Title: Word Learning by Preschool-Age Children with Developmental Language Disorder: Impaired Encoding and Robust Consolidation during Slow Mapping

Authors and Affiliations

Katherine R. Gordon (Center for Childhood Deafness, Language and Learning, Boys Town National Research Hospital)

Holly L. Storkel (Department of Speech-Language-Hearing, University of Kansas)

Stephanie Uglow (Center for Childhood Deafness, Language and Learning, Boys Town National Research Hospital)

Nancy B. Ohlmann (Center for Childhood Deafness, Language and Learning, Boys Town National Research Hospital)

Conflict of Interest The authors declare that there was no conflict of interest at the time of publication.

Corresponding Author: Katherine R. Gordon, Center for Childhood Deafness, Language, and Learning, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131, Katherine.Gordon@boystown.org

Abstract

Purpose: Learning novel words, including the specific phonemes that make up word forms, is a struggle for many individuals with Developmental Language Disorder (DLD). Building robust representations of words includes encoding during periods of input and consolidation between periods of input. The primary purpose of the current study is to determine differences between children with DLD and with typical development (TD) in the encoding and consolidation of word forms during the slow mapping process.

Method: Preschool-age children (DLD = 9, TD = 9) were trained on nine form-referent pairs across multiple consecutive training days. Children's ability to name referents at the end of training days indicated their ability to encode forms. Children's ability to name referents at the beginning of training days after a period of overnight sleep indicated their ability to consolidate forms. Word learning was assessed one-month after training to determine long-term retention of forms.

Results: Throughout training, children with DLD produced fewer forms correctly and produced forms with less phonological precision than children with TD. Thus, children with DLD demonstrated impaired encoding. However, children with and without DLD demonstrated a similar ability to consolidate forms between training days and to retain forms across a one-month delay.

Conclusions: Difficulties with word form learning are primarily driven by deficits in encoding for children with DLD. Clinicians and educators can support encoding by providing children with adequate exposures to target words via robust training that occurs across multiple sessions.

Introduction

A child's vocabulary knowledge at school entry is the foundation upon which academic, social, and literacy skills are built (Pace, Alper, Burchinal, Golinkoff, & Hirsh-Pasek, 2019). Children with Developmental Language Disorder (DLD)[1] are at risk for poor vocabulary knowledge at school entry (Norbury et al., 2016; Tomblin et al., 1997). DLD is a fairly common disorder, affecting 7% of the population (Tomblin et al., 1997). Most children with DLD are not diagnosed and do not receive clinical services during early childhood (Norbury et al., 2016). This is problematic. Through effective vocabulary intervention during the preschool years, children with DLD could enter school better equipped to access academic content, to build essential literacy skills, and to learn additional vocabulary from verbal and written content.

For vocabulary interventions to be successful, it is important to understand areas of relative strength and weakness in the word learning process for children with DLD. The Complementary Systems Account of word learning, which is based on a large body of neurological and behavioral evidence, posits that word learning occurs through two essential systems (Davis & Gaskell, 2009; McClelland, McNaughton, & O'Rilly, 1995). These systems manage encoding during periods of input and consolidation between periods of input. These systems could be differentially impaired in individuals with DLD relative to their peers with typical development (TD) (Storkel, 2015; Storkel, Komesidou, Fleming, & Romine, 2017).

During an initial experience with an unfamiliar word, the child may encode representations of the word form, which includes the specific phonemes that make up the word and their order; the word meaning; and the link between the two (Figure 1). This is often referred

[1] Developmental Language Disorder has been called by a variety of terms including Specific Language Impairment, Language Impairment, and Primary Language Disorder. We use the term, Developmental Language Disorder, following the recommendations of Bishop et al., (2017).

to as fast mapping (Swingley, 2010). At this stage, representations of the form, meaning, and link are associated with hippocampal activity (Dumay & Gaskell, 2007). After this initial period of input, these representations must be consolidated to be retained (Simon, Gomez, & Nadel, 2020). Consolidation entails the off-line mental activities related to the target information that occur between periods of input. Specifically, during sleep recently encoded information is often re-activated (Davis & Gaskell, 2009). Long-term memories associated with cortical activity that are related to the newly encoded information are also activated (James, Gaskell, Weighall, & Henderson, 2017). Thus, the newly encoded information can become associated with cortical activity and consolidated with long-term memories. During sleep, representations of the form, meaning, and link can be consolidated successfully in that they are strengthened and maintained (Henderson, Weighall, Brown, & Gaskell, 2012). Consolidation can be less successful, in that after a period of sleep children remember less about a word than they did before sleep (Malins et al., 2020). Consolidation can also be unsuccessful when encoded representations are forgotten completely (Storkel, 2015).

If a word is consolidated, the child can retrieve her stored representations of the form, meaning, and link during the next period of input. When representations are retrieved they become activated in working memory and become malleable (McKenzie & Eichenbaum, 2011). Re-encoding is similar to encoding in that children are building representations during a period of input. The key difference is that during re-encoding, the child refines a previously encoded representation by correcting, adding detail to, and linking additional information with the representation (Nader & Hardt, 2009). It is through the processes of encoding, consolidation, retrieval, and re-encoding across multiple experiences that the child slowly refines her representations associated with the word. This process is referred to as slow or extended mapping (Swingley, 2010).

Learning the forms and meanings of words can be challenging for individuals with DLD (Kan & Windsor, 2010). However, learning forms is a particular area of weakness (McGregor et al., 2013). Notably, adding a word to one's expressive vocabulary such that the form can be readily retrieved and produced accurately requires extensive input (McGregor, Sheng, & Ball, 2007). For a child to add a word to her expressive vocabulary, she must encode the specific phonemes that make up the form and their order, link the representation of the form to the word meaning, build the representation of the form across exposures to a high level of phonological precision, retrieve the representation of the form after a delay, and enact a motor plan to produce it (Benham & Goffman, 2020). Learning forms to this level is challenging even for children with TD. For example, Gray (2005) found that preschool-age children with and without DLD required 80-90 exposures before they reliably named referents. Given that learning forms is particularly challenging, in the current study we focus on the ability of preschool-age children with and without DLD to learn forms across sessions that occur on subsequent days. Critically, to understand differences between children with and without DLD in the slow-mapping process, it is essential to account for differences in encoding and in consolidation. Specifically, children with DLD could encode less phonological information about forms than children with TD during periods of input and/or could lose more phonological information about forms between periods of input. Below we review the current literature on differences between children with and without DLD in encoding and consolidation.

Encoding Words from Input

When first exposed to an unfamiliar word, children encode sparse and short-lasting representations of forms (Horst & Samuelson, 2008; Vlach, 2019). Initial representations of forms are sparse in that they lack phonological precision and short lasting in that they decay rapidly from memory (Munro, Baker, McGregor, Docking, & Arciuli, 2012). Given the short-lasting nature of the representation, a single exposure to a novel form is often insufficient to support learning that lasts. However, if the child hears multiple presentations of the target form during the same session, he is able to refine and strengthen his representation of the form (Bishop, Barry, & Hardiman, 2012). Thus, at the end of a session with multiple presentations, the child may have developed a robust enough representation of the form that it can be successfully consolidated and retrieved during the next experience with the word.

Children's ability to encode novel words from input is strongly associated with their working memory abilities (see Gathercole, 2006 for a review). Verbal working memory includes the ability to encode, temporarily store, and manipulate verbal information. Children with poorer working memory abilities are at risk for slower vocabulary growth during the preschool years and poorer vocabulary knowledge at school entry (Archibald, 2017). Many children with DLD demonstrate deficits in verbal working memory which makes learning forms from verbal input challenging (Alloway & Gathercole, 2006; Alt, 2011; Archibald & Gathercole, 2006; Montgomery, Magimairaj, & Finney, 2010). For example, when given one presentation of a novel form, individuals with DLD encode less phonologically precise representations than their peers (Graf Estes, Evans, & Else-Quest, 2007). Thus, during the next presentation of the form, individuals with DLD have a less phonologically precise representation to build upon. At the end of a session with a set number of exposures, children with DLD are likely to have encoded a less phonologically precise representation of the form than their peers with TD (Bishop et al., 2012).

Because verbal working memory is a common weakness for children with DLD, the stages of encoding and re-encoding are likely to be areas of weakness in the slow mapping process.

Consolidating Words Between Periods of Input

To date, the majority of research on word learning in both typical and atypical populations has focused on children's ability to encode words during periods of input. Thus, there is limited research on how consolidation varies across children and whether it is an area of weakness for individuals with DLD. Current results on this topic are mixed. Through a series of studies, McGregor and colleagues found that consolidation is an area of relative strength for college students with DLD (McGregor, Arbisi-Kelm, & Eden, 2017; McGregor, Arbisi-Kelm, Eden, & Oleson, 2020; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017). Leonard and colleagues have demonstrated a comparable pattern of results for preschool-age children with DLD when tested five minutes and one week after training (see Leonard & Deevy, 2020 for a review).

In contrast, other researchers have demonstrated impaired consolidation of word learning for individuals with DLD. When exposed to novel words through an incidental learning paradigm, preschool-age children with DLD demonstrated poorer retention than their peers (Rice, Oetting, Marquis, Bode, & Pae, 1994). In a more recent study, 3rd and 4th grade children with DLD demonstrated poorer retention of word learning than their peers when tested the following day (Malins et al., 2020). Children's phonological representations of forms were more stringently tested in this study than is typical for word learning studies. For every form in the set (e.g., pibu), there was a form that shared an onset with that form (e.g., pibo) and a form that rhymed with that form (e.g., dibu). Thus, children were required to encode and retain phonologically precise representations of forms to correctly identify the target referent. In this case, the fact that children with DLD demonstrated poorer retention of learning than their peers suggests that they had difficulty retaining the phonological specificity of their encoded representations of forms.

