?

Method

Participants

A summary of the participants in the five studies is provided in Table 1. We report data from 50 English-speaking children with DLD and 57 English-speaking

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 1. A summary of the participant characteristics in the five studies.

Study	Test/measure	Group	
1		DLD (*n* = 8)	TD (*n* = 10)
	Age in months	62.63 (6.99)	63.20 (4.89)
	Maternal education in years	15.38 (2.20)	16.90 (2.56)
	SPELT-P 2 SS	76.13 (4.91)	118.90 (7.48)
	KABC-2 or PTONI SS	107.50 (12.22)	121.60 (17.06)
	PPVT-4 SS	97.25 (10.28)	115.20 (13.21)
	Nonword repetition	0.56 (0.08)	0.77 (0.06)
2		DLD (*n* = 7)	TD (*n* = 13)
	Age in months	62.29 (6.78)	62.54 (6.70)
	Maternal education in years	15.00 (2.38)	16.69 (1.65)
	SPELT-P 2 SS	67.86 (17.96)	119.00 (8.03)
	KABC-2 SS	98.14 (11.36)	114.31 (11.06)
	PPVT-4 SS	104.14 (14.69)	118.62 (13.62)
	Nonword repetition	0.60 (0.06)	0.69 (0.11)
3		DLD (*n* = 10)	TD (*n* = 12)
	Age in months	59.60 (4.43)	60.00 (3.49)
	Maternal education in years	16.20 (2.04)	15.83 (2.12)
	SPELT-P 2 SS	74.50 (12.25)	115.33 (9.69)
	KABC-2 SS	108.70 (13.63)	112.50 (13.79)
	PPVT-4 SS	97.80 (16.78)	121.58 (10.72)
	Nonword repetition	0.66 (0.11)	0.71 (0.11)
4		DLD (*n* = 13)	TD (*n* = 13)
	Age in months	57.08 (5.38)	58.00 (6.03)
	Maternal education in years	17.08 (3.20)	17.42 (2.36)
	SPELT-P 2 SS	75.31 (9.90)	118.93 (5.29)
	KABC-2 SS	104.77 (6.97)	107.77 (10.58)
	PPVT-5 SS	97.46 (11.49)	116.69 (10.98)
	Nonword repetition	0.63 (0.11)	0.77 (0.07)
5		DLD (*n* = 12)	TD (*n* = 9)
	Age in months	60.42 (12.31)	55.67 (6.89)
	Maternal education in years	16.75 (4.16)	18.33 (2.69)
	SPELT-P 2 SS	65.75 (12.49)	119.67 (6.82)
	PTONI SS	109.50 (12.70)	123.44 (13.62)
	PPVT-4 or PPVT-5 SS	101.00 (9.96)	120.78 (12.60)
	Nonword repetition	0.53 (0.13)	0.76 (0.10)

Note. Values are means (and standard deviations). 1 = Leonard, Karpicke, et al. (2019); DLD = developmental language disorder; TD = typical language development; SPELT-P 2 = Structured Photographic Expressive Language Test–Preschool 2; SS = standard score; KABC-2 = Kaufman Assessment Battery for Children, Second Edition; PTONI = Primary Test of Nonverbal Intelligence; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; 2 = Leonard, Deevy, et al. (2019); 3 = Leonard et al. (2023); 4 = Leonard et al. (2024); 5 = Souto et al. (2025); PPVT-5 = Peabody Picture Vocabulary Test–Fifth Edition.

children with TD. Several of the children included in the original studies had participated in two studies; for the present study, we include these children's data from only one of the studies to preserve independence across studies. We included the data from the study with the smaller number of children to ensure we could include all five studies in the data analysis. None of the children showed a clear difference in how they performed in the two studies. Children in both groups fell within the age range of 4;0 (years;months) and 5;11. In the group with DLD, there were 44 children who were White, two who were Asian, one who was Black, and three who identified as having more than one race. Two of the children were Hispanic. In the group with TD, 49 children were White, three were Black, and five identified as having more than one race. Three were Hispanic. Children speaking more than one language were eligible to participate provided that English was one of the languages spoken from the outset of language development. All children in both groups passed a hearing screening and had no history of neurological damage or disease based on caregiver report. All children scored above 75 on the Nonverbal Index of

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

the Kaufman Assessment Battery for Children, Second Edition (KABC-2; Kaufman & Kaufman, 2004) or the Primary Test of Nonverbal Intelligence (PTONI; Ehrler & McGhee, 2008). Hereafter, we refer to these two tests as "NVIQ." Scores above 75 exceed the level of intellectual disability even after the standard error of measurement is taken into account. The children with DLD scored in the range of "minimal or no symptoms of ASD" on the Childhood Autism Rating Scale–Second Edition (Schopler et al., 2010). This test was not administered to the children with TD given their general age-appropriate developmental history.

The children with DLD were enrolled in a language intervention program or were scheduled to be enrolled. All 50 of these children scored below a standard score of 87 on the Structured Photographic Expressive Language Test–Preschool 2 (SPELT-P 2; Dawson et al., 2005). The cutoff of 87 has been found to show acceptable sensitivity (exploratory group = 90.6%; confirmatory group = 100%) and specificity (exploratory group = 100%; confirmatory group = 95.6%; Greenslade et al., 2009). Seven other children with DLD in the original studies met the selection criteria using a different test but scored just above 87 on the SPELT-P 2, so they were excluded from analyses here. All 57 children with TD scored above 87 on the SPELT-P 2.

Other measures obtained from the children and included as predictors were maternal education in years, age in months, standard scores on a standardized receptive vocabulary test, and the total percentage of phonemes correct on the nonword repetition measure. These measures are described in more detail in the Predictors section. To assist in the scoring of the nonword repetition task, we also administered a speech production task to each child to assess the child's use of the phonemes required for accurate repetition. This scoring procedure is also described in the Predictors section, where we provide our rationale for nonword repetition as a predictor.

General Design of the Studies

The five studies were (1) Leonard, Karpicke, et al. (2019); (2) Leonard, Deevy, et al. (2019); (3) Leonard et al. (2023); (4) Leonard et al. (2024); and (5) Souto et al. (2025). These studies will be referred to as Studies 1–5, respectively. The recruitment and experimental procedures used for these five studies were approved by the Purdue University Institutional Review Board (No. 1603017480). Written consent was obtained from the children's families, and verbal assent was provided by the children.

In all studies, a condition that included retrieval practice was compared to a condition in which the children heard the same number of presentations of the novel words but without the opportunities for retrieval. In two studies, children learned novel nouns; in two others, they learned novel verbs; and in the remaining study, novel adjectives were the focus of learning. As noted earlier, for this article, we focus only on the children's learning and recall of the novel word forms (e.g., /bog/, /jʌt/, /kudɪp/, /zogi/). The two studies involving novel nouns also presented arbitrary "meanings" of the novel words, but the recall of meanings was generally high for both conditions and both groups of children, often resulting in ceiling effects in both groups. All five studies also included a recognition task in which, upon hearing the novel word, the children had to point to the matching visual referent. This measure, too, yielded ceiling effects in both groups.

The overall design was the same for all five studies. In all studies, children learned eight novel words. Within each study, all novel words were presented an equal number of times, in the form of exposures during "study trials." Examples of the study trials in each study are shown in Table 2. Note that for each study trial, the novel word was presented three times for Studies 1, 2, 3, and 5 (e.g., /nɛp/ was heard three times in each study trial in Study 1). For Study 4, the novel word was heard two times per study trial. For half of the novel words in each study, "retrieval trials" were also included; examples of these also appear in Table 2. The number of study trials and retrieval trials differed slightly across studies but were consistent for both groups within each study. For each study, a within-participant design was used. Four- and 5-year-old English-speaking children with DLD and children with TD learned the novel words in two sets, with each set studied in two sessions held on consecutive days and tested directly after the second session and again 1 week later. In each set, half of the novel words were presented in a repeated spaced retrieval (RSR) condition and half were presented in a comparison condition involving the same number of presentations but no retrieval (the "repeated study" or RS condition). These details are shown in Table 3. The novel words in the two conditions were presented in alternating order with assignment of the novel words to a particular condition counterbalanced across the children in each group. The novel words and their referents were presented to the child in individual sessions via laptop computer. The experimenter sat next to the child and controlled the pace of the presentation. For each set, there were two learning sessions held on consecutive days. Each session was approximately 20 min in duration. In each study, the children learned eight novel words—four in the RSR condition and four in the study-only condition.

Table 4 provides a sketch of the way novel words in the two conditions appeared during learning. The sequence shown in Table 4 appeared three additional times over the two learning sessions. The same sequence and order of presentation was used in both learning sessions for Study 1 and Study 5. The other studies varied slightly in the number and

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 2. Examples of the auditorily presented study trials and retrieval trial prompts in each study.

