ASHA American Speech-Language-Hearing Association

LSHSS

Research Article

Interactive Book Reading to Accelerate Word Learning by Kindergarten Children With Developmental Language Disorder: Incorporating Retrieval-Based Practice Into Training

Katherine R. Gordon,[a] Danielle Moss,[b] Rebecca E. Swinburne Romine,[c] Kandace K. Fleming,[c] and Holly L. Storkel[d]

[a] Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital, Omaha, NE [b] Department of Speech-Language-Hearing: Sciences & Disorders, The University of Kansas, Lawrence [c] Life Span Institute, The University of Kansas, Lawrence [d] National Institute on Deafness and Other Communication Disorders, Bethesda, MD

ARTICLE INFO

Article History:
Received November 11, 2024
Revision received February 6, 2025
Accepted May 17, 2025

Editor-in-Chief: Karla N. Washington

https://doi.org/10.1044/2025_LSHSS-24-00147

ABSTRACT

Purpose: Laboratory-based word learning research demonstrates that retrieval-based practice supports learning and retention more than passive training strategies. The goal of the current study was to determine whether incorporating retrieval-based practice into an interactive book reading intervention contributed to better learning and retention for children with developmental language disorder (DLD) than a more passive version.
Method: Kindergarten children with DLD (N = 37) completed the intervention administered twice a week for a total of 15 sessions. The between-participants conditions varied in the proportion of retrieval opportunities (high, mid, low). Learning of word forms and meanings were assessed throughout training, immediately after training, and after 4-, 8-, and 12-week delays.
Results: The proportion of retrieval opportunities did not relate to learning or retention. Children who started the intervention with more language knowledge and skills produced more forms and meanings correctly during and after treatment. Few factors related to the learning and forgetting rates. Maternal education related positively to learning rate of forms. Children who produced more definitions during training and females, as opposed to males, demonstrated a shallower forgetting rate of definitions.
Conclusions: Both the more passive version and the retrieval-based version of the intervention were effective to support word learning in children with DLD. Overall, pretreatment characteristics did not provide a good indication of children's learning rate or posttraining forgetting rate. Thus, it is likely that frequent assessments during training are the best indicator of intervention effectiveness for a given child. Additionally, retention should be periodically assessed post-intervention to monitor forgetting.
Supplemental Material: https://doi.org/10.23641/asha.29696552

Learning vocabulary is a pervasive and persistent challenge for individuals with developmental language disorder (DLD; Jackson et al., 2019; Kan & Windsor, 2010). Despite this challenge, few effective treatments exist, especially for kindergarten-age children with DLD (Cirrin & Gillam, 2008). Language skills and knowledge at kindergarten entry predict success in various academic areas, including language, mathematics, and social skills (Pace et al., 2019). Additionally, vocabulary knowledge is strongly

Correspondence to Katherine R. Gordon: Katherine.Gordon@boystown.org. ***Disclosure:*** *This article was prepared while Holly Storkel was employed at The University of Kansas. The other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

related to children's ability to learn to read and to understand what they read (Catts et al., 2002; Kaiser & Roberts, 2011). To support school success for children with DLD, it is critical that we develop effective vocabulary interventions for kindergarten-age children when they are building essential academic skills (e.g., reading) and foundational academic knowledge (e.g., science, mathematics).

The long-term goal of the current project is to develop an effective interactive book reading intervention for kindergarten-age children with DLD. Interactive book reading is effective for children with typical development (TD; Marulis & Neuman, 2010; Mol et al., 2008, 2009). The specific components of interactive book reading to teach unfamiliar words vary (see Towson et al., 2021, for a review). However, the core components include an adult reading a storybook with a child and, at times, deviating from the text to provide explicit instruction or to engage the child in a dialogue to support the learning goals. Interactive book reading can be advantageous to teach vocabulary because words, and their meanings, are not taught in isolation. Instead, words are presented within a meaningful context, a key principle of rich vocabulary instruction (Beck et al., 2013).

As a first step to developing an effective interactive book reading intervention for children with DLD, Storkel and colleagues identified a version of the intervention that demonstrated high rigor and effectiveness. They identified Justice et al.'s (2005) clinical trial for kindergarten children from low-income households as a good candidate for adaptation for children with DLD. This version includes elaboration, in which a definition and a supportive context sentence are provided for each target word. Storkel included these elements and also included a synonym for each target word to increase elaboration. Notably, compared to children with TD, children with DLD require two to three times the number of exposures to learn words (Gray, 2003; Riches et al., 2005). Thus, Storkel and colleagues systematically varied the dosage across children with DLD to identify an adequate intensity to support word learning (Storkel et al., 2017). They determined that 36 exposures was the adequate intensity for the intervention to be effective. In a subsequent study, Storkel et al. (2019) determined that children's learning did not vary based on dose per session or dose frequency when they received a total dosage of 36 exposures per word. Given these findings, in the current study, all children were provided with 36 exposures to all target words.

Intervention Effectiveness: Posttraining Retention

Despite favorable learning during the interactive book reading intervention, children with DLD demonstrated difficulties with posttraining retention (Storkel et al., 2019). Developing a long-term memory of a word includes several stages (Davis & Gaskell, 2009). Word learning involves learning the word form (i.e., the phonemes of the word and their order), the word meaning, and the link between the two. During the first session with an unfamiliar word, the learner can encode initial representations of the word's form, meaning, and link via fast mapping (Gupta & Tisdale, 2009). Multiple training sessions are important to develop a phonologically precise representation of the word's form and a rich understanding of the word's meaning that can be retained long term (Beck et al., 2013; Gordon, Storkel, et al., 2021; McGregor et al., 2007; Munro et al., 2012). This process is known as slow mapping.

Our current theories of word learning are incomplete because we lack an understanding of how words are learned and retained over long timescales. Intervention effectiveness is most often determined by assessing knowledge immediately after the training is concluded (Marulis & Neuman, 2010). When retention of word learning is assessed, it is typically assessed between 24 hr to a week after training. However, it is common for learners to forget some or all of the trained information across extended posttraining delays (Cepeda et al., 2008; Ebbinghaus, 1885; Gordon et al., 2016). Current research indicates that training strategies can differentially affect how much material is retained long term (Booth, 2009; Vlach & Sandhofer, 2012). The goal of the current study is to modify the interactive book reading approach to support posttraining retention of taught words.

Retrieval-Based Practice

Long-standing research indicates that engaging in retrieval-based practice (i.e., learning through testing) supports better posttraining retention than more passive training strategies (see Eisenkraemer et al., 2013; Roediger & Butler, 2011, for reviews). During both passive instruction and retrieval-based practice, the learner is initially presented with the target information. For example, the learner could be presented with an object and a laboratory-created word form (e.g., This is a blicket.). During passive training strategies, the learner is then presented with the target information across subsequent training blocks. In contrast, during retrieval-based practice, the learner is asked to actively retrieve the information during subsequent training blocks. For example, the learner could be asked to label a previously presented object (e.g., What is this one called?). Retrieval-based practice can be effective even when the learner does not successfully retrieve the target information (Butterfield & Metcalfe, 2006). When feedback is given, the learner becomes aware of

what they do and do not know providing an opportunity for correction and elaboration of the memory (Lipowski et al., 2014).

Retrieval-based practice can be particularly effective when it is paired with spaced practice (see Gordon, 2020, for a review). Spacing can occur within a session as the learner cycles through target items, or across sessions, as there are gaps in time between sessions. Spaced practice introduces a desirable difficulty in that the learner must attempt to retrieve an item that is not readily accessible in working memory. Overall, as the learner attempts to retrieve target items across spaced training blocks within a session and/or across sessions, they slowly build the retrieval route, which supports successful retrieval after a posttraining delay (Karpicke et al., 2014).