The Current Study

The primary goal of the current study is to compare the abilities of preschool-age children with and without DLD to encode, consolidate, and re-encode forms during the slow mapping process. To date, the majority of studies on encoding and consolidation include one training session to assess encoding and one delayed session to assess consolidation. These studies do not capture how children refine representations across multiple training sessions and how children consolidate these representations between training sessions. However, these processes are essential for a child to add a word to her expressive vocabulary. To more accurately capture the slow-mapping process, in the current study we trained children on nine form-referent pairs across subsequent training days until each child demonstrated learning of all pairs or completed a total of six training days. We also assessed their ability to retain learning over a one-month post-training delay.

For this study, we assessed differences in encoding and consolidation across groups when they were given highly supportive training. We utilized retrieval-based practice as this strategy supports encoding and long-term retention in children with and without DLD (see Leonard & Deevy, 2020). We asked children to actively retrieve forms throughout training and provided feedback to their responses. We included tasks that varied in retrieval demands as this strategy can be particularly effective in supporting encoding and long-term retention of information (Adesope, Trevisan, & Sundararajan, 2017). Specifically, the learner can strengthen retrieval pathways when she successfully retrieves information via tasks that include cuing. However, she can also slowly enhance her ability to successfully retrieve information without cuing via free recall tasks (see Gordon, 2020 for an overview of this literature).

Our research questions are the following. First, do preschool-age children with DLD produce fewer forms correctly and produce forms with less phonological precision during periods of input in comparison to children with TD demonstrating poorer encoding and re-encoding? Second, do the number of forms produced correctly and the phonological precision of productions by preschool-age children with DLD decrease over periods of sleep more than children with TD demonstrating poorer consolidation? Third, do the number of forms produced correctly and the phonological precision of productions by preschool-age children with DLD decrease over a one-month delay more than children with TD demonstrating poorer long-term retention of learning?

Method

Participants

All reported protocols and recruitment methods were approved by the Institutional Review Board at Boys Town National Research Hospital. Methods of recruitment included: the Human Subject Research Core participant database, hearing and language screenings during public events (e.g., library story hour), local childcare providers, home school events, and other kindergarten readiness programs. Word of mouth, flyer distribution and social media campaigns supported additional recruitment. Parents and legal guardians gave written informed consent for their child to participate. Sessions were completed at a location that fit the family's needs: either the child's home, the child's daycare facility, a public place such as a local library, in the laboratory of the first author, or a speech-language clinic operated by Boys Town National Research Hospital. When testing occurred in participants' homes, families were asked to create a distraction-free environment. When testing occurred in libraries or clinics, private rooms or other quiet spaces were used.

Participants included nine children with DLD (females = 5, males = 4) and nine children with TD between the ages of 4:0 and 6:11 (mean age in months = 59.06, sd = 8.40) Table 1 lists participant characteristics, and Appendix A includes additional standardized test results. Each child with TD was matched to a child with DLD on biological sex, age (within three months difference), and number of years of maternal education (mean difference = .78 years, range = 0 to 3 years difference). Children's racial/ethnic backgrounds were as follows: white/non-Hispanic = 14, biracial/non-Hispanic = 2, biracial/non-reported ethnicity = 1, information not provided = 1. The data was collected between May 2018 and December 2019 and all children resided within 30 miles distance from Omaha, Nebraska during data collection. The data from two additional children were excluded due to experimenter error.

To be included in the study children had to demonstrate: normal hearing via a pure-tone audiometric screening and typical non-verbal IQ by achieving a standardized score of 4 or greater (equivalent to a standardized score of 70) on the Wechsler Preschool and Primary Scale of Intelligence-IV block design and matrix reasoning subtests (WPPSI-IV; Wechsler, 2012). To pass the hearing screening, responses to pure tones presented at 20 dB HL were required in both ears for frequencies of 1-4k Hz. All children spoke Mainstream American English with no reported exposure to a second language. All children lacked neurological or other developmental disorders based on parental report.

Additionally, to be included children had to demonstrate the ability to produce the majority of phonemes that made up the target forms included in the current study on the Goldman-Fristoe Test of Articulation-3 (GFTA-3; Goldman, 2015). For children with DLD, an average of 1.44 (sd = 1.51) of the target forms included phonemes that were not produced correctly on the GFTA-3. For children with TD, an average of 1.11 (sd = 0.93) target forms included phonemes that were not produced correctly on the GFTA-3. However, in many cases these forms resolved in that they were produced with 100% accuracy sometime during training (see Supplemental Materials, S1). Overall, the children with TD and DLD demonstrated a similar number of items missed on the GFTA-3 that related to target forms and a similar number of resolved forms. Thus, we elected to code each production throughout training and the long-term test based on the phonemes the child produced rather than making adjustments based on missed GFTA-3 items.

DLD status was determined via standardized scores on the Structured Photographic Expressive Language Test – 3 (SPELT-3; Dawson, Stout, & Eyer, 2003) given the high sensitivity and specificity of this measure (Perona, Plante, & Vance, 2005). Children who received a standardized score of 95 or below were included in the DLD group and children who received a standardized score of 96 or above were included in the TD group. Children completed the Peabody Picture Vocabulary Test – 4 (PPVT-4; Dunn & Dunn, 2007) to assess receptive vocabulary abilities and a non-word repetition test [NWR, (Dollaghan & Campbell, 1998)] to assess phonological working memory abilities (Table 1). Children with and without DLD did not differ significantly in age, maternal education, articulation abilities (GFTA-3), or the two subtests from the WPPSI-IV. Children with DLD demonstrated significantly lower performance than children with TD on the SPELT, t (15.11) = 10.30, $p < .001$, and on the non-word repetition test, t (15.97) = 5.90, $p < .001$. Notably, the scores from the two groups did not overlap on the SPELT or the NWR test. Children with DLD demonstrated a lower mean score on the PPVT, but this difference was not statistically significant, t (14.11) = 1.62, p = 0.13, and the scores overlapped substantially.

Stimuli

Stimuli included nine forms, three one-syllable and six two-syllable forms, created in the laboratory (Appendix B). The two-syllable forms contained a syllable structure of CV.CVC or CCV.CVC. All forms were composed primarily of early acquired sounds (McLeod & Crowe, 2018). All forms within the set varied in initial consonant or consonant cluster. Each form had an accompanying minimal pair that varied from the target in the final consonant for the one-syllable forms. For the two-syllable forms, three minimal pairs varied from the target in the medial consonant and three varied in the final consonant. Each form was paired with one of nine unfamiliar objects (Appendix B). Throughout training and testing, children never produced the real name for any of the objects demonstrating that they were unnamable to the children.

Assessments of Learning

We assessed children's representations of forms throughout training and after the long-term delay via three primary tasks. These include: a Free Recall Task, a Cued Recall Task, and a four alternative forced choice (4AFC) Recognition Task. Each of these tasks served the dual purpose of assessing children's representations of forms and providing retrieval-based learning opportunities throughout training. Administering tasks that varied in retrieval demands provided sensitivity to assess group differences across training days and after delays. The Free and Cued Recall Tasks proved to be sensitive assessments of children's learning. Thus, responses to these tasks are the primary outcome variables in the current analyses. Children's responses to the Recognition Task (4AFC task) administered near the end of each training day did reveal differences across groups and training days (see Appendix C). However, this task is less sensitive to assess the specific phonological precision of children's representations of forms. Thus, we do not include these results in the primary analyses.

Free and Cued Recall Tasks

During each trial of the Free Recall Task, the child was shown one of the trained objects and asked, "What is this one called?". After a three to five second wait time, encouragement to guess was offered. If the child did not produce the correct form, the experimenter administered the Cued Recall production probe which included the CV or CCV onset of the target form. For example, if the child was shown the object that was labeled a /mep/ during training and she labeled the object a /bınıg/, she was told, "It starts with /me/ ….". If the child responded by finishing the target word, in this case /p/, the examiner prompted the child to produce the entire form.

We administered the Cued Recall prompt to aid children in retrieving their representation of the form. Free recall of phonological forms (i.e., naming tasks) are traditionally difficult for young children (see Gordon & McGregor, 2014). We anticipated that children would often fail to respond to this task. Providing a cue of the first CV or CCV of the target form allowed us to assess if the child had some memory of the form. Notably, if children produced the correct form in the Free Recall Task, there was no need to administer the cue as we already assessed their representation of the form. In this case, their response to the Free Recall prompt was coded as both their Free Recall and Cued Recall response for that object at that time point. Additionally, during Free Recall if a child produced the correct CV/CCV onset of the target form, but failed to produce the entire form correctly, the Cued Recall probe was not administered. The reasoning behind this choice was that providing the first CV/CCV of the target form is unlikely to induce further retrieval of the correct form. In this case, their response to the Free Recall prompt was coded as both their Free and Cued Recall response for that object at that time point.

Children were provided with performance-contingent feedback during the Free and Cued Recall Tasks at specific timepoints in the protocol (Appendix D). The feedback was given after the child had responded to both the Free Recall prompt and Cued Recall prompt, when it was administered. When the child produced the form correctly, the experimenter would respond, "Yes that's right, this is a /mep/." When the child produced the form incorrectly, the experimenter would respond, "Actually this is a /mep/." When the child stated a phonologically similar form (e.g., /meb/ for /mep/) the examiner's response was, "That's really close but this one is a /mep/."[2]

Recognition Tasks

Two recognition tasks were used to provide additional opportunities for children to retrieve forms and receive feedback to their responses. These included two and four alternative forced choice recognition tasks (2AFC Dot Task; 4AFC Dot Task). In a given trial of the 2AFC Dot Task, an object as well as a piece of paper with two large dots on it were placed in front of the child (see Appendix C). The experimenter pointed to the object and asked, "What is this one called?". The experimenter then presented the child with two forms, the target form for the object present and another trained form from the set, order randomized across trials. As she produced the first form the experimenter pointed to one of the black dots on the paper. As she produced the second form she pointed to the other black dot. Children could indicate their response by producing one of the forms, by pointing to one of the black dots, or by doing both. The 4AFC Dot Task was similar but each trial included four forms as options: the target form, a

[2] One form, /grɑmɚ/, was particularly difficult for children to say with 100% accuracy. Thus, we adjusted what we considered a correct production during training when giving feedback (see Supplemental Materials, S1).

minimal pair of the target, another trained form (i.e., the alternative form), and a minimal pair of the alternative form. Feedback to the 2AFC and 4AFC Dot Tasks were given in a manner similar to the Free and Cued Recall Tasks.