Study	Referent type	Study trial	Retrieval trial
1	Noun	This is a /nɛp/. It's a /nɛp/. A /nɛp/ likes birds.	What's this called? What do we call this?
2	Adjective	This cat is /fɪm/. It's very /fɪm/. This cat is really /fɪm/.	Tell me about the cat. The cat is very ____.
3	Verb	This woman likes to /tɛb/. She really likes to /tɛb/. This woman can really /tɛb/.	Tell me about the woman. The woman likes to ____.
4	Verb	The woman is /pʌm/*ing* the cow. That woman is /pʌm/*ing* the cow.	What's happening here? The ____.
5	Noun	This is a /gɪs/. It's a /gɪs/. A /gɪs/ likes rocks.	What's this called? What do we call this?

Note. 1 = Leonard, Karpicke, et al. (2019); 2 = Leonard, Deevy, et al. (2019); 3 = Leonard et al. (2023); 4 = Leonard et al. (2024); 5 = Souto et al. (2025).

distribution of immediate retrieval and spaced retrieval trials. In addition, for these other studies, the order of novel words was changed (reversed) from the first learning session to the second learning session. (These slight variations across the five studies did not affect the outcome, as in all studies, more novel word forms were learned in the RSR condition than in the RS condition.)

For novel words in the RSR condition, the initial retrieval trial was an immediate retrieval trial; that is, directly after a study trial, the image of the referent was shown again and the child was asked to retrieve the novel word. Subsequent retrieval trials were spaced retrieval trials, in which there were three other novel words that intervened since the to-be-retrieved word was last heard. All retrieval trials were followed by another study trial of the same novel word. As noted earlier, novel words assigned to the RS condition involved study trials only, with the number of study trials equal to the number used in the RSR condition. To equate the two conditions for number of study trials, all the early trials in the RS condition had two study trials in succession (see Table 4).

Five minutes after the second learning session, a recall test was administered. This test was repeated 1 week later, along with a recognition test. (As noted earlier, only the word form recall responses are examined here, as they showed sufficient variability to warrant analysis.) The recall test was presented via laptop computer, controlled by the experimenter. The children's word form responses were scored as correct or incorrect using an adaptation of a system developed by Edwards et al. (2004). In this scoring system, each consonant produced is credited with 1 point each for correct place, manner, and voicing. Each vowel produced is awarded 1 point each for correct backness, height, and length. An extra point is given for correct syllable shape (e.g., consonant–vowel–consonant). For example, if the novel word /gɪn/ was produced as /dɪn/, the production would earn a score of 9 out of a possible 10, with 1 point deducted for the error in place of articulation in the initial consonant.

Table 3. Details of each study.

Study	Novel words	No. of exposures per word in study trials in each condition (RSR and RS)	No. of retrieval opportunities per word (RSR only)
1	/dɔik/, /pɑɪb/, /gɪf/, /nɛp/, /faʊn/, /jʌt/, /bog/, /wæd/	48	12
2	/fɪm/, /taɪmɪk/, /zogi/, /beɪp/, /næfi/, /mok/, /kudɪp/, /paɪt/	42	12
3	/nok/, /jæd/, /pʌm/, /tɛb/, /meɪp/, /faɪb/, /dik/, /gɪn/	32	12
4	/nok/, /jæd/, /pʌm/, /tɛb/, /meɪp/, /faɪb/, /dik/, /gɪn/	48	12
5	/bog/, /nɛp/, /paɪb/, /jʌt/, /faʊn/, /wæd/, /dɔm/, /gɪs/	48	12

Note. RSR = repeated spaced retrieval; RS = repeated study; 1 = Leonard, Karpicke, et al. (2019); 2 = Leonard, Deevy, et al. (2019); 3 = Leonard et al. (2023); 4 = Leonard et al. (2024); 5 = Souto et al. (2025).

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 4. An illustration of the spacing of retrieval trials and sequencing of the novel words during the learning period of Study 1.

Trial	Novel word	Condition	Trial type	Trial sequence
1	/fɑʊn/	RSR	Immediate retrieval	Study–retrieve–study
2	/wæd/	RS	Study only	Study–study
3	/bog/	RSR	Immediate retrieval	Study–retrieve–study
4	/jʌt/	RS	Study only	Study–study
5	/fɑʊn/	RSR	Spaced retrieval	Retrieve–study
6	/wæd/	RS	Study only	Study
7	/bog/	RSR	Spaced retrieval	Retrieve–study
8	/jʌt/	RS	Study only	Study
9	/fɑʊn/	RSR	Spaced retrieval	Retrieve–study
10	/wæd/	RS	Study only	Study
11	/bog/	RSR	Spaced retrieval	Retrieve–study
12	/jʌt/	RS	Study only	Study

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

This scoring method allowed for phonetic imprecision while significantly reducing the likelihood that the child was attempting a different novel word. First, a production was regarded as incorrect if it resembled an actual word that was a plausible name or description of the referent (e.g., "monkey," "kick," "shiny"). Second, any productions that appeared to be a potential attempt at the novel word were examined further. Productions that matched the adult pronunciation of the novel word were scored as correct. If the production deviated from adult pronunciation, it was scored according to the Edwards et al. (2004) system. We then compared this score with the score that would be awarded if we assumed the production was actually an attempt at a different novel word. In the preceding example, the production of /gɪn/ as /dɪn/ received 9 points. If, instead, /dɪn/ were actually an attempt at the novel word /dik/, for example, it would earn a score of only 6 points out of 10 due to differences between /dɪn/ and /dik/ in the vowel and final consonant. Assuming all other candidate novel words yielded scores below 9, /dɪn/ would be considered a correct attempt at the novel word /gɪn/.

Our rationale for allowing phonetic imprecision (provided there was confidence that the correct novel word was being attempted) is that we view word learning as mapping a word form to a referent and retaining the word form–referent association over time. Requiring adult-level accuracy in producing the novel word will miss the numerous instances in which a child learns and retains an approximation of the novel word and reliably applies it to the correct referent. An alternative scoring—simply using a 0%–100% phonetic accuracy continuum—does not provide a dependable way of distinguishing between an approximation of the correct novel word and a production that might not even be an attempt at the correct novel word. High interjudge scoring reliability was reported for each of the five studies. Percentage agreement for the DLD group across the five studies ranged from 93% to 100%. For the TD group, the range was 97%–100%.

Predictors

Several measures were used as predictors of the children's novel word form recall both across learning conditions and within each learning condition. These measures were among those frequently used in studies examining response-to-treatment by preschoolers with DLD (e.g., Kapa et al., 2020), including those focused on word learning (Gray, 2004; Kapa & Erikson, 2020). One predictor was the standard score on the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; L. Dunn & Dunn, 2007) or Fifth Edition (D. M. Dunn, 2019). Hereafter, we refer to these as "PPVT." This test—a measure of receptive vocabulary—has been used in previous studies, though with mixed results. For example, Gray (2004) found that the PPVT–III (L. Dunn & Dunn, 1997) was predictive of fast mapping, though, with the same task, Gray (2003, 2006) found that this same test was not predictive. The PPVT was included in the present investigation because children's static knowledge of words (as estimated by standardized tests) is likely to be affected not only by degree of previous exposure to words but also by the children's ability to make use of this exposure—their "intake" ability (Rowe, 2015). This intake ability might also be measurable within our novel word learning task. Further justification for the inclusion of the PPVT was the finding from the meta-analysis of Calabrese et al. (2025) that receptive vocabulary test scores were associated with the ability to encode new words by children with DLD.

Another language-related measure serving as a predictor is nonword repetition, specifically, the total proportion of phonemes correct on the nonword repetition task of Dollaghan and Campbell (1998). Nonword repetition is often described as a measure of working memory

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

(Gathercole & Baddeley, 1990). However, in recent years, performance on nonword repetition tasks is recognized as multiply determined, with vocabulary knowledge, perceptual skill, and speech–motor output processes making a contribution along with factors tied to memory (e.g., Gathercole, 2006). This measure, too, has been used in earlier studies with mixed success. Studies finding an association between nonword repetition and one or another spoken measure of novel word learning include Gray (2004), Alt and Plante (2006), Thordardottir and Brandeker (2013), Jackson et al. (2016), and Gordon et al. (2021). In contrast, Gray (2006), Hansson et al. (2004), Stokes et al. (2006), Thordardottir et al. (2011), and Storkel et al. (2019) did not find a relationship between the two. Both positive and negative findings have been reported for languages beyond English and for bilingual as well as monolingual children.

Given our focus in the present investigation on children's learning of novel word forms (the phonetic elements of the word) and numerous findings showing that encoding is an especially weak area in children with DLD, the use of nonword repetition ability as a predictor seems highly appropriate. A child's first encounter with a novel word form such as /kudɪp/ is comparable to the child hearing a two-syllable item on a nonword repetition test, though the latter will not be paired with a referent. Weak nonword repetition ability could suggest that when faced with a new word form, the child would need more exposures than other children to reliably associate the correct phonetic form of the word with the referent. Because our novel word learning task in each study had a fixed number of exposures, low nonword repetition ability might be associated with a child not having enough exposures to acquire the same number of novel words as other children. When scoring the nonword repetition test, we took the additional step of accounting for the children's articulation errors on real words. To do this, we administered a speech production task to each child designed to probe the phonemes used in the nonword repetition test. We eliminated from scoring any phoneme for which the child showed zero correct productions out of three opportunities. Note that because scores on the nonword repetition test took the form of the proportion of phonemes correct, when a phoneme was absent from a child's inventory, reductions were made to the denominator (the number of possible phonemes correct) as well as the numerator. This approach to scoring was different from the Edwards et al. (2004) system used to score the productions of the novel word forms described earlier. For novel word form scoring, we needed to ensure that a phonetically imprecise production was not an attempt at a different novel word. In scoring the nonword repetition test, we wanted to reduce the impact of developmental speech errors.