Most retrieval-based practice research has been conducted with college-age adults with TD. However, retrieval-based practice is effective for children with TD (Fazio & Marsh, 2019), adults with DLD (Gordon, McGregor, & Arbisi-Kelm, 2021; McGregor et al., 2020), and children with DLD (see Gordon, 2020). Most relevant to the current study, Leonard and colleagues demonstrated that retrieval-based practice supports word learning and posttraining retention for 4- to 6-year-old children with DLD across a series of studies (see Leonard & Deevy, 2020, for a review). In these studies, unfamiliar objects were verbally labeled with laboratory-created forms. During retrieval-based practice, they asked children to name objects. Of note, most research with retrieval-based practice, including work by Leonard, includes one or two training sessions and one session to assess posttraining retention.

To effectively incorporate retrieval-based practice in word learning interventions, it is important to conduct research using real words with meanings that vary in their level of abstractness. Currently, most studies on word learning and retrieval-based practice with children include concrete objects. Notably, concrete meaning information is primarily learned through sensorimotor experiences, but abstract meaning information is primarily learned through verbal input (Muraki et al., 2022). When taught via verbal input, children must use their current semantic and syntactic knowledge to learn word meanings. Because language processing and language knowledge are areas of marked weakness for children with DLD, learning words with abstract meanings is particularly difficult for them (McGregor et al., 2012). However, learning words with abstract meanings is essential to support academic content learning and to build reading comprehension skills (Beck et al., 2013; N. Dawson et al., 2021).

There is preliminary evidence that retrieval-based practice can be effective to teach words with abstract meanings to children with DLD in an intervention format. In Gordon et al. (2024), we taught a subset of the words included in the current study to seven children with DLD across 15 training sessions. We assessed their retention after 2- and 8-week posttraining delays. In Gordon et al. (2024), children were not trained with books. Instead, words were presented in researcher-created mini stories with pictures that were several sentences long. Similar to Storkel et al. (2017, 2019), children were presented with child-friendly definitions and synonyms to support word learning. Notably, in Gordon et al. (2024), the majority of time during each training session was dedicated to retrieval-based practice in that children were asked to retrieve forms and meanings with feedback across multiple training blocks. In contrast, past versions of Storkel's interactive book reading intervention included minimal retrieval opportunities (Storkel et al., 2017, 2019). Children's knowledge of word forms and meanings were assessed periodically during training, but not during every session. Overall, Gordon et al. (2024) provides preliminary evidence that retrieval-based practice can be effective for children with DLD in that six of the seven children demonstrated a response to the intervention.

The Current Study

Previous versions of the interactive book reading intervention were effective for children with DLD in that children demonstrated learning of targeted words (Storkel et al., 2017, 2019). However, Storkel et al. (2019) noted significant posttraining forgetting. In previous versions of Storkel's intervention, children were read books and provided with extratextual information about target words. Specifically, the administrator provided synonyms, definitions, and context sentences for the target words in the books. This intervention included spaced practice within sessions in that the instructor cycled through target words during pre–book reading, book reading, and post–book reading activities. Additionally, each book was read during six sessions, which introduced across-session spacing. However, the intervention was largely passive in that the child's primary job was to listen. The goal of the current study is to assess the effectiveness of incorporating retrieval-based practice into the intervention to support both learning and retention.

When designing a retrieval-based training approach, a critical design element is how many times a learner should be asked to retrieve target information. Current evidence with adult learners indicates that as the number of retrieval opportunities increases, the probability of posttraining retention also increases (Karpicke & Roediger, 2008; Lipowski et al., 2014; Roediger & Karpicke, 2006). As previously noted, each retrieval attempt is a learning opportunity.

The primary goal of the current study was to determine the within-session proportion of retrieval opportunities to optimally support learning and posttraining retention of words. Thus, we varied the proportion of retrieval opportunities relative to passive exposures across three between-participant conditions (high, mid, and low). Of note, children were given feedback to all retrieval prompts, whether they provided a correct or incorrect response. Thus, in all conditions, children received 36 exposures to each word form. They also received an equal number of exposures to synonyms, definitions, and context sentences across conditions. Because we were interested in both learning and retention rates, we increased the probes of learning during training and posttraining retention relative to past iterations of the intervention. Children's knowledge of words was probed during every reading of every book during treatment. Additionally, we included assessments immediately after training as well as 4, 8, and 12 weeks posttreatment. Overall, the current study provides a more thorough assessment of both word learning rates and retention rates by children with DLD than past studies.

There were several possible outcomes to varying the proportion of within-session retrieval opportunities for kindergarten-age children with DLD. Given the past research with adult learners, more retrieval opportunities might be optimal. However, word learning is difficult for children with DLD. Thus, repeated probing on words may be overly discouraging, and a mid or low proportion of testing may be optimal. Another possibility is that children will not demonstrate differences in learning and retention based on the proportion of within-session testing. Specifically, probing children's knowledge of words during every training session (whether it is high, mid, or low) may similarly support learning and retention. Regardless of the outcome, the current methodology should provide useful information for clinicians when incorporating retrieval-based practice into interactive book reading interventions.

In past versions of the intervention, Storkel et al. (2017, 2019) demonstrated variability across children with DLD in the number of words learned. Thus, it is important to determine which child-level factors are related to learning and retention in the current intervention. Standardized language assessments are often included in word learning intervention studies to determine how prior language knowledge affects children's response to the intervention (Loftus et al., 2010; Nash & Snowling, 2006; Zipoli et al., 2011). The current study is unique in that, in addition to language assessments, we administered a battery of standardized tests to assess children's memory skills including working memory, episodic learning (i.e., how children build representations across presentations), and decay (i.e., retention). This study is also unique because we assessed how individual factors differentially relate to learning during training and retention posttraining as these are different processes.

In summary, in the current study, we implemented a modified version of the interactive book reading intervention that incorporated retrieval-based practice during treatment. Our primary questions were as follows:

1. How does the proportion of within-session testing during treatment (low, mid, high) affect the number of word forms and meanings learned during the intervention and retained post-intervention?
2. Which pretreatment characteristics of children with DLD are associated with learning during the intervention and retention post-intervention?

Method[1]

Participants

Thirty-seven kindergarten children with DLD (M_{age} = 5;9 [years;months], SD = 4.46 months, range: 5;0–6;6) completed the intervention. Eight additional children were excluded because their treatment schedule was disrupted during the 2020 pandemic, and three additional children were excluded due to scheduling issues. All protocols were approved by the institutional review board at The University of Kansas. Parents and/or guardians provided written informed consent for their child to participate. Participants were recruited through language screenings (86.49%), the project study's website (8.11%), and referrals from speech-language pathologists (5.40%). Characteristics of our sample can be found in Table 1 and Supplemental Material S1. All included children were at least 5 years old and eligible for kindergarten. All children demonstrated normal hearing in that they passed a pure-tone audiometric screening if they completed the sessions in person (n = 34) or parents reported typical hearing for children who completed sessions virtually (n = 3). All children demonstrated average nonverbal reasoning in that they achieved a standard score at or above 78 on the Primary Test of Nonverbal Intelligence (Ehrler & McGhee, 2008).

All children demonstrated a language deficit in that they achieved a standard score at or below 94 on the Structured Photographic Expressive Language Test–Third Edition (J. A. Dawson et al., 2003) or a narrative index of 91 or lower on the Test of Narrative Language–Second Edition (TNL-2; Gillam & Pearson, 2017). All children

[1]Preregistration information can be found at https://doi.org/10.17605/OSF.IO/RH9PS.