Protocol

Each training day included six distinct blocks (Appendix D). Children participated in stretches or walks for one-minute intervals between each block. On each day, objects were presented in the same order during each block. However, object order was randomized across training days. The first block included the Free Recall Task without feedback for all nine form-referent pairs. Any object that was named with 100% of phonemes produced correctly was not included in the training blocks for that day (see Supplemental Materials S2). In this way, the training time each day focused on form-referent pairs that the child had not yet learned.[3] During Training Block 1, the child was shown each object individually, the experimenter labeled the object two times, and prompted the child to repeat the form. During Training Blocks 2 and 3, children completed the 2AFC Dot Task. During these blocks, each form served once as the target and once as the alternative form. Training Block 3 mirrored Training Block 2 with the exception that each target form was paired with a different distractor than Block 2.

Upon completion of the training blocks, children participated in two additional Testing Blocks in which all nine form-referent pairs were tested regardless of whether children named them correctly at the beginning of the session. During Testing Block 2, they completed the 4AFC Dot Task. Each form served once as the target and once as the alternative form. Children took a

[3] The original purpose of this study was to train all children to criterion on all nine words, without overtraining words they had successfully learned. This would allow us to assess children's ability to retain words over a post-training delay after training all words to the same level. However, not all children reached criterion after six subsequent training days. Regardless, children within and across groups demonstrated a similar ability to retain learning relative to their end of training performance. We realized the more interesting differences across groups resided in differences in encoding during training days and consolidation between training days.

one-minute break to stretch or walk between every three trials during this testing block to avoid fatigue. During the final testing block, the examiner administered the Free and Cued Recall probes for each object. As stated previously, the Cued Recall probe immediately followed the Free Recall probe for a specific object if the child did not produce the onset CV/CCV of the target form correctly.

The first training day followed a similar structure as the other training days with the following key differences. Testing Block 1 was not administered as children had not yet learned any of the form-referent pairs. Pilot testing revealed that presenting all nine pairs one after the other on the first day contributed to particularly poor performance on target tasks. Thus, form-referent pairs were presented in three sets of three. The first set were introduced via all training and testing blocks included in Appendix D (excluding Testing Block 1). During Training Block 2, each form was presented with a familiar noun (e.g., Is it a /plun/ or a flower?) to familiarize children with the 2AFC Dot Task. Children subsequently completing all training and testing blocks with the second and third set. After the three sets had been presented, children completed Testing Blocks 2 and 3 with all three sets intermixed.

Children completed training days until they reached criterion or completed a total of six training days. All children completed all training days consecutively without extra days between any of the sessions. To achieve criterion, a child had to demonstrate robust learning of all nine target forms and the link between each form and its referent at the end of a given training day. Thus, for each object the child had to respond correctly in the Free Recall Task or had to respond correctly in the Cued Recall Task and respond correctly for that same form in the 4AFC Dot Task, demonstrating a memory for both the form and the link for that object.

All children completed testing one-month after their last training day (range 27 to 30 days, mean = 28.17 days). During this session, words were tested in three sets of three (see below). Each set included three testing blocks with one-minute stretch breaks between each block. During the first block, the examiner presented each object one at a time and administered a Free Recall prompt immediately followed by a Cued Recall prompt if needed. No feedback was given. During the second block, the examiner administered the 4AFC Dot Task without feedback. The third testing block was identical to the first testing block however children were provided with feedback to their responses, similar to the feedback given during training. The 4AFC block and additional Free/Cued Recall testing block were added in case children demonstrated poor retrieval of target forms. However, the initial Free and Cued Recall prompts proved to be sensitive to assessing children's memory for forms after the one-month delay. Thus, only the first Free and Cued Recall prompts administered for each word were included in analyses.

Between the last training day and the one-month session we tested subsets of the forms at various delay intervals to determine whether a brief testing session aided long-term retention of forms. For each child, three forms were tested one week after the last training day (Set A), three forms were tested two weeks after the last training day (Set B), and three forms were tested at the beginning of the one-month session (Set C). These additional testing sessions followed the same protocol as the one-month testing session. Information about how forms were assigned to each testing timepoint is included in Supplemental Materials S2. We report results of this manipulation in the analyses below. However, this manipulation did not significantly relate to children's ability to retrieve and produce forms at the one-month delay. Thus, we do not discuss testing at these various timepoints in the introduction or discussion.

Scoring

Three research assistants formally trained in IPA transcription coded children's verbal responses in the Free and Cued Recall tasks based on high-quality video and audio recordings. Each production was independently transcribed by two research assistants before they compared their transcriptions. Intercoder reliability was 91%. Discrepancies were resolved by watching the video together to determine an accurate transcription. The third author is a certified speech-language pathologist and she aided with resolving disagreements when needed.

Each production yielded two scores: a phonological precision score and a whole word score. For the phonological precision score, the phonemes that children produced were compared to the target form (Edwards, Beckman, & Munson, 2004). Consonants were assigned a value up to three points for the correct production of manner, voicing, and place features. Vowels were assigned a value up to three points for correct production of height, backness, and tenseness features. Each production was coded as the total percentage of phonetic features produced correctly relative to the target form. Target forms were coded as zero features correct when the child's production more closely approximated another form in the set. For the whole word score, only productions with 100% of the phonetic features produced correctly were coded as correct. All other productions were coded as incorrect. See Supplemental Materials S1 for additional information about coding.

Analyses and Results

In all sections below we fit two mixed effects models: a model with the log odds of a correct response as the outcome variable using the whole word score and a model with the phonological precision of productions as the outcome variable using the phonological precision score. For the whole word score we used the log odds to account for a lack of homogeneity of variance inherent in analyses with binary responses (see Gordon, 2019). All analyses were conducted in an R environment using the lme4 package.

Encoding During Training

To determine children's ability to encode forms from input, we conducted a generalized mixed effects model with the log odds of a correct response to the Free and Cued Recall probes administered at the end of each training day as the outcome variable. The fixed effects of the maximal model included: group (DLD, TD), cue (free recall, cued recall), day, age (in months), and biological sex. We included a day by group interaction to determine whether participants with TD demonstrated a faster learning rate across days than participants with DLD. We also included a cue by group interaction to determine if either children with DLD or TD demonstrated a greater benefit from cuing. Mixed effects models include random effects to control for systematic variation in the data when identifying the relationships between the fixed effects and outcome variable (see Gordon, 2019). For this model, we included intercepts for participant and form as well as participant by day and form by day slopes as random effects. The random effect structure that best supported model fit included intercepts for participant and form and a participant by day slope. The minimal fixed effect structure that best supported model fit included group, cue, day, and a cue by group interaction (Appendix E, Table 1). Descriptive statistics of the number of words produced correctly at the end of each training day for children with TD and DLD are listed in Table 2.

Children with TD had a higher probability of producing the correct form than children with DLD (B = -1.23, $z = -2.63$, $p < 0.01$). Additionally, children had a higher probability of producing the correct form in the Cued as opposed to Free Recall Task (B = 1.67, $z = 9.52$, $p < 0.001$). For example, on Day 1 the TD group had an average probability of 12% of producing the form correctly in the Free Recall Task, and 41% of producing the form correctly in the Cued Recall Task. Thus, they produced a model based[4] average of 1.04 and 3.69 forms correctly in the Free and Cued Recall Tasks, respectively. In contrast, the DLD group had an average probability of 4% of producing the form correctly in the Free Recall Task, and 9% of producing the form correctly in the Cued Recall Task. Thus, they produced a model-based average of 0.33 and 0.78 forms correctly in the Free and Cued Recall Tasks, respectively. These averages reveal the nature of the cue by group interaction. Specifically, when given the cue children with TD increased the probability of producing forms correctly by 29%, the equivalent of 2.65 additional forms. When given the cue, the children with DLD increased the probability of producing forms correctly by 5%, the equivalent of .45 additional forms. Performance improved across days (B = 0.61, z = 7.32, $p < 0.001$). The group by day interaction did not improve model fit, thus, it was excluded from the final model. This indicates that the probability of producing the correct form increased to the same degree per training day across groups, an average increase of .72 forms per day. Children's individual performance during training is included in Supplemental Materials, S4.

To determine changes in the phonological precision of productions across training days, we conducted a linear mixed effects model. For this model, the percentage of phonological features produced correctly in response to Free and Cued Recall probes administered at the end of each training day was the outcome variable. This model had the same fixed and random effects as the previous model. The random effect structure that best supported model fit included intercepts for participant and form and a participant by day slope. The minimal fixed effect structure that best supported model fit included group, cue, training day, biological sex, and a

[4] These averages are derived from the final model in which systematic differences across participants and forms are accounted for via the random effects. Descriptive statistics of means and standard deviations are included in Table 2.

group by cue interaction (Appendix E, Table 2). Descriptive statistics of phonological precision of productions at the end of each training day for children with TD and DLD are listed in Table 3.

Productions by children with DLD were, on average, 11% less precise than productions by children with TD (B = -0.11, $t = -2.13$, $p = 0.05$). Children produced more phonological features correctly in the Cued as opposed to Free Recall Task (B = 0.44, $t = 18.09$, $p < .001$). The group by cue interaction revealed that children with TD benefited more from the cue than children with DLD (B = -0.08, $t = -2.45$, $p = 0.01$). Specifically, on Day 1 children with TD produced an average of 26% features correct in the Free Recall Task, and 71% features correct in the Cued Recall Task, reflecting an average increase of 45%. On Day 1 children with DLD produced an average of 15% features correct in the Free Recall Task, and 51% of features correct in the Cued Recall Task, reflecting an average increase of 36%. Examination of the interaction of cue revealed that children with DLD were more likely not to respond to the Cued Recall prompt than children with TD. Across all training days, children with TD were cued a total of 246 times (mean per child = 27.22, sd = 11.05). In 23 instances, or 9% of trials, children with TD did not produce a response. Across all training days, children with DLD were cued a total of 335 times (mean per child = 37.11, sd = 10.89). In 76 instances, or 22% of trials, children with DLD did not produce a response. When children did respond, they were highly accurate at producing the part of the word that was cued. Children with TD and DLD produced a respective average of 98% (sd = .08) and 98% (sd = .07) phonetic features of the first CV/CCV correctly.