Standard scores on the Non-Verbal Index of the KABC-2 (Kaufman & Kaufman, 2004) or the PTONI (Ehrler & McGhee, 2008) served as another predictor. Hereafter, we refer to these two tests as "NVIQ." As noted earlier, standardized vocabulary test scores do not consistently predict novel word learning outcomes. It seemed possible that NVIQ scores could reflect, in part, some relative strength or weakness that affects the process of word learning that is not tapped by standardized tests of vocabulary. For example, NVIQ tests require attention skills and working memory abilities that could be called into play during word learning. Other predictors were chronological age in months and maternal education in years. Maternal education is sometimes correlated with greater linguistic input directed toward the child (e.g., Hoff et al., 2002). However, this measure is more likely to reflect what the child has already acquired based on this input. Its role in predicting what the child might learn in a dynamic learning situation such as novel word learning is not known and its inclusion in the present investigation might prove informative.

Finally, we also used standard scores on the SPELT-P 2 as a predictor. These scores differed from the other measures used as a predictor because they were part of our selection criteria to distinguish children with DLD from their peers with TD (see the Participants section). Therefore, the two groups differed, by design, in their scores on this test. This fact was taken into consideration in the types of analysis conducted that included this predictor. Apart from the group differences, within each group, standard scores on this test varied considerably (see Table 1). The SPELT-P 2 emphasizes expressive morphosyntax; therefore, it was unlikely to reflect the same abilities seen on the PPVT.

The Individual Studies

Study 1

Study 1 was a comparison of an RSR learning condition and an RS learning condition (Leonard, Karpicke, et al., 2019). Eight novel one-syllable words were used. These represented nouns referring to exotic plants and rare animals. There were 16 study trials for each novel word. In each study trial, the novel word was presented three times (see Table 3), for a total of 48 exposures. For words in the RSR condition, there were 12 retrieval trials (see Table 2 for an example). In this study, two kinds of information were requested—the word form ("What's this called? What do we call this?") and meaning ("And what does this one like? What does it like?").

For the RSR condition, the spacing in Study 1 followed the pattern shown in Table 4. For the words in the RSR condition, the first retrieval trial involved immediate retrieval. And subsequent retrieval trials had spacing

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

created by three other novel words intervening between the retrieval attempt and the last time the child heard the novel word in a study trial. (As noted earlier, all retrieval trials were followed directly by another study trial.)

The pattern for the novel words in the RS condition is also shown in Table 4. These words alternated with words in the RSR condition. To ensure an equal number of exposures of the words in the two conditions, two consecutive study trials were used initially for words in the RS condition (matching the two study trials in the study–retrieval–study sequence used initially for words in the RSR condition).

Five minutes after the second learning session, the recall test was administered to evaluate the children's retention of the word form and meaning of each word. The prompts and video presentations were identical to those used in the retrieval trials that occurred during the learning sessions. Each word form was tested twice; the second item for a word occurred only after all words had already been tested with one item. There were eight items (four novel words, each tested twice) for each child for each learning condition. One week later, the same recall test was administered again.

Study 2

Study 2 (Leonard, Deevy, et al., 2019), like Study 1, compared RSR and RS with immediate retrieval trials presented initially, followed by spaced trials with three intervening novel words. However, the novel words in Study 2 were novel adjectives. Pictures were drawings of common objects with unusual characteristics (e.g., a toothbrush with multiple curls instead of a straight handle). Eight novel adjectives were used, four monosyllabic (e.g., /mok/) and four disyllabic (e.g., /kudɪ]p/). To enhance the interpretation that the novel words referred to attributes (rather than as an alternative name of the object), two exemplars were used for each novel word (e.g., a toothbrush and a pencil) during the learning session. Only word form information (e.g., /kudɪp/) was targeted in the present study. Examples of study trials and retrieval trials appear in Table 2. Across the two learning sessions, all words were heard an equal number of times; the two conditions differed because the spaced retrieval condition also included 12 retrieval trials. Five minutes after the second learning session, the recall test was administered, identical in format to the retrieval trials. Each exemplar was tested once. The recall test was repeated 1 week later.

Study 3

Study 3 (Leonard et al., 2023) was designed to resemble Study 1, with the important exception that it focused on novel verb learning rather than novel noun learning. All novel verbs were monosyllables. The novel verbs referred to video recordings of actors performing novel arm and leg movements. For each novel verb, two different actors were shown performing the action during the learning period. An RSR condition with spacing of three intervening novel words was compared to an RS condition. All novel verbs were heard an equal number of times in study trials. In the RSR condition, there were 12 retrieval trials for each novel verb. Table 2 provides examples of study trials and retrieval trials. Testing occurred 5 min after the second learning session and 1 week later. Each exemplar for each novel verb was included in the recall test.

Study 4

The goal of Study 4 (Leonard et al., 2024) was to determine if novel verb learning can be facilitated if the novel verbs appeared in transitive sentences with varying subjects and objects (see Table 2). As in other studies, an RSR condition was compared to an RS condition. Unlike the other studies examined here, in Study 4, the children needed to use the novel verbs in the transitive sentences in their recall attempts. Two learning sessions held on consecutive days were used. All novel verbs were heard an equal number of times in study trials. For the RSR condition, there were 12 retrieval trials. Testing occurred 5 min after the second session and 1 week later. As in Study 3, there were recall items involving the original subjects and objects and items involving new subjects and objects. We included here only those involving the original subjects and objects, resulting in eight test items per cell for each learning condition.

Study 5

The purpose of Study 5 (Souto et al., 2025) was to determine whether the results of Study 1 (Leonard, Karpicke, et al., 2019) could be replicated if the novel words and referents were incorporated within illustrated children's books. Placing the novel words and referents within a story represented a more stringent test of the benefits of RSR. That is, because the stories themselves would make the task more engaging, more attention might be devoted to the RS novel words than in Study 1, potentially reducing the advantage held by RSR over RS. All study trials and retrieval prompts were the same as in Study 1. Their frequency of presentation was also the same. The referent objects were the same as in Study 1. However, two of the novel words, /dɔik/ and /gɪf/, were replaced by /dɔm/ and /gɪs/. Again, spacing with three intervening novel words was used. The referent objects (drawings of exotic plants and animals) did not play an integral role in the story line; rather, they were objects that were encountered along the characters' journey. As in the earlier studies, two learning sessions were held on consecutive days and recall was tested 5 min and 1 week later.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Analysis Plan

A series of linear, mixed-effects regression models were estimated to determine which of the child factors were most important for word learning. The word learning outcome is a sum of the number of novel words learned out of eight novel words and is approximately normally distributed. Six child predictors were evaluated in the models: child age in months, mother's education in years, PPVT standard score, SPELT-P 2 standard score, NVIQ standard score, and the nonword repetition task proportion correct. SPELT-P 2 standard scores were different in the DLD and TD groups as these scores were part of the selection criteria. However, within each group, these scores varied considerably so including them as a predictor was potentially informative. Child variables were centered at zero.

Study differences were controlled for using dummy codes (Study 5 referent) as fixed-effects controls. This was an important analysis feature because the five studies differed somewhat in the number of novel word exposures provided and in the type of target word used (noun, verbs, adjectives). If word learning outcomes were different due to these characteristics, these differences would be captured by these fixed effects. Experimental design variables were included in the models as controls and tested as moderators of the child effects. The experimental variables included condition (RSR vs. RS), group (DLD vs. TD), and time (1 week vs. 5 min). A random intercept measuring between-children variability and random slopes for study condition and time were included in all models. Random effects indicated significant nesting (Intraclass Correlation Coefficient = .45).

First, each child factor was evaluated in a model with study and design controls but no other child factors. Next, all six child factors were included in a model with controls. A backward selection method was used to arrive at an optimal model. This procedure involved removing the child variable with the highest p value (lowest statistical significance) first; rerunning the model; and reevaluating statistical significance, model fit criteria (e.g., variance explained using r square), and information criteria (Akaike information criterion [AIC] and Bayesian information criterion [BIC]). Smaller values of AIC and BIC indicate a better model. (Supplemental Material S1) Nonlinear effects for each child factor were evaluated, but none were relevant. Next, each possible two-way interaction between the child factors was tested in models with all variables (child factors plus controls), and then significant interactions were tested together in models. The optimal model with two-way child factor interactions was determined through nested model testing. Finally, using the best model based on the informative two-way child factor interactions, three-way interactions were tested where the experimental design variables were treated as moderators of the two-way interactions. Important interactions are presented graphically using low (−1 *SD* below the average), average, and high (+1 *SD* above the average) values of the moderator. Simple effect estimates are also provided at these levels. Stata SE v.18 (StataCorp., 2025) was used for all statistical analyses.