Table 1. Demographic characteristics of participants

Characteristic	*M* (*SD*)
Age (years;months)	*M* = 5;9 (*SD* = 4.46), range: 5;0–6;6
Sex assigned at birth	Males, *n* = 15 Females, *n* = 22
SPELT-3	*M* = 18 (*SD* = 17), range = 1–84
TNL-2	*M* = 24 (*SD* = 22), range = 1–84

Note. SPELT-3 = Structured Photographic Expressive Language Test–Third Edition; TNL-2 = Test of Narrative Language–Second Edition.

demonstrated a word learning/vocabulary deficit in that they achieved a standard score of 6 or lower on one of the following subtests: the Semantic subtest of the Children's Communication Checklist–Second Edition (Bishop, 2006), which emphasizes real-world communication in context; the Semantic subtest of the Diagnostic Evaluation of Language Variation (DELV; Seymour et al., 2005), which emphasizes quick word learning; or the Word Classes subtest of the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5; Wiig et al., 2013), which emphasizes semantic richness of known words. Note that all children completed all of these measures. Children were excluded if they spoke more than one language per parental report and/or the parent reported a diagnosis that would exclude a diagnosis of DLD (e.g., autism spectrum disorder).

Each child completed a battery of standardized assessments that were not used for inclusion/exclusion criteria. The goal of including this battery was to identify the measures that were most related to children's intervention outcomes including learning and posttraining retention. This battery included: the Arizona Articulation and Phonology Scale–Fourth Edition (Fudala & Stegall, 2017), the Phonological Awareness Composite and Nonword Repetition subtest from the Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2; Wagner et al., 2013), and the California Verbal Learning Test–Children's Version (CVLT-C; Delis et al., 1994). Supplemental Material S1 contains a summary of participants' percentile scores on test measures. Demographic factors were assessed via parental report including sex, race, ethnicity, parent history of language disorder, and environmental variables, which included parent marital status, education, and income.

Testing Conditions

Participants were randomized to one of three testing conditions (i.e., low, mid, high). Eleven participants were in the low testing condition, 12 were on the mid testing condition, and 14 were in the high testing condition. In the original design, to balance narrative abilities across treatment conditions, we used stratified randomization based on the narrative index score achieved on the TNL-2. The stratified TNL groups were defined as follows: (a) low, standard score of 50–81; (b) mid, standard score of 82–91; and (c) high, standard score of 92 and higher. Due to the disruption of the pandemic in some children's treatment schedules, some children's data from each stratum were excluded from the final sample. Thus, we were unable to maintain the stratified randomization of TNL scores as originally planned. Supplemental Material S1 indicates how many children from each stratum remained in each treatment group. We compared TNL scores across treatment groups of the final sample, which did not reveal a significant difference across groups (chi-square = 1.187, *df* = 4, *p* = .880). However, to account for a potential influence of TNL scores on treatment outcomes, we included TNL score in the statistical models.

Treatment

Treatment Materials

As in previous studies (Storkel et al., 2017, 2019), treatment materials consisted of 10 commercially available books. Within each book, six words were selected as target words (see Supplemental Material S2). Each participant was assigned one set of five books as their treated books and the other set of five books as their untreated books, yielding a total of 30 treated and 30 untreated words. The 60 words included 25 verbs, 19 adjectives, and 16 nouns. Each word was paired with a definition and a synonym. The definitions were constructed by consulting multiple dictionaries and selecting high-frequency words children were likely to know. For example, the definition for *horrified* was "a feeling of being scared." A context sentence was created for each word that included information to support the inference of word meaning. The sentence for *horrified* was, "After going through the haunted house, he was horrified of ghosts." We selected two color pictures for each word, one that illustrated the definition and one that illustrated the context sentence.

Outcome Measures

We used a definition task to assess knowledge of word meanings and a naming task to assess knowledge of word forms. For the definition task, the research assistant provided an open-ended prompt (e.g., "Tell me what [target word] means"). Research assistants blinded to participant's testing condition scored the definition responses between 0 and 3 based on a standard rubric adapted from McGregor et al. (2013). The rubric was created by identifying the key elements of each definition and listing examples. Definitions were given a score of 3 points if they

included all key elements, 2 points if they included one key element but lacked others, 1 point for a vague definition or an appropriate use of the word in a sentence, and 0 points for an incorrect or absent definition. Responses that received 0 or 1 were classified as incorrect, and responses that received 2 or 3 were classified as correct. Two research assistants scored each definition independently and compared scores. Discrepancies were addressed via discussion. If a score could not be agreed upon, then a third scorer resolved the discrepancy.

For the naming tasks, participants were shown a picture specific to the target word and given a prompt. For example, for the target word *marsh*, children were shown a picture of a marsh and asked, "Where do ducks and beavers live?" Research assistants blinded to participant's testing condition scored the naming responses. Responses were scored as correct if the response matched the target word. Misarticulations consistent with the child's Goldman-Fristoe Test of Articulation–Third Edition (Goldman & Fristoe, 2015) performance and age-typical misarticulations were counted as correct. Changes in verb tense including overregularizations as well as variations in noun plurality were counted as correct. Naming scores were verified by a second research assistant, and discrepancies were resolved through discussion between the first and second scorers.

Treatment Form and Intensity

During a treatment session, a research assistant read two books that were assigned as treated books to the child. The provider cycled through the books across sessions, and after the fifth book was read, they started again with the first book. Each book reading included a pre–book reading, a book reading, and a post–book reading activity (see Supplemental Material S3). Regardless of testing condition (low, mid, high), when a book was read, each word was presented across six teaching episodes. Teaching included using the treated word in a context sentence (two teaching episodes), providing a synonym (two teaching episodes), and providing a definition (two teaching episodes). Each book was read six times during treatment, totaling 36 teaching episodes for each treated word (i.e., 6 teaching episodes × 6 readings of each book).

Treatment Delivery

Each child completed two sessions a week for a total of 15 sessions. Each session lasted between 20 and 30 min. Treatment was delivered both in person and online. Prior to the 2020 pandemic, research assistants saw participants in person and used print books to deliver treatment. For online sessions, the book pages were scanned and included in a PowerPoint document and presented as a slideshow via Zoom. Once research activity resumed after pandemic-related shutdowns, the PowerPoint slideshow was presented on tablets to facilitate safety protocols for disinfecting.

Slight adjustments were made for online sessions. First, a caregiver training session was conducted with directions for using Zoom and instructions on how to support their child during the session. A mock treatment session using a book and words that were not a part of the treatment was conducted. At the beginning of each treatment session, caregivers were reminded of their role (e.g., to support, not give answers). For in-person sessions prior to the 2020 pandemic, pre–book reading and post–book reading activities were presented on the research assistant's tablet computer. For online sessions and for in-person sessions, once research activities resumed after pandemic-related shutdowns, the pre–book reading and post–book reading activities were included in the same PowerPoint used to read the books. For the children whose data were included in the analyses, 32 children completed all sessions in person, four children completed all sessions via Zoom, and one child completed some sessions in person and some sessions via Zoom.