Performance improved across days (B = 0.08, $t = 7.57$, $p < .001$) with an average increase in phonological precision of 8% from one day to the next. Once again, the group by day interaction did not improve model fit, thus, it was eliminated from the final model. This indicates that across groups, children improved the phonological precision of productions to the same degree from one day to the next.

We ran two additional models to assess how children's NWR score (i.e., percentage of phonemes correct, Dollaghan & Campbell, 1998) and PPVT score were related to phonological precision of productions. The variables NWR score, PPVT score, and Group (DLD, TD) cannot be included in the same model due to multicollinearity. However, NWR score and PPVT score were significantly related to the phonological precision of children's productions, NWR (B = .003, t = 2.13, p = .05) and PPVT (B = .004, t = 2.43, p = .03) (Supplemental Materials, S3).

Consolidation During Training

We conducted a generalized mixed effects model to determine whether there was a significant change in the probability of a correct production at the end of one training session and at the beginning of the following training session and whether this differed across groups. This model only included responses to the Free Recall Task, as Cued Recall was not administered at the beginning of training sessions. The fixed effects of the maximal model included: assessment time (end of session, beginning of session), gap (1 to 5), group (DLD, TD), age (in months), and biological sex. The variable Gap is similar to the variable Day from the previous analyses (See Figure 1). For each child, there was up to six training days. Similarly, for each child there was up to five gaps between training days. Each gap (e.g., Gap 1) included performance at the end of a specific training day (e.g., Day 1) and performance at the beginning of the following training day (e.g., Day 2).[5] We included an assessment time by group interaction as a fixed effect to

[5] Scores included in Gap 1 assesses learning that occurred after the input from Training Day 1, but before input from training Day 2. Performance at the end of Training Day 1 are coded as Gap 1 and coded as the assessment time, end of session. Performance at the beginning of Training Day 2 are coded as Gap 1 and coded as the assessment time, beginning of session. Thus, the effect of Time allows us to assess how performance changed from the end of one session to the beginning of the next session. The effect of Gap allows us to assess changes across the five Gaps based on how much training the child had received at that point.

determine whether children with or without DLD demonstrated poorer consolidation of learning over a period of overnight sleep.

The random effects structure that best supported model fit included intercepts for participant and form. The minimal fixed effects structure that best supported model fit included fixed effects for gap and group (Appendix E, Table 3). Participants increased the probability of a correct response from one gap to the next (B = .66, $z = 10.38$, $p < .001$). This reflects an average increase of 4% probability of a correct response from one gap to the next. Also, participants with TD had a higher probability of producing forms correctly than participants with DLD (B = -1.33, $z = -2.10$, $p = .04$). Notably, including a fixed effect for assessment time did not improve model fit, indicating that participants' performance at the end of a training day and the beginning of the following training day did not differ significantly.

We conducted a similar model to determine whether there was a significant difference between the percentage of phonetic features produced correctly at the end of one training day and at the beginning of the following training day and whether this differed across children with DLD and TD. This model had the same fixed and random effects as the previous model. The random effects structure that supported model fit included intercepts for participant and form and a participant by gap slope. The minimal fixed effects structure that best supported model fit included fixed effects for gap and age in months (Appendix E, Table 4). Participants increased the percentage of phonetic features produced correctly from one gap to the next (B = .10, $t = 7.39$, $p < .001$). Including a fixed effect for assessment time did not improve model fit, indicating that the phonological precision of productions at the end of a training day and the beginning of the following training day did not differ significantly.

Post-Training Retention

Four children with TD and one child with DLD achieved criterion with all nine forms by the end of the training. Children with TD demonstrated criterion on an average of 7.56 (1.42) forms and children with DLD demonstrated criterion on 4.89 (2.57) forms. To assess retention of learning, we conducted a generalized mixed effects model with the log odds of a correct response as the outcome variable to determine changes in the probability of producing the correct form at the end of the last training day and the one-month delay. The fixed effects of the maximal model included: group (DLD, TD), cue (free recall, cued recall), age (in months), biological sex, delay (end of training, one-month delay) and a fixed effect for retest (retested after a one-week delay, a two-week delay, or not retested before the one-month test). We also included a delay by cue by group interaction to determine if the cue differentially affected children with and without DLD based on the delay (end of training, one-month delay). We included intercepts for participant and form as random effects. These random effects were retained as they best supported model fit. The minimal fixed effect structure that best supported model fit included delay, cue, and group (Appendix E, Table 5).

Participants had a higher probability of producing a form correctly at the end of the last training day (average probability = .64, 5.76 forms) than after the one-month delay (average probability = .48, 4.32 forms). Participants with TD had a higher probability of producing forms correctly than participants with DLD (B = -1.49, $z = -2.09$, $p = .04$). Additionally, participants had a higher probability of producing forms correctly in the Cued as opposed to Free Recall Task (B = 1.46, $z = 7.14$, $p < .001$). Specifically, at the end of training children with TD produced an average of 5.22 (2.11) forms and 7.67 (1.22) forms correctly in the Free and Cued Recall Tasks, respectively. After the one-month delay, they produced an average of 3.67 (2.45) forms and 6.78 (2.28) forms correctly in the Free and Cued Recall Tasks, respectively. At the end of training, children with DLD produced an average of 3.44 (2.35) forms and 5.00 (2.60) forms correctly in the Free and Cued Recall Tasks, respectively. After the one-month delay, they produced an average of 2.67 (1.94) forms and 4.33 (3.39) forms correctly in the Free and Cued Recall Tasks, respectively. There was not a significant delay by group interaction, indicating that the probability of a correct response decreased to a similar degree over the one-month delay for both groups. Including a fixed effect for retest did not improve model fit, thus, it was eliminated from the final model.

We conducted a linear mixed effects model to determine the percentage of phonological features produced correctly at the end of training and after the one-month delay. This model had the same fixed and random effects as the previous model. The random effect structure that best supported model fit included intercepts for participant and form. The minimal fixed effect structure that best supported model fit included group and cue (Appendix E, Table 6). Participants with TD produced more phonological features correctly than participants with DLD (B = -0.16, t = -2.13, p = .05). Additionally, participants produced forms with more phonological precision in the Cued as opposed to Free Recall Task (B = 0.35, t = 12.33, p < .001). Notably, there was not a significant effect of delay indicating that the phonological precision of productions did not differ significantly between the end of training performance and after the one month delay. Including a fixed effect for retest did not improve model fit, thus, it was eliminated from the final model.

Discussion

In the current study, our research questions were the following. First, do preschool-age children with DLD produce fewer forms correctly and produce forms with less phonological precision during periods of input in comparison to children with TD demonstrating poorer encoding? Second, do the number of forms produced correctly and the phonological precision of productions by preschool-age children with DLD decrease over periods of sleep more than children with TD demonstrating poorer consolidation? Third, do the number of forms produced correctly and the phonological precision of productions by preschool-age children with DLD decrease over a one-month delay more than children with TD demonstrating poorer long-term retention of learning? Overall, encoding was a relative weakness and consolidation was a relative strength for children with DLD. This was the case both for the number of forms produced and for the phonological precision of productions. Children with TD and DLD demonstrated robust long-term retention of learning. We discuss results and implications in more detail below.

Encoding and Phonological Precision

In the current study, preschool-age children with DLD demonstrated difficulty encoding forms from input relative to children with TD. Interestingly, children with DLD demonstrated similar gains in word learning across days as children with TD. These findings are consistent with Bishop et al. (2012). They presented forms multiple times to children and asked them to repeat the forms each time they heard them within a single session. Children with DLD demonstrated poorer productions of forms than children with TD. However, the degree to which their performance improved across trials did not differ across groups. Notably, children's performance in non-word repetition tasks, which capture their ability to encode an initial phonological representation of the form, are related to the rate of children's vocabulary growth during early childhood (see Gathercole, 2006 for a review). The results of Bishop et al., (2012) and the current study suggest that the phonological precision of children's initial representation of a form is what contributes to individual differences in the total amount of experience the child needs to add a form to her productive vocabulary. However, children's ability to refine a representation across additional experiences (i.e., re-encoding) is much less likely to contribute to individual differences in the rate of word learning.

An alternate explanation of the current findings and Bishop et al. (2012) is that children with DLD and TD do differ in their ability to refine representations of forms across trials/training sessions. However, because the initial representations of children with TD are more phonologically precise, they have less ability to refine those representations as they approach ceiling throughout training. In the current study and Bishop et al., the word set was limited to a specific number of word-referent pairs. However, in the real world, the number of words to learn is vast. Thus, children with DLD may demonstrate a slower rate of vocabulary growth than children with TD because they encode less phonological information overall during periods of input. Both encoding new forms and refining previously encoded forms may be impaired relative to children with TD.

Overall, there is robust evidence that encoding, which involves establishing an initial representation of a word form, is impaired in individuals with DLD relative to peers with TD. However, it is unclear if the process of re-encoding, which involves refining a previously encoded representation, is impaired in individuals with DLD. Furthermore, if both encoding and re-encoding are impaired in individuals with DLD, the relative contribution of each of these processes to the slowed rate of word learning is unclear. The current study and Bishop et al., suggest that impaired encoding abilities has a larger effect on the rate of word learning for individuals with DLD than possible impaired re-encoding abilities. However, further research in which ceiling effects are carefully controlled are needed to gain a better understanding of the encoding and re-encoding abilities of children with DLD relative to children with TD.