Results

Predictors of Novel Word Form Learning Operating Singly and in Interaction

Single child factor ("bivariate") models indicated that, after controlling for study, condition, group, and time, only child age and the nonword repetition task were associated with the number of novel word forms learned (see Table 5). The nonword repetition model was the best in terms of AIC and BIC, but the age model was best in terms of total variance explained. Each 6 months increase in age was associated with an increase of 0.52 (= 6 × 0.086, p = .003) in novel words learned, and each 10% increase in correct phonemes on the nonword repetition test was associated with an increase of 0.36 (= 0.10 × 3.558, p = .019) in novel words learned.

The multiple child factor models (all models in Supplemental Material S2) using backward selection indicated that only child age was statistically significant using α level .05 when including other child covariates. Supplemental Material S2 provides results for the series of main effects models examined in the backward selection process. While not all child factors were statistically significant, the models that included child age, nonword repetition, NVIQ, and mother's education (Models 1–3) revealed that these factors, in combination, explain more variance in the outcome at the trial and child levels and that removing these variables did not improve AIC and BIC, substantially indicating that Model 3 (see Supplemental Material S2) was a relatively good main effects model. Note that the PPVT and SPELT-P 2 variables overlapped and were mostly irrelevant in models controlling experimental group (DLD vs. TD) even though only the SPELT-P 2 score was used to assign the children to groups. (Without the group variable, both the PPVT and the SPELT-P 2 were important factors.)

Three 2-way, child factor interactions were statistically significant when evaluated individually in models with all the other child factors and controls. These interactions included child age by NVIQ, child age by nonword repetition, and mother's education by nonword repetition. A model with all three of these two-way interactions included simultaneously along with the child factors and controls resulted in a nonsignificant two-way interaction

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 5. Single child variable models ("bivariate") controlling for study and all experimental variables (subjects = 107, trials = 428).

Variable	*b*	*p* value	AIC	BIC	Total R^{2a}
Age (months)	0.086	.003	1804.1	1856.9	.495
Maternal education (in years)	0.099	.151	1809.2	1862.0	.440
PPVT	–0.005	.748	1814.3	1867.0	.449
SPELT-P 2	–0.008	.661	1813.7	1866.5	.436
NVIQ	–0.007	.595	1814.2	1866.9	.456
Nonword repetition	3.558	.019	1799.6	1852.4	.468

Note. Most variance in novel word learning is explained by the experimental factors. AIC = Akaike information criterion; BIC = Bayesian information criterion; PPVT = Peabody Picture Vocabulary Test–Fourth Edition or Peabody Picture Vocabulary Test–Fifth Edition; SPELT-P 2 = Structured Photographic Expressive Language Test–Preschool 2; NVIQ = Kaufman Assessment Battery for Children, Second Edition or Primary Test of Nonverbal Intelligence.

[a]Proportion of variance explained by all model variables.

for mother's education by nonword repetition. Nested testing was done comparing this model with all three interactions to the models with each possible pair of interactions (see Supplemental Material S3), including the child factors found to be important in the backward selection. The best model included the two interactions involving child age (see Supplemental Material S3, Model 2), which removed the Mother's Education × Nonword Repetition interaction. The complete results of this model are provided in Table 6. The simple effects for this model (see Table 7) indicate that age is positively associated with word form learning at low (−1 *SD*) and average levels of NVIQ, but not for children with high (+1 *SD*) NVIQ scores. That is, when NVIQ is high, recall scores do not increase appreciably with age unlike the case for average and low NVIQ. As NVIQ scores go down, the importance of age goes up. Conversely, the age effect increases as nonword repetition proportions go up. Age is not important at low (−1 *SD*) nonword repetition but is by average levels of nonword repetition. That is, for children with low nonword repetition ability, recall scores do not get higher with increasing age.

Table 6. Results from the best model (Model 2) with child factor two-way interactions (subjects = 107, trials = 428).

Variable	*b*	*p* value	95% CI	
Child factors				
Age (months)	0.07	.024	0.01	0.12
Mother's education (years)	0.12	.090	–0.02	0.26
NVIQ	0.00	.832	–0.03	0.04
Nonword rep (proportion)	2.48	.135	–0.77	5.74
Age × NVIQ	–0.01	.002	–0.01	0.00
Age × Nonword Rep	0.53	.014	0.11	0.95
Study and experiment controls				
Study 1	–0.63	.297	–1.81	0.55
Study 2	0.45	.415	–0.63	1.53
Study 3	0.07	.903	–1.08	1.22
Study 4	–0.80	.152	–1.90	0.30
Condition (RSR vs. RS)	1.28	.000	0.90	1.66
Group (DLD vs. TD)	–0.72	.000	–1.05	–0.38
Time (1 wk vs. 5 min)	–0.82	.062	–1.69	0.04
Intercept	3.82	.000	2.92	4.72
	σ^2		95% CI	
Random effects				
Condition	2.85		1.91	4.26
Time	2.07		1.34	3.20
Intercept	2.81		1.93	4.09
Residual	1.02		0.73	1.42

Note. CI = confidence interval; NVIQ = Kaufman Assessment Battery for Children, Second Edition or Primary Test of Nonverbal Intelligence; rep = repetition; RSR = repeated spaced retrieval; RS = repeated study; DLD = developmental language disorder; TD = typical language development; wk = week.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 7. Simple effects for the Age × NVIQ and the Age × Nonword Repetition interactions from the best interaction model (subjects = 107, trials = 428).

Variable	*b*	95% CI		*p* value
Age at low NVIQ	0.17	0.08	0.26	.000
Age at average NVIQ	0.07	0.01	0.12	.024
Age at high NVIQ	–0.03	–0.12	0.05	.428
NVIQ at low age	0.05	–0.01	0.10	.076
NVIQ at average age	0.00	–0.03	0.04	.832
NVIQ at high age	–0.04	–0.07	–0.01	.008
Age at low nonword rep	0.00	–0.08	0.08	.977
Age at average nonword rep	0.07	0.01	0.12	.024
Age at high nonword rep	0.13	0.05	0.21	.001
Nonword rep at low age	–0.74	–5.20	3.71	.744
Nonword rep at average age	2.48	–0.77	5.74	.135
Nonword rep at high age	5.71	1.89	9.53	.003

Note. High and low values are at +1 *SD* and −1 *SD*, respectively. NVIQ = Kaufman Assessment Battery for Children, Second Edition or Primary Test of Nonverbal Intelligence; CI = confidence interval; rep = repetition.

Table 6 also reveals an important finding regarding the learning condition. In each of the five studies, we had found that the RSR condition was associated with greater word form recall than the RS condition. Table 6 shows that this effect holds even after all other factors are taken into consideration, including age, maternal education, NVIQ, nonword repetition ability, and their interactions.

Interaction Effects Differ Across Groups (TD, DLD) and Learning Condition (RSR, RS)

Three-way interactions were then explored to test for moderation by experimental factors. Results reveal that the Age × Nonword Repetition interaction effects differed across groups ($p = .001$), learning conditions ($p = .017$), and time ($p < .001$). The Age × Nonword Repetition interaction is displayed separately for the TD and DLD groups in Figure 1. Figure 1 shows that for the TD group, age is consistently positively associated with word form learning (average effect is 0.11 additional words for a 1-month age increase), while, for the DLD group, the association of age with word form learning is different across levels of nonword repetition. There is a strong effect of age at high levels of nonword repetition ($b = 0.35$, $p = .001$), indicating an increase in 0.35 words learned for a month increase in age. At average levels of nonword repetition, the age effect is important but less strong ($b = 0.15$, $p = .012$). Age is not associated with word learning when a child is −1 *SD* below the mean on nonword repetition. Table 8 provides these simple effects.

The Age × Nonword Repetition interaction is also influenced by learning condition. Figure 2 displays the Age × Nonword Repetition interaction separately for the RSR and RS learning conditions. The positive association of age with word learning is apparent in the RSR condition across all levels of nonword repetition. The average age effect is 0.10 in this condition. The association of age with word learning is different across levels of nonword repetition in the RS condition where the association is not present at low levels of nonword repetition ($b = -0.02$, $p = .692$), moderate for children with average nonword repetition ($b = 0.06$, $p = .035$), and stronger for children who are +1 *SD* on nonword repetition ($b = 0.14$, $p < .001$). These simple effects are provided in Table 9. It is important to point out that this difference in how age and nonword repetition interact in the two learning conditions occurred even though the same children participated in both conditions. It is also noteworthy, as illustrated in Figure 2, that children at each age level and each level of nonword word repetition ability achieved higher recall scores when they were in the RSR condition than when they were in the RS condition. This observation indicates that the three-way interaction of this model was not reflecting an RSR advantage that was dependent on particular ages and/or nonword repetition abilities but rather that the relationship between age and nonword repetition ability was not the same in the two learning conditions.