Proportion of Testing During Treatment

As previously stated, during individual sessions, the child received six exposures to each target word during each book reading. For all children, the words of the book read during a given session were tested before the book was read via the definition test to assess word meaning and after the book was read via the naming test to assess word form (two testing trials). Feedback was not given during these beginning and end of book reading testing trials. Children were assigned to complete either low, mid, or high testing during the intervention. The low testing condition was similar to previous iterations of the intervention. Specifically, children were presented with word form and meaning information across six teaching trials for each word during a reading of the book (i.e., six teaching, two testing). For the mid testing condition, two of these teaching trials were replaced with testing trials in which the child was asked to actively retrieve target information (i.e., four teaching, four testing). For the high testing condition, four of the teaching trials were replaced with testing trials (i.e., two teaching, six testing). In cases when a retrieval prompt was given during a teaching episode, children were provided with the correct answer as feedback regardless of whether their response was correct or incorrect. Thus, all three testing conditions (low, mid, and high) had the same number of exposures to all target information. However, in all conditions, children did not receive feedback to beginning and end of book reading

tests. For all children in all conditions, the first reading of each book followed the format of the low testing condition. This was implemented to provide some initial knowledge of the words before more testing was implemented in the mid and high conditions.

Pretreatment and Posttreatment Sessions

Pretreatment sessions were administered to each child prior to beginning treatment during which each child was asked to name and define all 60 words. The definition task was administered first, followed by the naming task. Children completed two pretreatment sessions, scheduled within the same week, to minimize fatigue. During each session, the child was asked to name and define 30 words (15 treated and 15 untreated). Treated and untreated words were intermixed and tested in a randomized order. Posttreatment sessions followed a similar format to the pretreatment sessions, with two sessions scheduled within the same week. Children were tested across four posttreatment delays: immediate, 4 weeks, 8 weeks, and 12 weeks.

Treatment Providers and Fidelity

A total of 551 treatment sessions were conducted by a diverse group of paid research assistants, including two undergraduate communication sciences and disorders students (8.2% of sessions), 11 speech-language pathology graduate students (66.6% of sessions), four doctor of philosophy students (17.8% of sessions), one doctor of speech-language pathology student (5.8% of sessions), and one postdoctoral scholar (1.6% of sessions). All post master's level research assistants (n = 6) were also American Speech-Language-Hearing Association–certified speech-language pathologists. Research assistants were trained using the protocol described in Storkel et al. (2019).

Treatment fidelity was checked for 20% of sessions. Research assistants watched videos of sessions and used a checklist to verify the number of target word exposures, the number of testing opportunities, and the treatment form, resulting in three fidelity scores. The exposure fidelity score reflected whether the total target exposures were given correctly. For example, if a research assistant missed an exposure or repeated a target word (an extra exposure), a value of 1 was subtracted from the total number of exposures score. The exposure fidelity score, calculated by dividing the total number of exposures scores by the target number of exposures, was a mean of 98.55% (SD = 0.05, range: 75%–100%). The testing opportunity score indicates instances when the provider needed to administer the prompt and pause to give the child the opportunity to respond. The testing fidelity score, calculated by dividing the total number of testing opportunities provided by the specified number of testing opportunities required based on condition (i.e., low, mid, high), was a mean of 99.27% (SD = 0.19, range: 91%–100%). The treatment form score reflected that the provider stated the definitions and explanations correctly. The treatment form fidelity score, calculated by dividing the treatment form scores by the target number of treatment forms, was a mean of 99.32% (SD = 0.01, range: 95%–100%).

Analyses and Results

Treatment Effectiveness

Before addressing the primary questions, we first assessed treatment effectiveness through linear mixed-effects models in SAS Proc Mixed. We compared performance on treated and untreated words with data from the three testing groups (low, mid, and high) combined. The outcome variables at each timepoint during treatment and posttreatment assessments included the number of definitions and forms produced correctly out of the 30 treated and 30 untreated words. We conducted two models, one for definitions and one for naming of forms, with fixed effects for the within-participant effects of word category (treated, untreated), timepoint (pretest, Book Reading 1–6), and a Word Category × Timepoint interaction. We also included the between-participant fixed effect of TNL standard score, centered around the mean, to provide a control for language skills across participants. The models included random effects of intercept for participant, and participant by timepoint slope, thus allowing them to vary across participants.

The model for definitions revealed a significant interaction between word category and timepoint (t = −4.14, p < .0001) in that treated words demonstrated a steeper slope during treatment than untreated words (see Supplemental Material S4, Table 1, and Supplemental Material S5, Figure 1). There was also a significant positive effect for TNL (t = 2.26, p = .03). There was not a significant main effect for treated words compared to untreated words (t = −1.39, p = .17), indicating that there was not a significant difference between treated and untreated words at the intercept. For naming, there was a significant main effect for word category in that children produced more forms for treated as opposed to untreated words (t = −3.70, p = .0002; see Supplemental Material S4, Table 2, and Supplemental Material S5, Figure 1). There was a significant Word Category × Timepoint interaction (t = −7.54, p < .0001) in that treated words demonstrated a steeper slope than untreated words. There was also a significant positive effect for TNL (t = 4.10, p = .0002).

Learning During Treatment and Retention Posttreatment

We conducted linear mixed-effects models in SAS Proc Mixed to address the first question: How does the proportion of testing during treatment (low, mid, high) affect the number of word forms and meanings learned during the intervention and retained post-intervention? The testing conditions (low, mid, high) were randomly assigned group numbers such that the authors who analyzed the data (K.K.F., R.E.S.R.) were blind to testing condition when conducting analyses. We considered definition and naming data separately and learning and retention separately. Thus, we conducted four mixed-effects models.

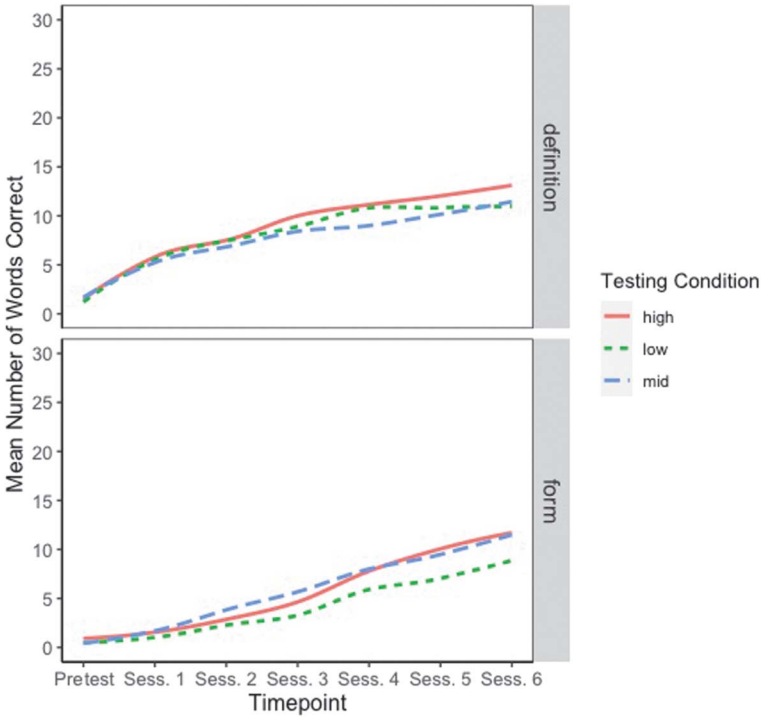
Models to assess learning included seven timepoints, pretreatment and each reading of each book (1–6). Models to assess forgetting included five timepoints, the last book reading for each treated word, and posttests administered immediately and 4, 8, and 12 weeks after the last treatment session. Some participants did not complete all the assessment timepoints (see Supplemental Material S6). Their data were still included as the model allows participants with any data to be included in the model estimation. Some participants only had partial data from some assessment points (see Supplemental Material S6). As the outcome variable was total number of words correct at each assessment point, we elected to exclude all data from assessment points with partial data.