Another finding from the current study is that children with TD demonstrated a greater improvement in their productions than children with DLD when they were given the first CV/CCV of the form as a cue. Children with TD were more likely to attempt a response when cued than children with DLD. The retrieval-based learning literature indicates that providing retrieval opportunities, even those that lead to wrong responses or no response, can support encoding and consolidation of target information (see Gordon, 2020). Notably, after the learner attempts to retrieve the target information, she becomes aware of what she does or does not know and is likely to attend to the target information when the correct response is given as feedback. Additionally, providing retrieval opportunities that vary in retrieval support (e.g., Free Recall, Cued Recall, Recognition) can enhance encoding and consolidation more than providing only one type of retrieval opportunity (Adesope et al., 2017). In the current study when the child was provided with the cue of the first CV/CCV, this may have benefited learning as it contributed to more successful retrievals than Free Recall prompts. Conversely, children may have put less effort into retrieving information in response to Free Recall prompts because they knew that they would receive the cue. Gordon et al., (2020) found that adults with DLD benefited more from Free Recall prompts that were not followed by a cue than Cued Recall prompts when learning word forms. Through further research, we can better understand the blend of retrieval tasks that contribute to the best learning and retention of forms for preschool-age children with DLD.

It is important to note that similar to children with TD, children with DLD varied in their verbal working memory abilities as measured by the NWR task (Dollaghan and Campbell, 1998). Working memory abilities and receptive vocabulary knowledge, as measured by the PPVT-4, were related to variability in children's ability to learn forms from input. The best learner with DLD achieved criterion, in that he demonstrated robust knowledge of all nine forms, on the fifth training day. The performance of this learner was similar to that of children with TD. The learner with DLD who struggled the most produced one and two forms correctly in the Free and Cued Recall Tasks respectively at the end of the last training day. Although verbal working memory is a common deficit for children with DLD (see Alloway & Gathercole, 2006; Montgomery, Magimairaj, & Finney, 2010 for reviews), not all children with DLD demonstrate impaired working memory (Gray et al., 2019). Furthermore, the specific nature and severity of this impairment varies across children with DLD. Additionally, children with DLD vary substantially in their receptive vocabulary knowledge, and often demonstrate overlapping receptive vocabulary scores with children with TD (Kan & Windsor, 2010). Thus, in line with Gray and colleagues (2019), we conclude that to support learning in an individual child with DLD, it is important to understand that child’s specific working memory and vocabulary knowledge profile.

Consolidation and Long-Term Retention

Results from the current study demonstrate that consolidation is an area of relative strength for children with DLD when learning words. This is the case for the number of forms children produced before and after a period of overnight sleep, and also for the phonological precision at which they produced forms. These results coincide with the results from McGregor and colleagues with college-age students with DLD (McGregor, Arbisi-Kelm, et al., 2017; McGregor et al., 2020; McGregor, Gordon, et al., 2017) and with the results from Leonard and colleagues with preschool-age children with DLD (see Leonard & Deevy, 2020 for a review). These results stand in contrast to Malins et al., (2020) and Rice et al., (1994) in which children with DLD demonstrated poorer consolidation of word learning than children with TD. Notably in

Malins et al.,(2020) children were taught words that contained a high degree of phonological similarity across the set. Given that children with DLD are likely to encode less phonologically precise representations of forms, these representations may not have been robust enough to support consolidation and fine-grained distinctions the following day. Furthermore, during training children were shown two potential referents and were asked to find the target referent when given a specific form (e.g., Find the pibu). Children were provided with direct feedback after their selection. However, having two referents present when they heard the form may have contributed to difficulty building a stable representation of the form-referent link. In Rice et al., (1994) children attempted to learn words via incidental exposures, which can be particularly challenging for children with DLD (Steele & Mills, 2011). Thus, children with DLD may have had more difficulty building representations that were robust enough to support consolidation of learning.

In the current study, we selected training that was most likely to support encoding, consolidation, and long-term retention of words. Children were given direct instruction via retrieval-based practice in which they were repeatedly asked to retrieve forms and were given feedback to their responses. Additionally, forms within the set were phonologically distinct from one another to support learning across the set. In a similar manner, McGregor and colleagues as well as Leonard and colleagues provided highly supportive training to individuals with DLD in that they both utilized retrieval-based practice. One limitation of the current study is that children with DLD and TD may have differed in their ability to consolidate forms, but we would need a larger sample size to capture this difference. We note that the current measures were sensitive enough to demonstrate differences in encoding given the current sample size. Thus, if differences in consolidation abilities across groups exists, the effect of consolidation is smaller than the effect of encoding when children are given highly supportive training. Overall, further research is needed to determine how the level of support during training contributes to differences in encoding and consolidation across children with and without DLD. This would lead to a better understanding of differences across groups when training is less supportive, such as through incidental exposures to target words when multiple referents are present, and when training is highly supportive, such as during robust vocabulary instruction in educational and clinical contexts. Additionally, as noted previously researchers typically include one training session and one session after a delay to assess consolidation and retention processes. However, robust learning is built across multiple sessions of input with multiple intervening periods of consolidation. Thus, it is important to understand how different levels of support during training affect changes of representations of words across these interleaving periods of input and consolidation (Simon et al., 2020)

In the current study we used some principles of robust vocabulary instruction (McKeown, 2019) such as providing explicit information about words, providing opportunities to actively retrieve encoded information, and targeting the same words across multiple sessions. However, a key aspect of robust vocabulary instruction includes teaching words in meaningful contexts given the semantic meanings of words. To focus on children's ability to learn forms, we constrained the semantic meaning of the words. We trained children with concrete objects and did not provide additional semantic information such as object function. During real-world word learning, children are gradually building rich semantic representations of word meanings while they are refining their phonological representations of forms. On the one hand, phonological and semantic representations may compete with each other in that children can only encode a limited amount of information during an exposure to a form-referent pair (Crystal, 1987). On the other hand, learning semantic information may support the learning of phonological information and vice versa (see Benham & Goffman, 2020). For example, children with DLD demonstrate better semantic knowledge for referents that they can more readily name (McGregor, Newman, Reilly, & Capone, 2002). Additional research is needed to understand how semantic representations (i.e., word meanings) and phonological representations (i.e., word forms) interact throughout the slow-mapping process.

An additional limitation of the current study is that we excluded training trials when children retrieved and produced forms correctly at the beginning of a training session. Because learning forms to the level that they can be retrieved and produced accurately is difficult for children with TD and DLD, we interpreted correct productions at the beginning of a training session as an indication that children had learned these forms. Overall, this was the case. When children retrieved and produced a form at the beginning of a session, they tended to retrieve and produce that form correctly at the beginning of subsequent sessions (Supplemental Materials, S2). Both children with and without DLD received the same training in that additional training trials were consistently excluded for forms they retrieved and produced correctly at the beginning of sessions. However, future research in which children with DLD and TD receive the same dosage to all words regardless of performance would illuminate whether receiving additional exposures to words children can retrieve and produce correctly contributes to differential effects on consolidation and long-term retention.

Implications for Educational and Clinical Practice

Overall, results from the current study and past research demonstrate that children with DLD are likely to successfully consolidate and later retrieve encoded information when provided with highly supportive training. In fact, principles of robust vocabulary instruction are likely to benefit both children with and without DLD (see Beck, McKeown, & Kucan, 2016; Gordon, 2020). Alarmingly, information about how to implement robust vocabulary instruction is often missing from Pre-K curricula (Neuman & Dwyer, 2009). This is reflected in educational practice. In an assessment of vocabulary instruction in Kindergarten classrooms, Wright and Neuman (2014) found that instructors tend to select target words incidentally. Instructors provided explicit definitions of words when they appeared in storybooks or a student asked about a word. When words are targeted incidentally, poorer word learners may not receive supportive enough training or enough exposures to the target word during a specific lesson to build representations that can be consolidated and retrieved during the following lesson.

Another danger of not implementing systematic robust vocabulary instruction is that poorer word learners may not receive an adequate total dosage to target words (i.e., cumulative intensity) to add words to their vocabularies. For example, preschool-age children with DLD demonstrate a better response to a vocabulary intervention when they were given 36 total exposures to target words as opposed to 12 or 24 exposures (Storkel et al., 2016). Critically, as demonstrated by the current study, children with DLD require more exposures to target words than children with TD to develop precise phonological representations of the words. For children with DLD, classroom instruction can be supplemented by an SLP who provides additional supportive exposures to target words during therapy sessions. Additionally, supporting word learning in children with DLD is particularly effective when SLPs and teachers partner together (Throneburg, Calvert, Sturm, Paramboukas, & Paul, 2000) and when SLPs involve parents (Marulis & Neuman, 2013). These partnerships include the benefits of increasing the exposures that children receive as well as providing opportunities for one-on-one instruction that can be tailored to the child's word learning abilities.

Children with DLD are likely to enter formal schooling with poorer vocabulary knowledge than their peers with TD. This poorer vocabulary knowledge affects the child's ability to succeed academically and to build essential literacy skills. Effective word learning interventions during preschool could improve children's access to academic content, set them on a path for better literacy development, and increase the likelihood that they will learn additional words from verbal and written content. Word learning interventions for children with DLD are only effective if they support encoding, consolidation, and long-term retention of learning. Children with DLD must be provided with supportive training during sessions and adequate exposures across sessions to foster learning that is likely to be retained long-term. Through partnerships between educators, clinicians, and parents we can achieve the important goal of improving long-term trajectories for children with DLD.

Acknowledgements This research was supported by the National Institute on Deafness and Other Communication Disorders Grant R21 5R21DC016130 awarded to Katherine Gordon and the National Institute of General Medical Sciences Grant P20 GM109023 awarded to Lori Leibold. Sara Kneifl, Kabao Lor, and Sandra McCoy transcribed and scored children's productions and were integral to this project.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). *Rethinking the Use of Tests : A Meta-Analysis of Practice Testing*. *XX*(X), 1–43. https://doi.org/10.3102/0034654316689306

Alloway, T. P., & Gathercole, S. E. (2006). *Working memory and neurodevelopmental disorders*. Psychology Press.