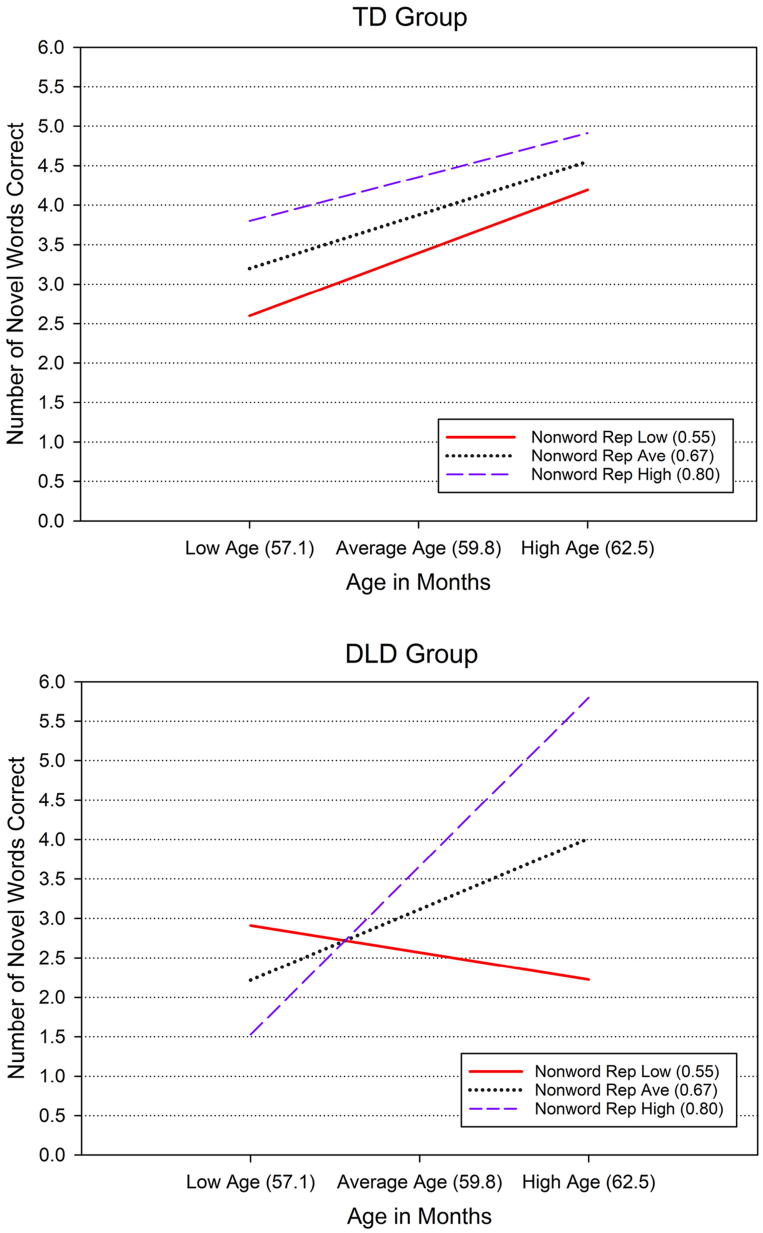
Finally, the Age × Nonword Repetition interaction was also influenced by time where the consistent positive association between age and word learning was present at 1 week post intervention regardless of nonword repetition ability (average effect: $b = 0.10$, $p < .05$). At 5 min after the intervention, the association of age with word learning differed across child levels of nonword repetition where the association was stronger at higher nonword repetition levels. Table 10 provides these results.

Discussion

The goal of our investigation was to answer three questions: Were there factors that helped to predict degree

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 1. Age × Nonword Repetition interaction for the typically developing (TD) group and the developmental language disorder (DLD) group. Rep = repetition; Ave = average.



of novel word form learning in children with DLD? Were these factors the same as in children with TD? Do the factors serving as predictors change if retrieval practice is incorporated into the learning sessions? Each question received an affirmative answer, though the answers emerged through interactions among the factors. Three-way interactions involving age and nonword repetition scores served as a central element in the findings. We begin with a discussion of these main findings and then turn to other findings that arose from the analysis.

Age and Nonword Repetition Interacted Differently in Children With DLD and Children With TD

A key to understanding how novel word form learning outcomes differed between children with DLD and children with TD is to consider how age and nonword repetition interacted in the two groups. For the children with TD, there was a slight advantage in recall for higher versus lower nonword repetition ability at each age. However, increases

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 8. Simple effects for the Age × Nonword Repetition × Group interaction (subjects = 107, trials = 428).

Variable	*b*	95% CI		*p* value
TD group				
Age at low nonword rep	0.13	0.01	0.25	.040
Age at average nonword rep	0.11	0.03	0.19	.005
Age at high nonword rep	0.09	0.00	0.18	.041
Nonword rep at low age	4.83	–1.64	11.30	.143
Nonword rep at average age	3.87	–0.54	8.28	.086
Nonword rep at high age	2.91	–1.93	7.74	.239
DLD group				
Age at low nonword rep	–0.06	–0.14	0.03	.216
Age at average nonword rep	0.15	0.03	0.26	.012
Age at high nonword rep	0.35	0.14	0.56	.001
Nonword rep at low age	–5.57	–11.63	0.48	.071
Nonword rep at average age	4.41	–0.60	9.42	.084
Nonword rep at high age	14.39	5.95	22.83	.001

Note. High and low values are at +1 *SD* and –1 *SD*, respectively. CI = confidence interval; TD = typical language development; rep = repetition; DLD = developmental language disorder.

in recall with age did not appear to depend on nonword repetition ability (see Figure 1).

For the children with DLD, age and nonword repetition ability interacted in a different way. Increases in recall with age were seen only for the children with average or high nonword repetition ability. Children with DLD with low nonword repetition ability did not differ in recall at the different age levels (see Figure 1). At the oldest ages, recall scores were much higher when nonword repetition ability was high.

Although higher ability with nonword repetition augurs well for greater success with novel word form learning, it is doubtful that a remediation procedure aimed directly at improving nonword repetition ability is a reasonable direction for intervention. Instead, for children with low ability in nonword repetition, an increase in the number of word learning sessions and/or other dose-related adjustments might be built into the intervention program. It is well-documented that children with DLD require more exposure to new words than their same-age peers with TD (e.g., Alt et al., 2004; Gray et al., 2014). It may prove to be the case that these are the children with DLD with the poorest nonword repetition ability.

Age and Nonword Repetition Interacted Differently Depending on Whether Retrieval Practice Was Provided

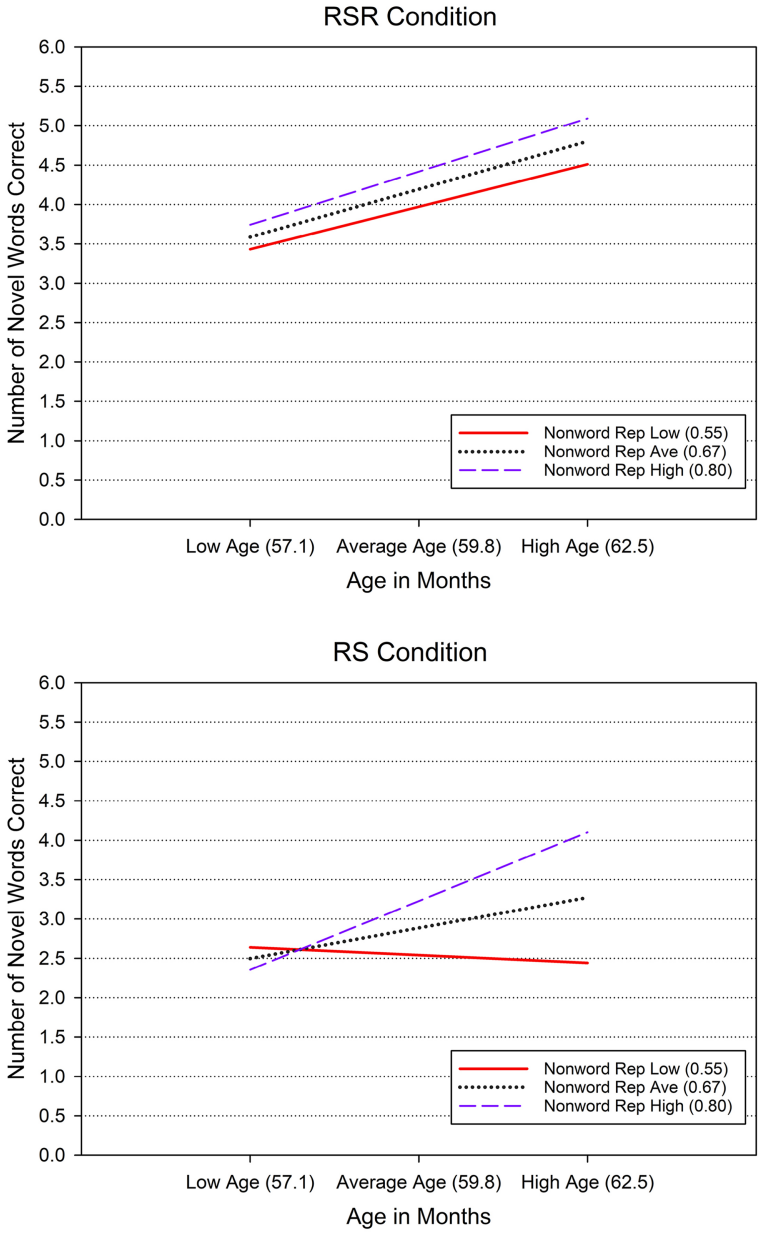
In each of the five studies providing data for the present investigation, the inclusion of retrieval practice (the RSR condition) was associated with more accurate novel word form learning than the same number of novel word exposures without opportunities for retrieval (the RS condition). These studies employed within-child designs with the same children participating in both conditions. Given the differences between the RSR and RS conditions in word learning outcomes, we asked if there were factors that played larger or smaller roles in these outcomes depending on the condition.