Fixed effects in all four models included testing condition (low, mid, high), timepoint, TNL standard score, and a Testing Condition × Timepoint interaction. For the learning models, pretest was set as the intercept. For forgetting models, Book Reading 6 was set as the intercept. TNL standard score was included to provide a control for language abilities across participants and was centered around the mean. Random effects included intercept for participant and participant by timepoint slope, allowing each individual to vary from the group mean. To assess the effect of testing condition (low, mid, high), all pairwise comparisons between testing conditions were assessed.

Learning During Treatment

For the models for definitions and naming during treatment, there were significant effects of timepoint and TNL, but not for testing condition (see Figure 1 and Supplemental Material S4, Tables 3 and 4). The number of definitions ($t = 8.45$, $p < .0001$) and forms ($t = 13.36$, $p < .0001$) children produced increased across timepoints. Additionally, TNL standardized score was positively related to performance for definitions ($t = 3.34$, $p = .002$)

Figure 1. Average number of word definitions and forms produced correctly at each timepoint during training based on testing condition (high, mid, low).



and forms (t = 3.80, p = .0006). There was not a significant Timepoint × Testing Condition interaction in either model. Group means of definitions and forms correct for each condition at each timepoint are included in Supplemental Material S4, Tables 7 and 8.

Retention Posttreatment

For the models for definitions and forms posttreatment, there were significant effects of timepoint and TNL, but not for testing condition (see Figure 2 and Supplemental Material S4, Tables 5 and 6). The number of definitions (t = −5.64, p < .0001) and forms (t = −3.55, p < .001) children produced decreased across posttest timepoints. However, there was not a significant Timepoint × Testing Condition interaction in either model. TNL standardized score was positively related to performance for both definitions (t = 3.50, p = .002) and forms (t = 3.69, p < .001).

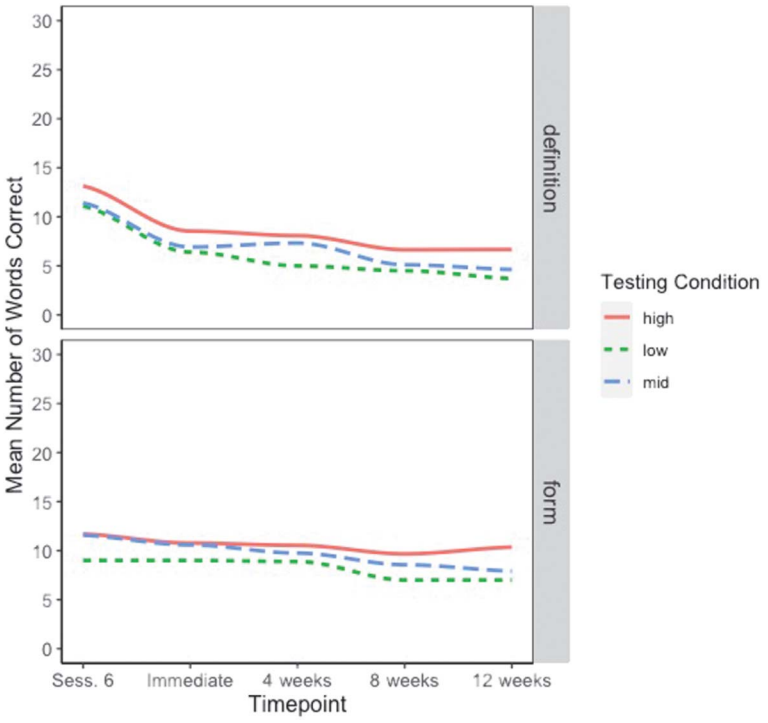
Demographic Factors Related to Learning and Retention

To determine how pretreatment factors related to learning during treatment and forgetting posttreatment, we first investigated the effects of participant demographics. We reran the learning and retention models with definitions and forms as the dependent variables with the following demographic factors: sex assigned at birth (female, male), parent history of language disorder (yes, no), parent marital status (married, all other categories), maternal education (not a high school graduate, high school graduate, some college/college graduate). For each factor, we ran a separate model. Results from demographic factors are summarized in Supplemental Material S7.

For definitions during treatment, there was not a significant effect of sex or parent history of language disorder. There was a significant effect of marital status (t = 2.21, p = .03) in that children whose parents were married produced more definitions correctly than all other categories. However, there was not a significant Timepoint × Marital Status interaction (t = −1.59, p = .11), indicating that there was not a difference in learning slope based on parental marital status. For maternal education, children whose mother was a high school graduate (t = 3.31, p = .001) or who had completed some college/college graduate (t = 3.44, p < .001) produced more definitions correctly than children whose mother had not completed high school. However, there was not a difference in learning slopes based on maternal education.

For forms during treatment, there was not a significant effect of sex, parent history of language disorder, or marital status. For maternal education, children whose

Figure 2. Average number of word definitions and forms produced correctly at the last book reading and each posttraining test point based on testing condition (high, mid, low).



mother was a high school graduate ($t = 2.82$, $p < .01$) or who had completed some college/college graduate ($t = 3.54$, $p < .001$) produced more forms correctly than children whose mothers had not completed high school. Additionally, children whose mother was a high school graduate ($t = -2.83$, $p = .005$) or who had completed some college/college graduate ($t = -2.55$, $p = .011$) demonstrated a steeper slope during treatment than children whose mother had not completed high school.

For definitions posttreatment, there was not a significant main effect of sex, parent history of language disorder, marital status, or maternal education. However, there was a significant Timepoint × Sex interaction in that males forgot more definitions than females ($t = 4.19$, $p < .0001$; see Supplemental Material S5, Figure 2).

For forms posttreatment, there was not a significant main effect of sex, parent history of language disorder, or marital status. There was a significant main effect of maternal education ($t = 3.75$, $p = .03$). Children whose mothers had completed some college/college graduate performed better than children whose mother had not completed high school ($t = 2.22$, $p = .03$). There was no difference between the categories of not a high school graduate and a high school graduate ($t = .64$, $p = .52$). There was not a significant Maternal Education × Timepoint interaction ($t = 0.92$, $p = .50$) indicating that maternal education was not related to forgetting rate.

Standardized Test Scores Related to Learning and Retention

In addition to demographic factors, we investigated how standardized test scores were related to learning and retention. For the learning models, test scores included three main categories: language (CELF, DELV), working memory (CTOPP, CVLT list free recall trial 1), and episodic learning (CVLT list free recall trial 5, CVLT slope of list free recall). For forgetting models, test scores included two main categories, learning during treatment (mean words correct across all book reading sessions during treatment, and words correct at last treatment session) and decay (CVLT short-delay free recall, CVLT long-delay free recall). Results from standardized test scores are summarized in Supplemental Material S7.

For definitions and forms during treatment, there was not a significant effect of CELF, CTOPP, or any of the CVLT scores. There was a significant main effect of DELV in that children with a higher DELV score produced more definitions correctly ($F = 6.97$, $p = .01$) and more forms correctly ($F = 7.08$, $p = .01$). However, there was not a significant DELV by timepoint slope during treatment for definitions ($F = 0.51$, $p = .48$) or for forms ($F = 2.46$, $p = .12$).

For retention posttreatment of definitions, there was not a significant effect of CVLT short-delay free recall. Children with a higher mean number of definitions correct during treatment demonstrated higher performance posttreatment ($t = 4.25$, $p = .0001$) and demonstrated better retention posttreatment via a shallower slope ($t = 3.03$, $p = .003$). There was a significant main effect of number of definitions at the last treatment (Book 6) and retention ($F = 33.53$, $p < .0001$), but there was not a significant timepoint by last treatment slope ($F = 1.22$, $p = .27$). There was a significant main effect of CVLT long-delay free recall score in that children with a higher score demonstrated a higher posttreatment score ($F = 6.01$, $p = .02$). There was not a significant Timepoint × CVLT long-delay interaction ($F = 3.28$, $p = .07$).