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

Archibald, L. M. D. (2017). Working memory and language learning: A review. *Child Language Teaching and Therapy*, *33*(1), 5–17. https://doi.org/10.1177/0265659016654206

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*, 675–693. https://doi.org/10.1080/13682820500442602

Beck, I. L., McKeown, M. G., & Kucan, L. (2016). Bringing words to life. In *Bringing words to life*.

Benham, S., & Goffman, L. (2020). Lexical – Semantic Cues Induce Sound Pattern Stability in Children With Developmental Language Disorder. *Journal of Speech Language & Hearing Research*, 1–18.

Bishop, D. V. M., Barry, J. G., & Hardiman, M. J. (2012). Delayed retention of new word-forms is better in children than adults regardless of language ability: A factorial two-way study. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037326

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of

CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *10*, 1068–1080. https://doi.org/10.1111/jcpp.12721

Crystal, D. (1987). Towards a “bucket” theory of language disability: Taking account of interaction between linguistic levels. *Clinical Linguistics and Phonetics*, *1*(1), 7–22.

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

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

Dollaghan, C. A., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136–1146.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words: Research report. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

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

Edwards, J., Beckman, M. E., & Munson, B. (2004). Vocabulary size and phonotactic production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, *47*(2), 421–436. https://doi.org/10.1044/1092-4388(2004/034)

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

Goldman, R. (2015). *GFTA-3: Goldman-Fristoe test of articulation 3*.

Gordon, K. R. (2019). How mixed-effects modeling can advance our understanding of learning and memory and improve clinical and educational practice. *Journal of Speech, Language, and Hearing Research*, *62*(3), 507–524. https://doi.org/10.1044/2018_JSLHR-L-ASTM-18-0240

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*, 1–11.

Gordon, K. R., & McGregor, K. K. (2014). A spatially supported forced-choice recognition test reveals children's long-term memory for newly learned word forms. *Frontiers in Psychology*, *5*(MAR), 1–12. https://doi.org/10.3389/fpsyg.2014.00164

Gordon, K. R., Mcgregor, K. K., & Arbisi-Kelm, T. (2020). Optimising word learning in post-secondary students with Developmental Language Disorder : The roles of retrieval difficulty and retrieval success during training Optimising word learning in post-secondary students with and retrieval success during traini. *International Journal of Speech-Language Pathology*, *0*(0), 1–14. https://doi.org/10.1080/17549507.2020.1812719

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children wtih and without specific language impairment: A meta-analysis. *Journal of Speech, Language & Hearing Research*, *50*(February), 177–196.

Gray, S. (2005). Word Learning by Preschoolers With Specific Language Impairment. *Journal of Speech Language and Hearing Research*, *48*(6), 1452. https://doi.org/10.1044/1092-4388(2005/101)

Gray, S., Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., & Cowan, N. (2019).

Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674–687. https://doi.org/10.1111/j.1467-7687.2012.01172.x

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/1525O(xH)701795598

James, E., Gaskell, M. G., Weighall, A., & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? In *Neuroscience and Biobehavioral Reviews* (Vol. 77). https://doi.org/10.1016/j.neubiorev.2017.01.054

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

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

Malins, J. G., Landi, N., Ryherd, K., Frijters, J. C., Magnuson, J. S., Rueckl, J. G., … Morris, R. (2020). Is that a pibu or a pibo? Children with reading and language deficits show difficulties in learning and overnight consolidation of phonologically similar pseudowords. *Developmental Science*, (July), 1–17. https://doi.org/10.1111/desc.13023

Marulis, L. M., & Neuman, S. B. (2013). How Vocabulary Interventions Affect Young Children at Risk: A Meta-Analytic Review. *Journal of Research on Educational Effectiveness*, *6*(3), 223–262. https://doi.org/10.1080/19345747.2012.755591

McClelland, J. L., McNaughton, B. L., & O'Rilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419.

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*. *19*(1), 43–57. https://doi.org/10.1080/10937404.2015.1051611.INHALATION

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

McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research*, 1–15.

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why Words Are Hard for Adults With Developmental Language Impairments. *Journal of Speech Language and Hearing Research*, *56*(6), 1845. https://doi.org/10.1044/1092-4388(2013/12-0233)

McGregor, K. K., Newman, R. M., Reilly, R., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *45*(October), 998–1014.

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)

McKenzie, S., & Eichenbaum, H. (2011). Consolidation and Reconsolidation: Two Lives of Memories? *Neuron*, *71*(2), 224–233. https://doi.org/10.1016/j.neuron.2011.06.037

McKeown, M. G. (2019). Effective Vocabulary Instruction Fosters Knowing Words, Using Words, and Understanding How Words Work. *Language, Speech, and Hearing Services in Schools*, *50*(4), 466–476. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0126

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, *27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, *19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, *3*(FEB), 1–10. https://doi.org/10.3389/fpsyg.2012.00041

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience*, *10*(3), 224–234. https://doi.org/10.1038/nrn2590

Neuman, S. B., & Dwyer, J. (2009). Missing in Action: Vocabulary Instruction in Pre-K. *The Reading Teacher*, *62*(5), 384–392. https://doi.org/10.1598/rt.62.5.2

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

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

Perona, K., Plante, E., & Vance, R. (2005). Diagnostic accuracy of the structured photographic expressive language test: (SPELT-3). *Language, Speech, and Hearing Services in Schools*, *36*(2), 103–115. https://doi.org/10.1044/0161-1461(2005/010)

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, *37*(1), 106–122. https://doi.org/10.1044/jshr.3701.106

Simon, K. C., Gomez, R. L., & Nadel, L. (2020). Sleep's role in memory reconsolidation. *Current Opinion in Behavioral Sciences*, *33*, 132–137.

Steele, S. C., & Mills, M. T. (2011). Vocabulary intervention for school-age children with language impairment: A review of evidence and good practice. *Child Language Teaching and Therapy*, *27*(3), 354–370. https://doi.org/10.1177/0265659011412247

Storkel, H. L. (2015). Learning from input and memory evolution: points of vulnerability on a pathway to mastery in word learning. *International Journal of Speech-Language Pathology*, *17*(1), 1–12. https://doi.org/10.3109/17549507.2014.987818

Storkel, H. L., Komesidou, R., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying adequate progress and successful learning patterns. *Language, Speech, and Hearing Services in Schools*, *48*(2), 108–124. https://doi.org/10.1044/2017_LSHSS-16-0058

Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Rominea, R. S. (2016). Interactive Book Reading to Accelerate Word Learning by Kindergarten Children With Specific Language Impairment: Identifying an Adequate Intensity and Variation in Treatment Response. *American Journal of Speech-Language Pathology*, *25*(October), 1–15. https://doi.org/10.1044/2016

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development*, *6*(3), 179–183. https://doi.org/10.1080/15475441.2010.484412

Throneburg, R. N., Calvert, L. K., Sturm, J. J., Paramboukas, A. A., & Paul, P. J. (2000). A Comparison of Service Delivery Models: Effects on Curricular Vocabulary Skills in the School Setting. *American Journal of Speech-Language Pathology*, *9*(1), 10–20.

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 : JSLHR*, *40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Vlach, H. A. (2019). Learning to Remember Words: Memory Constraints as Double-Edged Sword Mechanisms of Language Development. *Child Development Perspectives*, *13*(3), 159–165. https://doi.org/10.1111/cdep.12337

Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence-fourth ediction* (4th ed.). San Antonio, TX: The Psychological Corporation.

Wright, T. S., & Neuman, S. B. (2014). Paucity and disparity in kindergarten oral vocabulary instruction. *Journal of Literacy Research*, *46*(3), 330–357. https://doi.org/10.1177/1086296X14551474

Figure 1: The memory processes involved in slow mapping.

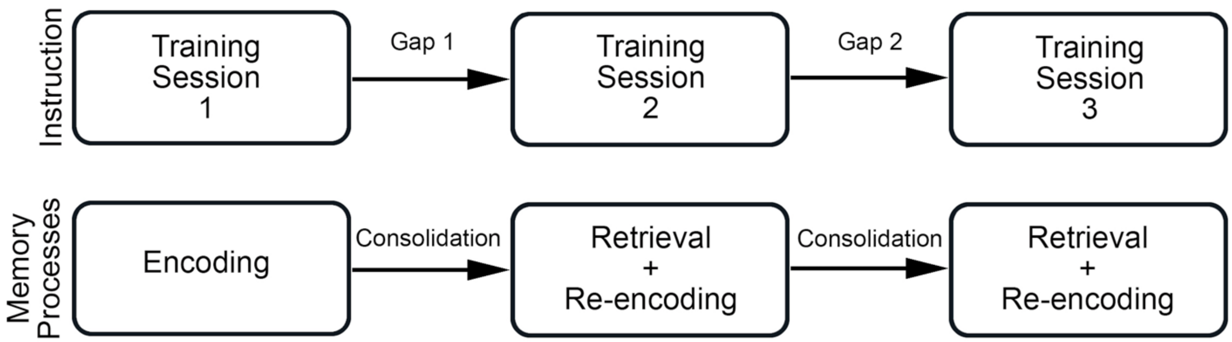


Table 1: Participant characteristics including mean standardized scores and standard deviations (sd). Minimum and maximum scores are listed in parentheses.