At the interaction level, this proved true. The three-way interaction showed that the age by nonword repetition relationship operated in different ways in the two conditions. For the RS condition (involving novel word exposure only), nonword repetition ability did not appear relevant at the youngest ages (see Figure 2). However, at higher age levels, those with nonword repetition ability at average or + 1 *SD* levels showed greater novel word form learning. In addition, at the oldest age level, novel word form recall was very clearly superior for children with the higher nonword repetition ability. These results suggest that a child's level of nonword repetition ability can have a large impact on gains in novel word form learning with age during more passive learning experiences as in the RS condition.

The nature of the Age × Nonword Repetition interaction was different for novel words learned with the addition of retrieval practice (the RSR condition). In this condition, nonword repetition ability appeared to be less crucial. Gains could be seen with increasing age to a marginally significant degree even by children with the lowest nonword repetition ability (p = .087; see Table 9 and Figure 2). When nonword repetition ability was average or higher, these changes with age were especially clear. Nevertheless, unlike the case for the RS condition, there were no significant differences in recall scores as a function of nonword repetition ability even at the oldest age level.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 2. Age × Nonword Repetition interaction for the repeated spaced retrieval (RSR) condition and the repeated study (RS) condition. Rep = repetition; Ave = average.



It should be noted that although the relationship between age and nonword repetition ability was different in the RSR and RS conditions, this three-way interaction had no bearing on the overall novel word recall advantage that RSR had over RS. As noted earlier, and as is clear from the *y*-axis in Figure 2, at each age level and level of nonword repetition ability, recall was higher for the RSR condition compared to the same children's recall in the RS condition. The most influential element in the interaction was that low nonword repetition ability was associated with limited to no improvement with age but only in the RS condition.

Though tempting, it is probably an exaggeration to conclude from these findings that retrieval practice serves a protective function against weak nonword repetition ability. However, it does seem reasonable to conclude that weak nonword repetition ability could have less adverse effects when retrieval practice is provided.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 9. Simple effects for the Age × Nonword Repetition × Learning Condition interaction (subjects = 107, trials = 428).

Variable	*b*	95% CI		*p* value
RS condition				
Age at low nonword rep	–0.02	–0.10	0.06	.692
Age at average nonword rep	0.06	0.00	0.12	.035
Age at high nonword rep	0.14	0.06	0.22	.000
Nonword rep at low age	–1.15	–5.66	3.36	.617
Nonword rep at average age	2.77	–0.55	6.09	.102
Nonword rep at high age	6.69	2.74	10.64	.001
RSR condition				
Age at low nonword rep	0.09	–0.01	0.19	.087
Age at average nonword rep	0.10	0.03	0.17	.009
Age at high nonword rep	0.11	0.01	0.21	.032
Nonword rep at low age	1.26	–4.37	6.88	.662
Nonword rep at average age	1.80	–2.19	5.80	.376
Nonword rep at high age	2.35	–2.44	7.14	.336

Note. High and low values are at +1 *SD* and –1 *SD*, respectively. CI = confidence interval; RS = repeated study; rep = repetition; RSR = repeated spaced retrieval.

How might retrieval practice serve to reduce the relative importance of nonword repetition ability? One possibility relates to the assumption that retrieval practice enhances the child's encoding of new words (e.g., Jonsson et al., 2021; Karpicke & Grimaldi, 2012). With each retrieval, the representation of the word form is strengthened. When this additional encoding experience is not available, as in the RS condition, the children's word form learning will depend more on their more basic ability to retain new word forms (e.g., /kudɪp/, /wæd/)—an ability that probably overlaps considerably with the ability measured in nonword repetition.

We have highlighted the possible importance of children's nonword repetition ability, especially when the word learning activity does not include retrieval practice. Yet, the operative factor in the role played by nonword repetition may go beyond nonword repetition itself. This measure is often thought of as just one measure among others that tap into working memory—an idea that can be traced back at least as far as Gathercole and Baddeley (1990). In recent years, more comprehensive measures of working memory have proven informative in studies of word learning. For example, Gray et al. (2022) found that a broader battery of working memory measures that included nonword repetition accounted for large amounts of variance in the phonological word learning of 7- to 9-year-olds with TD. In fact, when scores for the working memory measures were entered into analysis before the children's scores on the Expressive Vocabulary Test–Second Edition (EVT-2) and KABC-2, the latter could

Table 10. Simple effects for the Age × Nonword Repetition × Time interaction (subjects = 107, trials = 428).

Variable	*b*	95% CI		*p* value
5 min				
Age at low nonword rep	–0.03	–0.11	0.05	.528
Age at average nonword rep	0.06	0.00	0.12	.046
Age at high nonword rep	0.15	0.07	0.23	.000
Nonword rep at low age	–1.66	–6.18	2.85	.470
Nonword rep at average age	2.57	–0.74	5.88	.128
Nonword rep at high age	6.80	2.88	10.73	.001
1 week				
Age at low nonword rep	0.11	0.01	0.20	.026
Age at average nonword rep	0.10	0.03	0.17	.004
Age at high nonword rep	0.10	0.00	0.19	.047
Nonword rep at low age	2.98	–2.34	8.30	.272
Nonword rep at average age	2.68	–1.12	6.47	.167
Nonword rep at high age	2.37	–2.17	6.91	.307

Note. CI = confidence interval; rep = repetition.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

account for only 0.1% of the additional variance in phonological word learning. It could be that our interpretation about the importance of nonword repetition could apply to an even greater extent if other working memory measures were included.

Yet, even if working memory can explain some of the variance in novel word learning, it is not likely to have a bearing on the robust learning advantage that RSR has over RS. As a case in point, Fordyce et al. (2025) divided young adults into different levels based on their performance on working memory tasks. They found that, even though better working memory was associated with better recall, repeated retrieval produced greater recall than RS at all levels of working memory.

Other Findings

Our initial analysis examined the role of each potential predictor separately using as controls only the experimental variables of group (DLD, TD), learning condition (RSR, RS), and time (5 min, 1 week). Only age and nonword repetition proved significant. Standard scores on the SPELT-P 2 played no role, though this was likely due, in part, to group being included as a control, which could have neutralized the fact that the two groups differed in SPELT-P 2 scores by design. (These scores were relevant when group was removed from the analysis as a control.) Nevertheless, the finding for SPELT-P 2 scores was somewhat surprising because, within each group, there was considerable variation in SPELT-P 2 scores. For example, for the children with DLD, the average range of standard scores on the SPELT-P 2 across the five studies was 34.6 (smallest range = 17). The TD group also showed wide standard score ranges on the SPELT-P 2. For these children, the average range was 21.4 (smallest range = 15). Therefore, there was ample opportunity for this test to be a significant predictor. Yet, when the group factor and other factors were considered, SPELT-P 2 scores provided no unique insight into the children's novel word learning, suggesting that the kind of language abilities tapped on this test (primarily morphosyntactic) are not vitally tied to the kind of abilities needed to learn newly encountered word forms.

PPVT scores, like SPELT-P 2 scores, did not serve as a predictor. Despite not being part of the selection criteria, the PPVT proved to be colinear with SPELT-P 2 scores, which likely contributed to our findings when group was included as a variable. Yet, within each group, standard score ranges on the PPVT varied considerably. For the children with DLD, the average range of standard scores was 36 (smallest range = 28); for the TD group, the average was 38.6 (smallest range = 34).

Two characteristics of the PPVT suggest reasons for the lack of a close connection to novel word learning. The first is that it measures word knowledge accumulated over time rather than a dynamic measure reflecting the process of new learning. Second, it measures recognition of words and their meanings as assessed through pointing at the correct referent picture, whereas our novel word task required the children to retrieve and produce novel word forms. A measure requiring production such as the EVT-2 (Williams, 2007) might have shown a tighter connection to novel word outcomes in this study. However, because the EVT-2 involves the retrieval of accumulated words whose word forms have had time to stabilize in the children's lexicon, this test may capture only a piece of the type of ability involved in our novel word learning task. Adding to the plausibility of this idea is the Calabrese et al. (2025) finding that expressive vocabulary test scores (including the EVT) were not associated with the encoding of new words by children with DLD.