For naming retention posttreatment, children with a higher mean number of words correct during treatment demonstrated higher performance posttreatment ($t = 9.91$, $p < .0001$). However, there was not a significant Timepoint × Mean Number of Words During Treatment interaction ($t = 0.15$, $p = .88$), indicating that the mean number of words was not related to forgetting rate. There was a significant main effect of CVLT short-delay free recall score. Children with a higher score demonstrated a higher posttreatment score ($F = 7.00$, $p = .01$). However, there was not a significant Timepoint × CVLT short-delay interaction ($F = 0.28$, $p = .60$). Similarly, there was a main effect of CVLT long-delay free recall and posttraining performance. Children with a higher score produced more forms correctly posttraining ($F = 4.59$, $p = .03$). However, there was not a significant Timepoint × CVLT long-delay interaction ($F = 0.54$, $p = .46$). There was a Timepoint × Last Treatment interaction (i.e., slope) that approached but did not achieve significance ($F = 3.56$, $p = .06$) in that children with a higher number of words at the last treatment session demonstrated less posttraining forgetting.

Discussion

In the current study, we modified Storkel's interactive book reading intervention to include retrieval-based practice opportunities throughout training. All children received 36 exposures to the target words. However, the proportion of retrieval opportunities varied across three conditions: low, mid, and high testing. Of note, the low testing condition was similar to past iterations of the intervention. For the current study, we asked the following key questions:

1. How does the proportion of within-session testing during treatment (low, mid, high) affect the number of word forms and meanings learned during the intervention and retained post-intervention?

2. Which pretreatment characteristics of children with DLD are associated with learning during the intervention and retention post-intervention?

With regard to the first question, we found that the proportion of testing during treatment did not significantly relate to the number of word meanings and forms learned during the intervention. The descriptive results trended in the expected direction. Based on group means, children learned the most meanings and forms in the high, then mid, then low conditions. A lack of statistical significance across conditions may be related to high heterogeneity across children with DLD and a smaller final sample size than planned due to the disruption of the 2020 pandemic. However, note that if we had observed statistically significant differences with a larger sample size, these differences may not be clinically meaningful. For definitions and forms, the means of the testing conditions (high, mid, low) varied in one to three words (see Supplemental Material S4) out of the 30 possible words to learn in the set. Overall, we cannot conclude that more retrieval-based learning opportunities (mid, high) contributed to better learning than the more passive version of the intervention (low). However, children demonstrated intervention effectiveness in that they learned more treated than untreated words. Thus, based on the current results, retrieval opportunities neither helped nor hurt learning during treatment.

With regard to posttraining retention, we found that the proportion of retrieval opportunities during treatment did not significantly relate to the number of word meanings and forms retained post-intervention. The descriptive results trended in the expected direction in some cases, but not all. For example, if we subtract the mean number of words produced correctly at the last book reading from the mean number of words produced correctly at the final posttest assessment point, children forget, on average, 6.47, 6.79, and 7.40 definitions for the high, mid, and low testing conditions, respectively. For forms, children forget, on average, 1.35, 3.67, and 2.00 words for the high, mid, and low testing conditions, respectively. This follows the expected directions of results, in that high testing led to more retention but does not follow the expected direction when comparing mid to low. Given the current data, we cannot conclude that more retrieval-based learning opportunities contributed to better posttraining retention. However, once again, incorporating retrieval-based practice appears to have neither helped nor hurt posttraining retention.

Although the low testing condition was similar to past iterations of the intervention, there is a critical difference. Specifically, in the current version of the intervention, all children (in low, mid, and high testing conditions) were tested at least once on word meanings and forms during each reading of each target book to have a more precise understanding of learning trajectories. This differed from past versions of the intervention in which children were taught primarily through passive exposure, but children's knowledge of meanings and forms were tested periodically during training.

Given that all children were tested during every training session in the current version, this may have obscured differences across testing groups. In fact, in research with adults with TD, Rawson and Dunlosky (2012) found that when training occurs during a single session, multiple successful retrievals during that session relate to better posttraining retention than a single successful retrieval. However, when learning extends across many sessions, the number of sessions with successful retrievals is a better predictor of posttraining retention than retrieving target items multiple times during a session. The current results converge with findings from Rawson and Dunlosky in that the proportion of opportunities to retrieve words during sessions did not relate to differences in learning and posttraining retention. Instead, the opportunity to retrieve words during all training sessions may have similarly supported children across the three testing conditions. Past research with adults and children with TD has investigated how many sessions learners should successfully retrieve each target item during training to support retention (Gordon & Lowry, 2024; Rawson & Dunlosky, 2012). A natural extension of this work is to investigate how the number of sessions with successful retrievals during an intervention supports retention by children with DLD.

Individual Factors

To address our second question, we assessed how children's demographic characteristics and standardized test scores related to learning and retention. Similar to past research on word learning in children with DLD (e.g., Gordon, Storkel, et al., 2021; Jackson et al., 2016), we included standardized assessments of language knowledge and abilities and working memory. However, this study was unique in that we included a larger battery of memory assessments than are typically included. We first discuss how demographic factors related to learning and retention, followed by language standardized test scores, and then memory standardized test scores.

Regarding the relationship between demographic factors and learning, parental marital status and maternal education were related to performance during treatment. The specific findings varied somewhat across definitions and forms. Children whose parents were married produced more definitions correctly than all other marital status categories. Additionally, maternal education was positively related to the number of definitions produced. However,

neither factor was related to learning rate for definitions. For naming, marital status was not a significant predictor, but maternal education was positively related to number of forms produced correctly and to learning rate. These findings provide some information to speech-language pathologists about how children are likely to respond to the intervention. However, it is difficult to identify the specific mechanisms that led to better performance. Maternal education and marital status are typically used as proxy measures for the language input and economic resources (e.g., number of books) that children receive in the home (Vernon-Feagans et al., 2020). However, to understand the specific mechanisms that drive the current findings, research in which we more directly measure these factors instead of using proxy measures is important.

Regarding the relationship between demographic factors and posttraining retention. For naming, maternal education was positively related to the number of forms produced. However, maternal education was not related to forgetting rate. For definitions, there was not a main effect of sex. However, boys forgot more definitions than girls. As long-term retention is rarely assessed in vocabulary research, there is sparse evidence of whether sex differentially affects retention. However, the current findings coincide with McGregor et al. (2020) in that they found better overall performance by adult females than males when word learning retention was measured 1 day, 1 week, and 1 month after training. Of note, in McGregor et al. (2020), each learner continued the training via retrieval-based practice until they achieved mastery of the word set. Thus, it was not the case that learners who learned more words retained more words. Instead, the finding that women performed better than men captures differences in retention when all learners were trained to a similar level. Girls with DLD do exhibit more breadth and depth of vocabulary knowledge than boys with DLD (McGregor et al., 2013). Thus, our current findings may be explained by richer semantic and phonological networks of the girls. If they can link new learning to these richer networks in long-term memory, this may lessen posttraining forgetting.

For the relationship between the language standardized test scores and learning during treatment, the DELV score was positively related to the overall number of definitions and forms produced correctly but was not significantly related to learning rate. Additionally, the TNL score was positively related to the number of definitions and forms produced correctly during treatment and posttreatment. Overall, these results are not surprising. Children with richer language knowledge and abilities tend to learn more words during vocabulary interventions than children with less rich language knowledge (Coyne et al., 2019). However, the fact that language knowledge was not significantly related to learning rate is promising. Specifically, the current findings indicate that children who start the intervention with more language knowledge and abilities are likely to encode more of the words taught, but children who vary in language knowledge and abilities will not differ in learning rate during the supportive intervention. We observed a similar finding in Gordon, Storkel, et al. (2021) when we taught laboratory-created form–object pairs to children with DLD and TD across six training days. Children with less language knowledge, in this case the DLD group, demonstrated less encoding during the initial session. However, the groups did not differ in their learning rates across training days.

Regarding posttraining retention, children who produced a higher mean number of definitions correctly during treatment demonstrated better overall posttraining performance and a shallower forgetting slope for definitions. For forms, children who produced a higher mean number of forms correctly during treatment performed better overall posttreatment but did not demonstrate a difference in forgetting slope. This finding for forms is similar to findings from Gordon, Storkel, et al. (2021). Specifically, across a 1-month delay, children with DLD and TD tended to retain what they had learned during treatment. Notably, different factors appear to be related to posttraining retention of meanings and forms. We discuss this in more detail in the form and meaning section below.

The standardized tests assessing working and episodic memory were not related to learning forms or definitions during treatment. With regard to posttraining retention, for definitions, children with a higher CVLT long-delay free recall score demonstrated a higher posttreatment score. For forms, CVLT short-delay free recall score and CVLT long-delay free recall score were related to a higher posttreatment score. None of the memory scores related significantly to forgetting rate. The finding that these tasks are not related to the learning or forgetting rate is not particularly surprising as standardized tests assessing encoding and retention of information are administered in a single session. In the current study, learning is assessed over a prolonged slow-mapping process, which includes encoding during sessions, consolidation between sessions, and re-encoding during subsequent sessions. Additionally, retention is assessed in standardized tests after a very short delay, such as 10 min in the CVLT. In contrast, in the current study. retention is assessed over weeks and months.

Overall, our findings suggest that clinicians should not rely on pretreatment standardized test scores or demographic factors to predict learning rates by individual children engaged in an interactive book reading intervention. Instead, the child's own progress in the intervention during training should be used to determine whether the

intervention is effective for them. In a subsequent publication, we will determine the earliest timepoint during training that children can reliably be identified as responders or nonresponders to inform clinical decision making. Additionally, standardized test scores and demographic factors provide little information about children's ability to retain taught information post-intervention. Instead, retention of taught words should be periodically assessed during posttraining retention intervals to determine how much information children forgot. Research with adults suggests that periodic posttreatment assessments and retraining can boost long-term retention of taught information (Cepeda et al., 2009). Through future research, we can determine how often learning should be probed and retrained post-intervention to support optimal long-term retention by children with DLD.

Learning and Retaining Forms and Meanings

During the current study, children demonstrated similar learning rates of forms and meanings during training. This finding stands in contrast to past literature in that children with DLD tend to demonstrate more difficulty learning forms than meanings (Gray, 2004; McGregor et al., 2017). Notably, in most past word learning studies, children with DLD are trained on forms linked to concrete objects. For example, in a review of word learning abilities by individuals with DLD (Jackson et al., 2019), 50 out of the 70 studies included nouns linked to concrete objects, and 14 included highly picturable verbs. Additionally, most word learning studies include only one or a small number of training sessions. In contrast, in the current study, learning occurred across many sessions. Overall, the current study suggests that children with DLD may have just as much difficulty learning meanings as well as forms when word meanings are more abstract.

Of note, children in the current study appear to demonstrate differential retention of forms and meanings during the extended posttraining delay period. Most individual factors were related to overall performance during posttraining, not to forgetting rate. However, children who produced a higher mean number of definitions correctly during treatment demonstrated a shallower forgetting rate of definitions. Additionally, girls demonstrated a shallower forgetting rate of definitions than boys. Research investigating the relationship between children's ability to encode novel words and individual factors, such as verbal working memory skills, abound (Adlof & Patten, 2017; Jackson et al., 2016). However, posttraining retention is rarely assessed, and when it is assessed, it is typically measured over relatively short timeframes. There is sparse, but notable, research assessing the relationship between vocabulary knowledge on children's consolidation of newly learned words. Across a series of studies, Henderson, James, and colleagues found that vocabulary knowledge is positively related to better consolidation of new word learning (Henderson & James, 2018; Henderson et al., 2021; James et al., 2020). Note that these studies typically include one training session and assessments of retention up to a week posttraining. Also, note that their findings are variable in which aspect of word learning is affected by prior vocabulary knowledge, forms or meanings, and in some cases no effect of prior vocabulary knowledge was observed.

There are two likely explanations for why we observed individual factors related to the forgetting of definitions, but not forms, across the posttraining delay. As previously noted, children with richer semantic networks may be able to retain word meanings better than children with less rich semantic networks. During the consolidation process, which often occurs during overnight sleep, new learning is linked to knowledge stored in the child's long-term memory (Davis & Gaskell, 2009). Given that meaning information was more complex in the current study (i.e., abstract definitions) than in past studies (i.e., concrete objects), these semantic networks may have been particularly important to support both learning and posttraining retention of word meanings.

A second explanation for why we observed individual factors related to meanings and not forms is that our assessment of meaning was quite challenging. Children were asked to produce multiple key features of the definition and were not given a visual cue (i.e., a picture) to support retrieval of word meaning. In contrast, children were shown a picture and given a sentence prompt when asked to retrieve forms. Producing correct definitions for words with abstract meanings is a challenging task for children and adults (Dourou et al., 2020). Thus, children who were no longer able to produce a correct definition across long-term delays may have, nevertheless, retained an understanding of the word meaning. In future research, it is important to incorporate assessments after long-term delays that vary in their level of cuing. In this way, we can determine what children retain long term and how much support they need to retrieve that information. This will also inform how much retraining they need at these posttraining assessment points to support the continued retention of word meanings.

Conclusions

The current study and past studies in this line of research demonstrate that interactive book reading is effective to support word learning in children with DLD. However, we observed that retrieval-based practice opportunities

neither helped nor hurt learning and posttraining retention. Additionally, few demographic factors or standardized test scores were related to learning rate or posttraining retention rate. These results are promising in that children with DLD who vary on a number of factors can benefit from the intervention in that they learn words during treatment. Similarly, children who vary on a number of factors are likely to demonstrate similar retention of learning across posttreatment intervals.

Data Availability Statement

De-identified data are available from the first author upon request.

Acknowledgments

This research was supported by the National Institute on Deafness and Other Communication Disorders Grant R01DC012824, awarded to Holly L. Storkel. The opinions expressed in this article are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government. When H. L. Storkel transitioned to the National Institutes of Health, Danielle Moss became the principal investigator of the project, and Katherine R. Gordon became the consultant. At the completion of the project, Danielle Moss and Katherine R. Gordon collaborated on data cleaning, statistical analyses with Rebecca E. Swinburne Romine and Kandace K. Fleming, and manuscript preparation. We would like to acknowledge the research assistants who assisted with administration and scoring of the intervention: Amelia Brown, Grace Buckalew, Holly Maddalena, Lily Shaw, and Lauren Tigner.

References

Adlof, S. M., & Patten, H. (2017). Nonword repetition and vocabulary knowledge as predictors of children's phonological and semantic word learning. *Journal of Speech, Language, and Hearing Research, 60*(3), 682–693. https://doi.org/10.1044/2016_JSLHR-L-15-0441

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford.

Bishop, D. V. M. (2006). *CCC-2: Children's Communication Checklist–Second Edition*. Pearson.

Booth, A. E. (2009). Causal supports for early word learning. *Child Development, 80*(4), 1243–1250. https://doi.org/10.1111/j.1467-8624.2009.01328.x

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning, 1*(1), 69–84. https://doi.org/10.1007/s11409-006-6894-z

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

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice. *Experimental Psychology, 56*(4), 236–246. https://doi.org/10.1027/1618-3169.56.4.236

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Cirrin, F. M., & Gillam, R. B. (2008). Language intervention practices for school-age children with spoken language disorders: A systematic review. *Language, Speech, and Hearing Services in Schools, 39*(1), S110–S137. https://doi.org/10.1044/0161-1461(2008/012)

Coyne, M. D., McCoach, D. B., Ware, S., Austin, C. R., Loftus-Rattan, S. M., & Baker, D. L. (2019). Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children, 85*(2), 163–179. https://doi.org/10.1177/0014402918789162

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Dawson, J. A., Stout, C. E., & Eyer, J. A. (2003). *SPELT-3: Structured Photographic Expressive Language Test–Third Edition*. Janelle Publications.

Dawson, N., Hsiao, Y., Banerji, N., Tan, A. W. M., & Nation, K. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research, 1*(1), 9–53. https://doi.org/10.34842/5we1-yk94

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *CVLT-C: California Verbal Learning Test: Children's Version*. Pearson.

Dourou, C., Gavriilidou, Z., & Markos, A. (2020). Definitional skills and preferred definition types according to age, gender, educational level and career orientation. *International Journal of Research Studies in Education, 9*(2), 29–49. https://doi.org/10.5861/ijrse.2020.5021

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology (Educational Reprints No. 3)*. Teacher's College, Columbia University.

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

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A systematic review of the testing effect in learning. *Paideia, 23*(56), 397–406. https://doi.org/10.1590/1982-43272356201314

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

Fudala, J. B., & Stegall, S. (2017). *Arizona Articulation and Phonology Scale–Fourth Edition (Arizona-4)*. Western Psychological Services.

Gillam, R. B., & Pearson, N. A. (2017). *Test of Narrative Language*. Pro-Ed.

Goldman, R., & Fristoe, M. (2015) *Goldman-Fristoe Test of Articulation–Third Edition (GFTA-3)*. AGS.

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., & Lowry, S. L. (2024). Fostering retention of word learning: The number of training sessions children retrieve words positively relates to post-training retention. *Journal of Child Language, 51*(3), 710–719. https://doi.org/10.1017/S0305000923000752

Gordon, K. R., McGregor, K. K., & Arbisi-Kelm, T. (2021). Optimising word learning in post-secondary students with developmental language disorder: The roles of retrieval difficulty and retrieval success during training. *International Journal of Speech-Language Pathology, 23*(4), 405–418. https://doi.org/10.1080/17549507.2020.1812719

Gordon, K. R., McGregor, K. K., Waldier, B., Curran, M. K., Gomez, R. L., & Samuelson, L. K. (2016). Preschool children's memory for word forms remains stable over several days, but gradually decreases after 6 months. *Frontiers in Psychology, 7,* Article 1439. https://doi.org/10.3389/fpsyg.2016.01439

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

Gordon, K. R., Storkel, H. L., Lowry, S. L., & Sultani, M. J. (2024). A word-learning intervention pilot study utilizing principles of retrieval- and criterion-based learning for children with developmental language disorder. *American Journal of Speech-Language Pathology, 33*(2), 530–551. https://doi.org/10.1044/2023_AJSLP-23-00037

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders, 36*(2), 129–151. https://doi.org/10.1016/S0021-9924(03)00003-0

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)

Gupta, P., & Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1536), 3755–3771. https://doi.org/10.1098/rstb.2009.0132

Henderson, L. M., & James, E. (2018). Consolidating new words from repetitive versus multiple stories: Prior knowledge matters. *Journal of Experimental Child Psychology, 166,* 465–484. https://doi.org/10.1016/j.jecp.2017.09.017

Henderson, L. M., van Rijn, E., James, E., Walker, S., Knowland, V. C. P., & Gaskell, M. G. (2021). Timing storytime to maximize children's ability to retain new vocabulary. *Journal of Experimental Child Psychology, 210,* Article 105207. https://doi.org/10.1016/j.jecp.2021.105207

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

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2019). The evaluation of word-learning abilities in people with developmental language disorder: A scoping review. *International Journal of Language & Communication Disorders, 54*(5), 742–755. https://doi.org/10.1111/1460-6984.12490

James, E., Gaskell, M. G., & Henderson, L. M. (2020). Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: It'll be alright on the night? *The Journal of Child Psychology and Psychiatry, 61*(10), 1104–1115. https://doi.org/10.1111/jcpp.13253

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools, 36*(1), 17–32. https://doi.org/10.1044/0161-1461(2005/003)

Kaiser, A. P., & Roberts, M. Y. (2011). Advances in early communication and language intervention. *Journal of Early Intervention, 33*(4), 298–309. https://doi.org/10.1177/1053815111429968

Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 53*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of learning and motivation—Advances in research and theory* (Vol. 61, pp. 237–284). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968. https://doi.org/10.1126/science.1152408

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

Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology, 50*(4), 994–1000. https://doi.org/10.1037/a0035202

Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice, 25*(3), 124–136. https://doi.org/10.1111/j.1540-5826.2010.00310.x

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. https://doi.org/10.3102/0034654310377087

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

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders, 42*(1), 35–47. https://doi.org/10.1007/s10803-011-1210-4

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/10.3851/IMP2701. Changes

McGregor, K. K., Sheng, L., & Ball, T. (2007). Complexities of expressive word learning over time. *Language, Speech, and Hearing Services in Schools, 38*(4), 353–364. https://doi.org/10.1044/0161-1461(2007/037)

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*(2), 979–1007. https://doi.org/10.3102/0034654309332561

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education & Development, 19*(1), 7–26. https://doi.org/10.1080/10409280701838603

Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology, 3,* Article 41. https://doi.org/10.3389/fpsyg.2012.00041

Muraki, E. J., Reggin, L. D., Feddema, C. Y., & Pexman, P. M. (2022). The development of abstract word meanings. *Journal of Child Language, 52*(1), 195–207. https://doi.org/10.1017/S0305000923000569

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders, 41*(3), 335–354. https://doi.org/10.1080/13682820600602295

Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly, 46,* 112–125. https://doi.org/10.1016/j.ecresq.2018.04.001

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review, 24*(3), 419–435. https://doi.org/10.1007/s10648-012-9203-1

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research, 48*(6), 1397–1411. https://doi.org/10.1044/1092-4388(2005/097)

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Seymour, H. M., Roeper, T. W., & De Villiers, J. G. (2005). *Diagnostic Evaluation of Language Variation–Norm-Referenced Test*. The Psychological Corporation.

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

Towson, J. A., Akemoglu, Y., Watkins, L., & Zeng, S. (2021). Shared interactive book reading interventions for young children with disabilities: A systematic review. *American Journal of Speech-Language Pathology, 30*(6), 2700–2715. https://doi.org/10.1044/2021_AJSLP-20-00401

Vernon-Feagans, L., Bratsch-Hines, M., Reynolds, E., & Willoughby, M. (2020). How early maternal language input varies by race and education and predicts later child language. *Child Development, 91*(4), 1098–1115. https://doi.org/10.1111/cdev.13281

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology, 3,* Article 46. https://doi.org/10.3389/fpsyg.2012.00046

Wagner, R. K., Torgensen, J., Rashotte, C., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing–Second Edition*. Pro-Ed.

Wiig, E. H., Semel, E., & Secord, W. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition*. Pearson.

Zipoli, R. P., Coyne, M. D., & McCoach, D. B. (2011). Enhancing vocabulary intervention for kindergarten students: Strategic integration of semantically related and embedded word review. *Remedial and Special Education, 32*(2), 131–143. https://doi.org/10.1177/0741932510361262