	TD	DLD	*t*-test
Age in Months	58.78 (50, 80) sd = 9.05	59.33 (50, 78) sd = 8.25	t (15.86) = -0.14, p = 0.89
Biological Sex	Females = 5 Males = 4	Females = 5 Males = 4	___
Maternal Education (in years)	16.11 (14, 18) sd = 1.45	16.22 (13, 20) sd = 2.17	t (13.99) = -0.13, p = 0.90
PPVT-4*	112.44 (83, 142) sd = 17.81	100.78 (88, 121) sd = 12.13	t (14.11) = 1.62, p = 0.13
SPELT-3*	120.22 (111, 130) sd = 6.61	83.33 (81, 95) sd = 8.47	t (15.11) = 10.30, p < .001
GFTA-3*	87.89 (59, 115) sd = 19.39	85.22 (64, 104) sd = 14.26	t (14.70) = 0.33, p = .74
Non-word repetition (percentage of phonemes correct)	75.35 (64.58, 88.54) sd = 10.35	47.11 (28.13, 60.42) sd = 9.93	t (15.97) = 5.90, p < .001
WPPSI-IV, Block Design*	9.89 (8, 13) sd = 1.76	8.78 (7, 11) sd = 1.48	t (15.54) = 1.45, p = 0.17
WPPSI-IV, Matrix Reasoning*	10.00 (7, 13) sd = 1.80	9.00 (7, 14) sd = 2.40	t (14.85) = 1.00, p = 0.33

*Scores are standardized scores.

Table 2: Descriptive statistics for the mean number of words produced with 100% phonological precision at each assessment point during training. Standard deviations in paratheses.

Children with Typical Development						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Beginning of Session, Free Recall		1.11 (0.93)	1.75 (1.28)	3.38 (2.20)	3.67 (3.27)	3.00 (1.41)
End of Session, Free Recall	1.00 (1.00)	2.22 (1.72)	3.00 (1.77)	4.63 (2.56)	3.67 (1.97)	4.20 (1.10)
End of Session, Cued Recall	2.56 (1.88)	6.11 (1.05)	6.38 (1.60)	7.13 (1.55)	6.17 (2.04)	6.80 (0.84)
Children with Developmental Language Disorder						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Beginning of Session, Free Recall		0.33 (0.50)	0.78 (1.30)	1.44 (1.51)	2.11 (2.62)	2.88 (2.36)
End of Session, Free Recall	0.33 (0.50)	1.22 (1.56)	1.33 (1.22)	2.67 (2.00)	3.00 (2.00)	3.25 (2.43)
End of Session, Cued Recall	0.67 (0.71)	2.22 (1.86)	2.67 (1.87)	3.56 (2.19)	4.89 (2.37)	4.50 (2.27)

Table 3: Descriptive statistics for the percentage of phonological features produced correctly at each assessment point during training. Standard deviations in paratheses.

Children with Typical Development						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Beginning of Session, Free Recall		0.18 (0.37)	0.32 (0.44)	0.46 (0.48)	0.50 (0.48)	0.49 (0.47)
End of Session, Free Recall	0.17 (0.36)	0.32 (0.45)	0.41 (0.48)	0.60 (0.48)	0.55 (0.47)	0.58 (47)
End of Session, Cued Recall	0.68 (0.34)	0.85 (0.30)	0.90 (0.23)	0.95 (0.12)	0.91 (0.18)	0.92 (0.20)
Children with Developmental Language Disorder						
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
Beginning of Session, Free Recall		0.08 (0.26)	0.18 (0.36)	0.34 (0.45)	0.31 (0.45)	0.43 (0.48)
End of Session, Free Recall	0.08 (0.25)	0.25 (0.41)	0.25 (0.41)	0.44 (0.48)	0.38 (0.48)	0.45 (0.48)
End of Session, Cued Recall	0.50 (0.32)	0.61 (0.37)	0.68 (0.35)	0.73 (0.34)	0.74 (0.36)	0.73 (0.37)

Appendix A. Additional standardized test scores for children with DLD and TD.

Table 1: Results from additional standardized tests administered to further characterize the current sample of children with DLD and TD. Minimum and maximum scores are listed in parentheses.

		TD	DLD	*t*-test
Automated Working Memory Assessment (AWMA)	Digit Recall	109.44 (87, 137) sd = 14.31	86.33 (77, 97) sd = 7.79	*t* (12.36) = 4.25, $p < .01$
	Dot Matrix	97.22 (75, 126) sd = 15.05	85.67 (70, 98) sd = 10.87	*t* (14.57) = 1.87, $p = 0.08$
	Counting Recall	100.89 (82, 130) sd = 15.36	85.22 (69, 97) sd = 12.03	*t* (15.13) = 2.41, $p = 0.03$
	Counting Recall, Processing Score	98.67 (80, 131) sd = 15.57	85.56 (82, 94) sd = 6.13	*t* (10.42) = 2.35, $p = 0.04$
	Odd-one-out	109.00 (90, 125) sd = 15.86	93.56 (85, 115) sd = 9.81	*t* (13.34) = 2.48, $p = 0.03$
	Odd-one-out Processing Score	106.44 (89, 124) sd = 14.76	93.33 (84, 122) sd = 11.72	*t* (15.22) = 2.09, $p = 0.05$
Diagnostic Evaluation of Language Variation (DELV)*		11.11 (7, 18) sd = 2.93	7.22 (5, 10) sd = 2.28	*t* (15.08) = 3.14, $p < .01$

*Normed referenced scoring

To further assess children's verbal and non-verbal working memory abilities, we administered four subtests of the Automated Working Memory Assessment (AWMA; Alloway, 2007): digit recall, dot matrix, counting recall, and odd-one-out. During digit recall children repeated strings of digits of increasing length to assess verbal short-term memory. The dot matrix subtest required children to view a red dot on a four-by-four grid and then indicate the dot's prior position on an empty grid to assess visuospatial short-term memory. We obtained two scores from the counting recall subtest in which children counted the number of red circles in an array of shapes and colors (percentage of correct responses coded as their processing score) and recalled the correct number of red circles (recall score) to assess verbal working memory. We also obtained two scores from the odd-one-out task in which children indicated the different shape in a set of three across a number of trials (percentage of correct responses coded as their processing score) and then indicated the former positions of these shapes on an empty grid (recall score) to assess visuospatial working memory.

Appendix B. Word-referent pairs

Table 1. Word forms and characteristics, phonotactic probability. Phonotactic probability was calculated with the English calculator (Vitevitch and Luce, 2004; https://calculator.ku.edu/phonotactic/about).

Target Form	Phonotactic Probability, Sum of Positional Segment Frequency	Phonotactic Probability, Biphone Probability	Minimal Pair	Phonotactic Probability, Sum of Positional Segment Frequency	Phonotactic Probability, Biphone Probability
/bɪnɪp/	0.3117	.0191	/bɪnɪg/	0.303	.0197
/dob/	0.1271	.0027	/dof/	0.1208	.0016
/grɑmɚ/	0.2104	.0186	/grɑtɚ/	0.2703	.0250
/kinɪt/	0.3488	.0093	/kigɪt/	0.2706	.0033
/mep/	0.1235	.0045	/mev/	0.11	.0043
/nedɪg/	0.1505	.0099	/nedɪp/	0.1592	.0093
/plun/	0.1905	.0092	/plub/	0.1617	.0088
/sibl̩/	0.183	.0061	/sifl̩/	0.1767	.0041
/topɪn/	0.2547	.0078	/topɪf/	0.1929	.0058

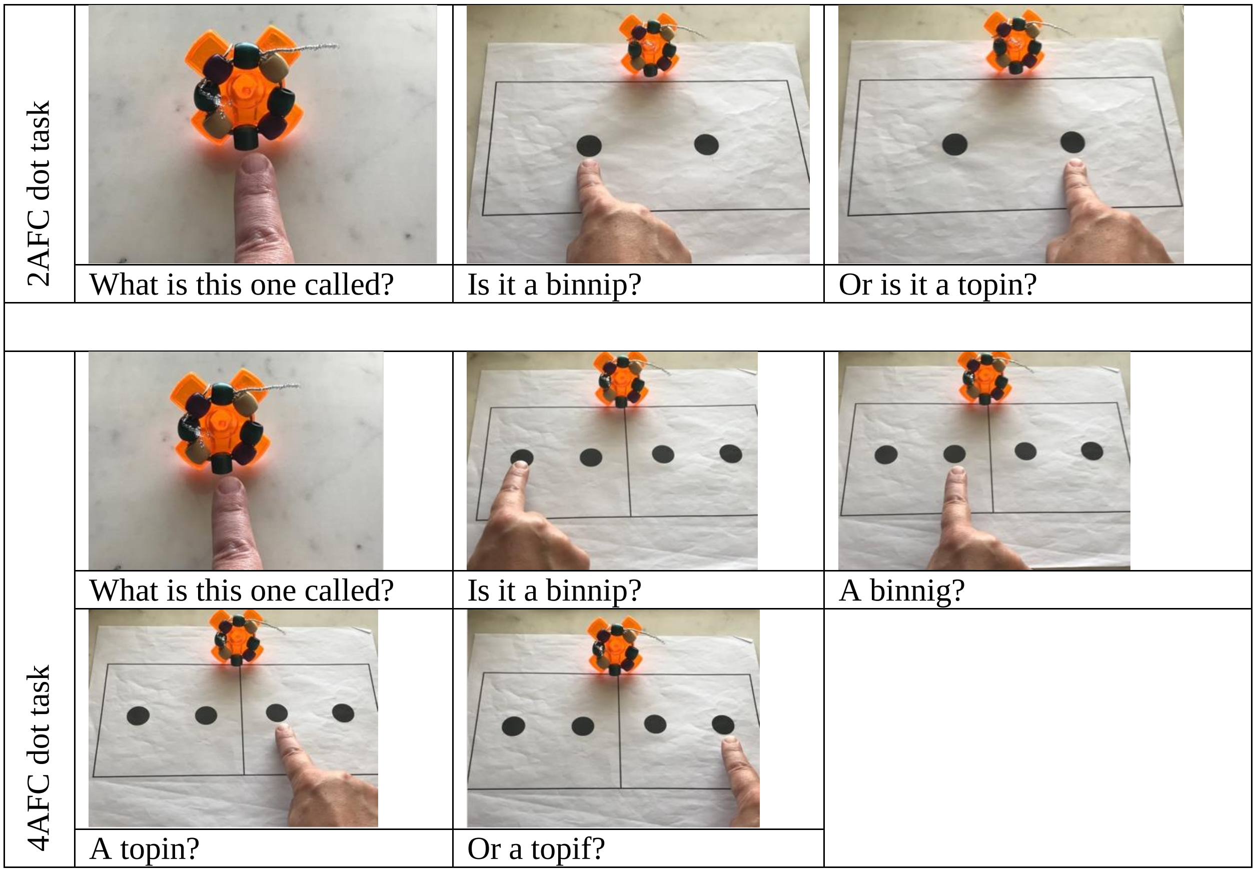
Table 2. Word forms and characteristics, neighborhood density. Neighborhood density was calculated using the English child corpus calculator (Vitevitch and Luce, 2004; https://calculator.ku.edu/density/about).