A subsequent analysis examined the role of each potential predictor when the other predictors served as a control in addition to the experimental variables. Age proved to be the dominant factor, though its interaction with other factors, especially nonword repetition ability, provided a clearer sense of its role. Other potential predictors were not significant in this analysis, though when NVIQ and maternal education were added to age and nonword repetition, more variance in learning outcomes could be explained. Maternal education also added some predictive ability in the interaction model (see Table 10). This measure is often taken to provide a rough estimate of the amount of input provided to the child, though it is recognized that the degree of input does not translate directly to the children's ability to make use of this input. Other factors beyond input frequency are clearly at work. Even so, linear trends have been reported between maternal education and measures of children's spontaneous speech such as number of different words and total number of words produced in addition to measures of receptive vocabulary (Dollaghan et al., 1999). We acknowledge, however, that in our studies, the role of maternal education may have been constrained because the education range of the children's mothers was not wide—the great majority in each group had at least some college experience.

Although all children in each group scored above the intellectual disability level, NVIQ did play a role in interaction with age. Specifically, there were greater increases in word form recall with age for children whose NVIQ levels were −1 *SD* or at average age levels. For children with NVIQ levels of +1 *SD*, changes with age were not significant. It is difficult to understand just how nonverbal intelligence played a role in this interaction.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

The NVIQ levels of the children showed differences between the TD and DLD groups that were in line with the broader literature. For example, in the meta-analysis of Gallinat and Spaulding (2014), children's standard scores on NVIQ tests were on average 10 points lower than their TD counterparts. Of our five studies reported in this article, the average gap was larger than 10 points in three studies but smaller than 10 points in the other two (see Table 1). Based on the studies in the Gallinat and Spaulding meta-analysis that used the KABC-2, standard scores for the DLD group were somewhat lower than we found in this investigation, but then, the standard scores for the TD group were also lower than we observed. If future work shows that the role played by NVIQ that we report here was not a spurious finding, research into the particular components of NVIQ might be examined. For example, all nonverbal tests of intelligence seem to involve more than one type of cognitive ability. Lancaster et al. (2025) completed a study of several nonverbal intelligence tests and found that each involves some combination of attention, statistical learning, receptive language, and working memory ability. One might speculate, then, that in our study, the Age × NVIQ interaction reflected something akin to the following: When NVIQ was higher, children at younger ages had additional types of cognitive abilities to support word learning. When NVIQ was lower, only children at older ages had acquired some of these cognitive abilities, leading to the increases in word learning skill we observed in this group. However, we currently have no basis to assume that this kind of developmental pattern is underlying the interaction observed.

Why Age?

Along with its role in important interactions, age was a main effect even when all other potential predictors and experimental variables served as controls. Why was age such a central player in the analyses? Provided they met our other selection criteria, children could range in age from 4;0 to 5:11, though of the 107 children, only four were younger than age 4;2 and in no study was the age range wider than 20 months. This age range is not out of line with other studies of children with DLD (e.g., Gray, 2003; Kapa, 2025).

Because age was a predictor across five different studies, there may be some important development that occurs within this particular age range—one that is apparent even when standard scores measuring expressive morphosyntax and receptive vocabulary are taken into account. It is possible that age was reflecting some language-related ability not assessed by our standardized language tests, such as knowledge of English phonotactics. We might also speculate that age, separate from language ability as measured in tests, might have been reflecting abilities related to particular types of executive functioning—areas of documented weakness at least in children with DLD (see meta-analyses in Niu et al., 2024; Pauls & Archibald, 2016). Some of these studies examined executive function abilities as predictors of learning. For example, Kapa and Erikson (2020) reported that a measure of inhibition was predictive of the novel word learning performance of preschoolers with DLD, including their ability to produce the novel words when presented with depictions of the referent. Notably, in the Kapa and Erikson (2020) study, executive function measures explained significant variance even after controlling for the KABC-2 and PPVT-4 scores of the children.

It will be important to discover which developing abilities are being reflected in this age range. Uncovering these abilities will help us better understand how age contributes to the greater importance of nonword repetition ability in children with DLD than in children with TD and in novel word form learning with no retrieval practice than in learning when retrieval practice is included.

Some Provisos

Our conclusions should be tempered by several important design features embedded in this investigation. First, our examination of the role played by each predictor took into account each of the other predictors. We do not argue, for example, that all or even most standardized language tests are unhelpful in estimating children's novel word learning. Rather, we noted that, when all other factors are taken into account—including the grouping (TD, DLD) factor—age and nonword repetition could explain much of the variance that might otherwise be attributed to the particular language tests that we employed. Other standardized language tests might hit closer to the heart of the ability that is underlying novel word learning and may explain unique variance even when age and nonword repetition are taken into account.

Second, when we examined the role of age in our data, the different age levels were made up of different children. This was not a longitudinal study of the same children tested at different points across time. It is possible that subtle differences between the children representing the different ages affected our results despite our use of selection criteria designed to keep each group relatively homogenous at each age level. (This potential limitation does not apply to the Age × Nonword interaction differences between the RSR and RS conditions, as the same children at each age level participated in both conditions.)

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Finally, there were features built into the original design of the five studies that could have contributed to the results but were not isolated in a way to determine their impact. For example, study trials directly followed retrieval trials in the RSR condition and we do not know if the children made use of this input. In addition, early in the learning period, novel words in the RS condition had two consecutive study trials and we do not know if the second of these was helpful or was disregarded as uninformative. These issues have more to do with the reasons for the overall learning advantages of the RSR condition over the RS condition in the five studies. However, it is plausible that they contributed to the predictors functioning somewhat differently in the two learning conditions.

Conclusions

With these points in mind, we believe our findings suggest that, (a) when children with DLD have relatively high nonword repetition ability, increases in novel word form learning seem especially likely with increasing age; (b) success with novel word form learning seems more closely related to nonword repetition ability when children must learn novel words without the opportunity for retrieval; and (c) low nonword repetition ability may be less of an impediment to gains in novel word form learning when retrieval practice is provided. Do these conclusions provide any guidance for treatment aimed at bolstering children's acquisition of new word forms? Our word learning procedure was probably more structured than most procedures that would be employed clinically with preschoolers, though, across studies, a variety of word types were included—nouns, verbs, and adjectives. In addition, Study 5 incorporated shared book reading, albeit via computer presentation. The growing appearance of children's e-books notwithstanding, similar research with physical books is needed. Further testing of alternative spacing schedules and materials is also warranted. If through these additional steps our results continue to apply, the contribution of this investigation becomes clear: Along with the overall learning benefits offered by RSR, for preschoolers with DLD and especially those with weak nonword repetition skills, including retrieval practice in the treatment plan might help the children avert a period of especially slow word learning.

Data Availability Statement

The data sets used for this study are available upon request to the corresponding author.

Acknowledgments

This research was supported by funding from National Institute on Deafness and Other Communication Disorders Grant R01 DC014708 (awarded to Laurence B. Leonard). The authors thank the children and their families for participating in this research and the research assistants in the Child Language Lab in Speech, Language, and Hearing Sciences for their assistance.

References

Accardo, J., & Shapiro, B. K. (2005). Neurodevelopmental disabilities: Beyond the diagnosis. *Seminars in Pediatric Neurology, 12*(4), 242–249. https://doi.org/10.1016/j.spen.2005.12.006

Adlof, S. M., Scoggins, J., Brazendale, A., Babb, S., & Petscher, Y. (2017). Identifying children at risk for language impairment or dyslexia with group-administered measures. *Journal of Speech, Language, and Hearing Research, 60*(12), 3507–3522. https://doi.org/10.1044/2017_JSLHR-L-16-0473

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

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)

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,* Article 3. https://doi.org/10.1186/s40359-015-0062-7

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Calabrese, P., Hedger, N., Pritchard, K., Stojanovik, V., & Pagnamenta, E. (2025). Word learning in children with developmental language disorder: A meta-analysis testing the encoding hypothesis. *Journal of Memory and Language, 145,* Article 104678. https://doi.org/10.1016/j.jml.2025.104678

Catts, H. W., Fey, M. C., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142–1157. https://doi.org/10.1044/1092-4388(2002/093)

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

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M.