Target Form	Neighborhood Density	Minimal Pair	Neighborhood Density
/bɪnɪp/	0	/bɪnɪg/	0
/dob/	7	/dof/	7
/grɑmɚ/	0	/grɑtɚ/	0
/kinɪt/	0	/kigɪt/	0
/mep/	13	/mev/	15
/nedɪg/	0	/nedɪp/	0
/plun/	2	/plub/	0
/sibl̩/	2	/sifl̩/	0
/topɪn/	0	/topɪf/	0

Figure 1. Nine target objects

Appendix C. The 2AFC and 4AFC dot tasks

Figure 1: Illustrations of the 2AFC and 4AFC dot tasks used throughout training and the long-term session.



Description of Tasks

In a given trial of the 2AFC dot task, an object as well as a piece of paper with two large dots on it were placed in front of the child. The experimenter first pointed to the object and asked, “What is this one called?”. The experimenter then presented the child with two forms to choose from, the target form for the object present and another trained form from the set. As she produced the first form the experimenter pointed to one of the black dots on the paper. As she produced the second form she pointed to the other black dot. Children could indicate their response by producing one of the forms, by pointing to one of the black dots, or by doing both. If the child produced a form and pointed to the dot not associated with that form, their stated form took precedence. If the child said a form that was phonologically similar but not identical to one of the options, this was coded as a production of the target form (e.g., /meb/ for /mep/). If the child produced a form that was not phonologically similar to either of the options or blended two forms, the experimenter re-administered the trial and encouraged the child to select one of the stated options. The pairing of target and distractor forms as well as the order they were presented within trials (e.g., target form produced first or second) were randomized throughout training. Feedback of the correct form after each trial was given in a manner similar to the feedback provided during the Free and Cued Recall tasks.

The 4AFC dot task was similar to the 2AFC dot task, but each trial included four forms as options: the target form, a minimal pair of the target, another trained form (i.e., the distractor form), and a minimal pair of the distractor form. The 4AFC dot task was administered with the object placed at the top of a paper that included four large dots on it arranged with two dots each in two large squares. For example, the experimenter would show the child the object labeled as a /mep/ during training, and would ask the child, “What is this one called? Is it a /mep/, a /mev/, a /bɪnɪg/, or a /bɪnɪp/?” The experimenter pointed to each dot, left to right from the child’s point of view, simultaneous to the production of each form. The order that the four options were presented within trials was balanced such that the correct response was presented a similar number of times in each position (produced first, second, third or fourth). The minimal pair for each form was always either produced directly before or directly after that form. Responses were coded in the same way as the 2AFC dot task.

Analyses

For the 4AFC task administered at the end of training sessions, we conducted a generalized mixed effects model with the log odds of a correct response as the outcome variable.

The fixed effects of the maximal models included: group (DLD, TD), day, age (in months), and biological sex. We included a day by group interaction to determine whether participants with TD demonstrated a faster learning rate across days than participants with DLD. For this model, we included intercepts for participant and form as well as participant by day and form by day slopes as random effects. The random effect structure that best supported model fit included intercepts for participant and form. The minimal fixed effect structure that best supported model fit included group and day (Table 1).

Children with TD had a higher probability of selecting the correct form than children with DLD (B = -1.15, z = -3.38, p < 0.001). For example, on Day 1 the TD group had an average probability of 74% of selecting the target form. They selected the target form for an average of 6.69 out of the nine trained objects. In contrast, the DLD group had an average probability of 48% of selecting the target form. They selected the target form for an average of 4.30 objects.

Overall, performance improved across days (B = 0.36, z = 6.84, p < 0.001). The probability of selecting the target form increased by an average of 8% the equivalent of .73 forms per day. The group by training day interaction did not improve model fit, thus, it was excluded from the final model. This indicates that the probability of selecting the correct form increased to the same degree per training day across groups.

Table 1: Final models for word learning performance across training days, 4AFC dot task

Log Odds of Selecting the Target Form				
	Estimate	Standard Error	z value	Pr(>|z|)
Intercept	1.06	0.29	3.62	<.001
Group[a]	-1.15	0.34	-3.38	<.001
Day	0.36	0.05	6.84	<.001

[a]Reference group is typically developing.

Appendix D, Training and Testing Blocks

Table 1. Protocol during each training session

Block	Task	Objects	Script	Feedback	Exposures to Form with Target Referent Present
Testing Block 1	Free Recall Task	All objects	What is this one called?	None	0
Training Block 1	Explicit Labeling	Only objects that were not named correctly during Testing Block 1.	This is a /bɪnɪp/. You say /bɪnɪp/.	None	2
Training Block 2	2AFC Dot Task	Only objects that were not named correctly during Testing Block 1.	Is this one a /bɪnɪp/ or a /mep/?	Correct = "Yes that's right this is a /bɪnɪp/." Incorrect = "Actually this is called a /bɪnɪp/." Incorrect/close = "That's really close, but this one is a /bɪnɪp/."	2
Training Block 3	2AFC Dot Task	Only objects that were not named correctly during Testing Block 1.	Is this one a /grɑmɚ/or a /bɪnɪp/?	The same as Training Block 2.	2
Testing Block 2	4AFC Dot Task	All objects	Is this one a /bɪnɪp/, a /bɪnɪg/, a /dob/, or a /dof/?	The same as Training Block 2.	2
Testing Block 3	Free/Cued Recall Task	All objects	What is this one called? It starts with /bɪ/….	The same as Training Block 2.	1

Appendix E. Final models for all analyses

Table 1: Final models for word learning performance across training days, whole word score

Log Odds of Producing a Word Correctly					
	Estimate	Standard Error	z value	Pr(>|z|)	Cohen's *d*[c]
Intercept	-2.04	0.41	-4.99	<.001	-
Group[a]	-1.23	0.47	-2.63	<.01	0.43
Day	0.61	0.08	7.32	<.001	0.26[d]
Cue[b]	1.67	0.18	9.52	<.001	0.22
Group x Cue	-0.76	0.25	-3.06	<.01	-

[a]Reference group is typically developing. [b]Reference group is Free Recall. [c]Cohen's *d* is calculated with means and standard deviations of each group/condition. [d]Comparison is between Day 1 and Day 2.

Table 2: Final models for word learning performance across training days, phonological precision score

Percentage of Phonetic Features Produced correctly					
	Estimate	Standard Error	t value	Pr(>|z|)	Cohen's *d*[d]
Intercept	0.22	0.06	4.02	<.001	-
Group[a]	-0.11	0.05	-2.13	0.05	0.35
Day	0.07	0.01	7.56	<.001	0.17[e]
Cue[b]	0.44	0.02	18.09	<.001	0.51
Biological Sex[c]	0.08	0.05	1.56	0.14	0.18
Group x Cue	-0.08	0.03	-2.45	0.01	-

[a]Reference group is typically developing. [b]Reference group is Free Recall. [c]Reference group is Female. [d]Cohen's *d* is calculated with means and standard deviations of each group/condition. [e]Comparison is between Day 1 and Day 2.

Table 3: Final models for consolidation during training, whole word score

Log Odds of Producing a Word Correctly					
	Estimate	Standard Error	z value	Pr(>|z|)	Cohen's *d*[b]
Intercept	-2.36	0.52	-4.51	<.001	-
Group[a]	-1.33	0.63	-2.10	0.04	0.23
Gap	0.66	0.06	10.38	<.001	0.10[c]

[a]Reference group is typically developing. [b]Cohen's *d* is calculated with means and standard deviations of each group/condition. [c]Comparison is between Gap 1 and Gap 2.

Table 4: Final models for consolidation during training, phonological precision score

Percentage of Phonetic Features Produced correctly					
	Estimate	Standard Error	t value	Pr(>|z|)	Cohen's *d*[a]
Intercept	0.14	0.05	2.90	<0.05	-
Age	0.01	0.003	1.96	0.06	-
Gap	0.10	0.01	7.39	<.001	0.16 [b]

[a]Cohen's *d* is calculated with means and standard deviations of each group/condition.
[b]Comparison is between Gap 1 and Gap 2.

Table 5: Final models assessing the effect of delay (end of training, one-month delay) on production of words, whole word score.

Log Odds of Producing a Word Correctly					
	Estimate	Standard Error	z value	Pr(>|z|)	Cohen's *d*[d]
Intercept	0.58	0.56	1.05	0.30	-
Delay[a]	-0.67	0.20	-3.39	<.001	0.12
Group[b]	-1.49	0.71	-2.09	0.04	0.45
Cue[c]	1.46	0.20	7.14	<.001	0.26

[a]Reference group is end of training. [b]Reference group is typically developing. [c]Reference group is Free Recall. [d]Cohen's *d* is calculated with means and standard deviations of each group/condition.

Table 6: Final models assessing the effect of delay (end of training, one-month delay) on production of words, phonological precision score.

Percentage of Phonetic Features Produced correctly					
	Estimate	Standard Error	t value	Pr(>|z|)	Cohen's *d*[c]
Intercept	0.58	0.06	10.26	<.001	-
Group[a]	-0.16	0.07	-2.13	0.05	0.37
Cue[b]	0.35	0.03	12.33	<.001	0.19

[a]Reference group is typically developing. [b]Reference group is Free Recall. [c]Cohen's *d* is calculated with means and standard deviations of each group/condition.