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

(1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research, 42*(6), 1432–1443. https://doi.org/10.1044/jslhr.4206.1432

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test–Fifth Edition*. NCS Pearson.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Third Edition*. AGS. https://doi.org/10.1037/t15145-000

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. AGS/Pearson. https://doi.org/10.1037/t15144-000

Edwards, J., Beckham, M. E., & 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)

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

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

Finneran, D. A., Francis, A. L., & Leonard, L. B. (2009). Sustained attention in children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 52*(4), 915–929. https://doi.org/10.1044/1092-4388(2009/07-0053)

Fordyce, A. L., Redick, T. S., Bedwell, J. P., & Karpicke, J. D. (2025). Individual differences in working memory and the benefit of retrieval practice. *Journal of Memory and Language, 144,* Article 104664. https://doi.org/10.1016/j.jml.2025.104664

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 57*(4), 1363–1382. https://doi.org/10.1044/2014_JSLHR-L-12-0363

Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513–543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336–360. https://doi.org/10.1016/0749-596X(90)90004-J

Gordon, K. R. (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. R., Storkel, H. L., Lowry, S. L., & Ohlmann, N. B. (2021). Word learning by preschool-age children with developmental language disorder: Impaired encoding and robust consolidation during slow mapping. *Journal of Speech, Language, and Hearing Research, 64*(11), 4250–4270. https://doi.org/10.1044/2021_JSLHR-21-00046

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

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

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

Gray, S. I., Levy, R., Alt, M., Hogan, T. P., & Cowan, N. (2022). Working memory predicts new word learning over and above existing vocabulary and nonverbal IQ. *Journal of Speech, Language, and Hearing Research, 65*(3), 1044–1069. https://doi.org/10.1044/2021_JSLHR-21-00397

Gray, S., Pittman, A., & Weinhold, J. (2014). Effect of phonotactic probability and neighborhood density on word-learning configuration by preschoolers with typical development and specific language impairment. *Journal of Speech, Language, and Hearing Research, 57*(3), 1011–1025. https://doi.org/10.1044/2014_JSLHR-L-12-0282

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

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with development language disorder II: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research, 62*(4), 944–964. https://doi.org/10.1044/2018_JSLHR-L-18-0071

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

Hoff, E., Laursen, B., & Tardiff, T. (2002). Socioeconomic status and parenting. In M. Bornstein (Ed.), *Handbook of parenting: Ecology and biology of parenting* (Vol. 2, pp. 161–188). Erlbaum.

Iverson, J. M., & Williams, D. L. (2026). Annual research review: Developmental language disorder—A hidden condition with lifelong impact. *The Journal of Child Psychology and Psychiatry, 67*(4), 546–565. https://doi.org/10.1111/jcpp.70067

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

Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology, 113*(5), 972–985. https://doi.org/10.1037/edu0000627

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)

Kapa, L. L. (2025). Does executive function performance predict treatment response in preschoolers with developmental language disorder? *American Journal of Speech-Language Pathology, 34*(4), 2040–2052. https://doi.org/10.1044/2025_AJSLP-24-00396

Kapa, L. L., & Erikson, J. A. (2020). The relationship between word learning and executive function in preschoolers with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(7), 2293–2307. https://doi.org/10.1044/2020_JSLHR-19-00342

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Kapa, L. L., Meyers-Denman, C., Plante, E., & Doubleday, K. (2020). Predictors of treatment response for preschool children with developmental language disorder. *American Journal of Speech-Language Pathology, 29*(4), 2082–2096. https://doi.org/10.1044/2020_AJSLP-19-00198

Karpicke, J. D. (2025). Retrieval-based learning. In J. Wixted (Ed.), *Learning and memory: A comprehensive reference* (3rd ed., Vol. 4, pp. 411–441). Elsevier. https://doi.org/10.1016/B978-0-443-15754-7.00074-2

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24*(3), 401–418. https://doi.org/10.1007/s10648-012-9202-2

Kaufman, A., & Kaufman, N. (2004). *Kaufman Assessment Battery for Children, Second Edition*. AGS. https://doi.org/10.1037/t27677-000

Lancaster, H. S., Smolak, E., Milne, A., Gordon, K. R., Emerson, S. N., & Selin, C. (2025). Analyzing the impact of four cognitive constructs on nonverbal intelligence test performance: Implications for children with neurodevelopmental disorders. *Language, Speech, and Hearing Services in Schools, 56*(3), 834–846. https://doi.org/10.1044/2025_LSHSS-24-00056

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

Leonard, L. B., Deevy, P., Christ, S. L., Karpicke, J. D., Kueser, J. B., & Fischer, K. (2024). Learning verbs in sentences: Children with developmental language disorder and the role of retrieval practice. *Journal of Speech, Language, and Hearing Research, 67*(11), 4446–4465. https://doi.org/10.1044/2024_JSLHR-24-00321

Leonard, L. B., Deevy, P., Horvath, S., Christ, S. L., Karpicke, J., & Kueser, J. B. (2023). Can retrieval practice facilitate verb learning in children with developmental language disorder and their peers with typical language development? *Journal of Speech, Language, and Hearing Research, 66*(4), 1309–1333. https://doi.org/10.1044/2022_JSLHR-22-00509

Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S., Weber, C., Kueser, J. B., & 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. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., Souto, S., Kueser, J. B., & 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

McGregor, K., Arbisi-Kelm, 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. K., Goffman, L., Plante, E., & Werfel, K. (2025). Developmental language disorder as a multidimensional neurodevelopmental spectrum: Implications for diagnosis. *Journal of Speech, Language, and Hearing Research, 68*(12), 6069–6088. https://doi.org/10.1044/2025_JSLHR-25-00089

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

McGregor, K. K., 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

Nation, K., Dawson, N. J., & Hsiao, Y. (2022). Book language and its implications for children's language, literacy, and development. *Current Directions in Psychological Science, 31*(4), 375–380. https://doi.org/10.1177/09637214221103264

Niemi, M. E. K., Martin, H. C., Rice, D. L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E. J., Yu, S., Gecz, J., Martin, N. G., Wright, C. F., Fitzpatrick, D. R., Firth, H. V., Hurles, M. E., & Barrett, J. C. (2018). Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature, 562,* 268–271. https://doi.org/10.1038/s41586-018-0566-4

Niu, T., Wang, S., Ma, J., Zeng, X., & Xue, R. (2024). Executive functions in children with developmental language disorder: A systematic review and meta-analysis. *Frontiers in Neuroscience, 18*. https://doi.org/10.3389/fnins.2024.1390987

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Nudel, R., Christiansen, R. V., Kalnak, N., Schwin, M., Banasik, K., Dinh, K. M., DBDS Genomic Consortium, Erikstrup, C., Pedersen, O. B., Burgdorf, K. S., Ullum, H., Ostrowski, S. R., Hansen, T. F., & Werge, T. (2023). Developmental language disorder—A comprehensive study of more than 46,000 individuals. *Psychiatry Research, 323,* Article 115171. https://doi.org/10.1016/j.psychres.2023.115171

Pauls, L. J., & Archibald, L. M. D. (2016). Executive functions in children with specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 59*(5), 1074–1086. https://doi.org/10.1044/2016_JSLHR-L-15-0174

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

Rowe, M. L. (2015). Input versus intake—A commentary on Ambridge, Kidd, Rowland, and Theakston's 'The ubiquity of frequency effects in first language acquisition'. *Journal of Child Language, 42*(2), 301–305. https://doi.org/10.1017/S030500091400066X

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

Sack, L., Dollaghan, C., & Goffman, L. (2022). Contributions of early motor deficits in predicting language outcomes among preschoolers with developmental language disorder. *International Journal of Speech-Language Pathology, 24*(4), 362–374. https://doi.org/10.1080/17549507.2021.1998629

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

Souto, S., Leonard, L. B., Deevy, P., Christ, S. L., Karpicke, J. D., & Schroeder, M. L. (2025). Word learning in children with developmental language disorder: The use of retrieval practice during shared book reading. *Journal of Speech,*

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Language, and Hearing Research, 68(7), 3305–3321. https://doi.org/10.1044/2025_JSLHR-24-00809

StataCorp. (2025). *Stata Statistical Software Release 18*. StataCorp LLC. https://www.stata.com/

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

Storkel, H. L., Komesidou, R., Pezold, M. J., Pitt, A. R., Fleming, K. K., & Romine, R. S. (2019). The impact of dose and dose frequency on word learning by kindergarten children with developmental language disorder during interactive book reading. *Language, Speech, and Hearing Services in Schools, 50*(4), 518–539. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0131

Storkel, H., Voelmle, K., Fierro, V., Flake, K., Fleming, 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

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders, 46*(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research, 54*(2), 580–597. https://doi.org/10.1044/1092-4388(2010/09-0196)

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

van Wijngaarden, V., de Wilde, H., van der Molen, D. M., **Petter, J., Stegeman, I., Gerrits, E., Smit, A. L., & van den Boogaard, M.-J.** (2024). Genetic outcomes in children with developmental language disorder: A systematic review. *Frontiers in Pediatrics, 12*. https://doi.org/10.3389/fped.2024.1315229

Williams, K. T. (2007). *Expressive Language Test–Second Edition*. NCS Pearson. https://doi.org/10.1037/t15094-000

Ziegenfusz, S., Paynter, J., Flückiger, B., & Westerveld, M. F. (2022). A systematic review of the academic achievement of primary and secondary school-aged students with developmental language disorder. *Autism & Developmental Language Impairments, 7*. https://doi.org/10.1177/23969415221099397

Downloaded from: https://pubs.asha.org Purdue University - Library on 10/30/2026, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions