**Author manuscript**

Pre-print of article accepted to *JSLHR*

Published in final edited form as:

*J Speech Lang. Hear. Res. 2022. doi: https://doi.org/10.1044/2022_JSLHR-22-00300*

# Direct instruction improves word learning for children with Developmental Language Disorder

Ron Pomper[1], Karla K. McGregor[1, 2], Timothy Arbisi-Kelm[1], Nichole Eden[1], and Nancy Ohlmann[1]

[1] Center for Childhood Deafness, Language and Learning, Boys Town National Research Hospital

[2] Department of Communication Sciences and Disorders, University of Iowa

## Author Note

Ron Pomper https://orcid.org/0000-0001-5595-4192

Karla McGregor https://orcid.org/0000-0003-0612-0057

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Ron Pomper, Center for Childhood Deafness, Language and Learning, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131, United States. Email: ronald.pomper@boystown.org

**Abstract**

**Purpose:** The current study compared the effects of direct instruction vs. indirect exposure on multiple aspects of novel word learning for children with Developmental Language Disorder (DLD) and children with typical language development (TLD).

**Method:** Participants included 36 children with DLD and 45 children with TLD. All children were in the first grade and 6 to 8 years of age (median = 7 years; 2 months). Using a between-subjects design, children were randomly assigned to be exposed to novel words and their unfamiliar referents via either direct instruction (each referent presented in isolation with an explicit goal of learning) or indirect exposure (multiple referents presented with the goal of answering yes/no questions).

**Results:** In alternative forced choice measures of recognition, children with DLD were less accurate than their TLD peers in linking words to referents, encoding semantic categories for words, and encoding detailed representations of word forms. These differences in word learning were accounted for by a constellation of cognitive measures, including receptive vocabulary, phonological memory, visuo-spatial memory, and sustained attention. All children were similarly accurate in retaining word forms over a 24- to 48-hour delay. Children with TLD were more accurate in all aspects of word learning following direct instruction compared to indirect exposure. Benefits from direct instruction were observed for children with DLD in link and semantic, but not word form, learning.

**Conclusions:** These results suggest that vocabulary interventions with direct instruction can help children with DLD learn some, but not all, aspects of novel words. Additional support is necessary to help children with DLD encode rich phonological representations.

# Direct instruction improves word learning for children with Developmental Language Disorder

People with Developmental Language Disorder (DLD) present with vocabularies that are smaller and less richly elaborated than their peers with typical language development (TLD; McGregor, Oleson et al., 2013), a gap that may disadvantage them academically (Biemiller & Slonim, 2001; Cunningham & Stanovich, 1997; Dockrell et al., 2007; Ehri et al., 2001). Given the same opportunity for learning, individuals with DLD, be they children (Kan & Windsor, 2010) or adults (McGregor, Arbisi-Kelm et al., 2020), will learn fewer words than their age-matched peers, a reliable effect of moderate size. These difficulties extend across multiple aspects of word learning. Relative to their peers, however, people with DLD tend to have more difficulty learning the word forms themselves than linking words to referents (Gray, 2004; Jackson et al., 2021; McGregor, Arbisi-Kelm, et al., 2020; McGregor, Licandro et al., 2013). Not surprisingly, then, it is often deemed necessary to provide additional opportunities for word learning to individuals with DLD in the form of language intervention (Steele & Mills, 2011).

In the current project, we investigate how two training contexts used in language interventions—direct instruction and indirect exposure—affect children's success in learning new words and whether the effect differs for children with DLD and their peers TLD. We examine multiple aspects of word learning, including how well children link words to referents, encode semantic category information, encode phonological representations of the word forms, and retain these phonological representations over a delay. Our use of multiple measures provides a more holistic exploration of word learning, moving beyond the tendency to focus solely on how children link words to referents (Wocjik, Zettersten, & Benitez, 2022). We focus on the early stages of word learning – what information children are able to encode after only a few exposures, but not how this process begins (triggering; e.g., Hoover, Storkel, & Hogan, 2010) or how it extends over time as children learn to approximate the full meaning of a word (e.g., Carey, 2010). Our measures test what children know about a word (lexical configuration), but not how this knowledge interacts with other words in their vocabulary (lexical engagement; e.g., Leach & Samuel, 2007). While we include a measure of retention, we do not systematically examine the process of consolidation (e.g., stabilization and enhancement; Walker, 2005). We focus on the early stages of word learning, because a compelling body of research demonstrates that the root of the word learning problem for children and adults with DLD often lies with their initial encoding of new words into long-term memory and not their ability to consolidate and retain this information over a delay (Gordon et al., 2021; Leonard, Deevy et al., 2019; Leonard, Karpicke et al., 2019; McGregor, Licandro et al., 2013; McGregor, Gordon et al., 2017).

**Individual differences in word learning**

For clinical purposes, individuals receive a categorical diagnosis – they either do or do not have DLD. Recent work suggests that DLD, however, should be conceptualized as a spectrum disorder (Lancaster & Camarata, 2019). Categorical grouping can mask significant heterogeneity in language ability amongst individuals with DLD. As a group, individuals with DLD perform worse than their peers with TLD on many measures of word learning. This does not mean, however, that every person with DLD experiences the same level of difficulty in learning new words. In fact, not all individuals with DLD struggle to learn new words (e.g., McGregor, Arbisi-Kelm, & Eden, 2017). When comparing accuracy in tests of word learning, there is often a high degree of overlap between the DLD and TLD groups and, in some instances, nearly completely overlapping ranges (e.g., McGregor et al., 2013). Prior research involving both children and adults with DLD has identified a range of cognitive factors that account for individual differences in word learning. Specifically, children and adults with DLD who have weaker phonological memory (Jackson et al., 2019; 2021), visuospatial memory (Kan & Windsor, 2010), and sustained attention (McGregor et al., 2022) tend to perform worse on measures of word learning. For children with TLD, measures of working memory (combining phonological and visuospatial) account for a substantial amount of variability in word learning over and above variability that is accounted for by differences in vocabulary size and nonverbal intelligence (Gray et al., 2022). For these reasons, it is important to compare word learning outcomes not only at the group level, but also across individuals.

**Learning words in different contexts**

People learn words in many different contexts. Indirect exposures are those that occur naturally in the world around us as we engage in conversations, watch television, and read books and other media. In these daily activities, direct instruction is not necessary—at least for typical language learners—because it is possible to infer the meanings of new words from visual and linguistic contexts. Speech-language pathologists (SLPs) who provide vocabulary interventions to toddlers and preschoolers frequently maximize opportunities for incidental exposures by using strategies such as focused stimulation (Cable & Domsch, 2011; Girolametto et al., 1996) or shared book reading (Ezell & Justice, 2005; Noble et al., 2019). Nevertheless, effects are often small, and hybrids that incorporate some direct teaching before or after the incidental exposures yield more robust outcomes (Pollard-Durodola et al., 2011).

By the early school years, most SLPs provide vocabulary interventions via direct instruction (Steele and Mills, 2011); they select a set of vocabulary targets that are educationally relevant and provide child-friendly definitions and synonyms, elicit productions in response to comprehension questions, and guide the child through exercises such as category sorting and semantic mapping (Beck, McKeown, & Kucan, 2013; Justice et al., 2014; McGregor & Duff, 2015). These intervention practices make explicit the meanings of the words and the contexts in which they can be used.

A large body of research has demonstrated that direct instruction is more effective than indirect exposure for children *without* DLD (*hedge's g* ~ 0.5; Marulis & Neumann, 2010). This benefit may be greater for typically-developing children with smaller compared to larger vocabularies (Coyne et al., 2004). Like their typically-developing peers, children with hearing loss benefit the most from vocabulary interventions with direct instruction compared to indirect exposure (Lund & Douglas, 2016). Given the prevalence of vocabulary intervention for children with DLD, it is somewhat surprising that we have a limited understanding of the extent to which direct instruction boosts vocabulary gains in these learners compared to indirect exposure. This lack of knowledge exists, in part, because there are many components of direct instruction. Only some of these components have been included in prior research comparing word learning outcomes for children with DLD to their peers with TLD. One compelling line of research demonstrates that practicing retrieval during learning is particularly helpful: compared to passive exposure, repeated spaced retrieval boosts success in word learning similarly for children with DLD and TLD (Haebig et al., 2019; Leonard, Deevy et al. 2019; Leonard, Karpicke et al., 2019). Another important aspect is providing child-friendly definitions of words: compared to indirect exposure via picture book reading, explicitly labeling referents and providing definitions boosts success in word learning similarly for children with DLD and TLD (Nash & Donaldson, 2005).

**Purpose of the present study**

We do not question the utility of practicing retrieval or providing definitions. Rather, we aim to isolate the essential core of direct instruction vs. indirect exposures to better understand their effects on children with DLD. Because indirect exposures take place in naturalistic contexts, the word and the referent (specified visually or linguistically) are available for the learner, but the goal of word learning is not specified. With direct instruction, the word and the referent are also available for the learner, in fact they are often the focal point of attention, and the goal of learning the word is made explicit. We hypothesize that the act of isolating the intended referent and explicitly identifying the goal of learning a new word during direct instruction (without active retrieval or explicit definitions) improves success in word learning for children with DLD.

To address our hypothesis, we use a protocol developed by Countache and Thompson-Schill (2014). They exposed adults with TLD to novel names for unfamiliar animals. For half of the participants, these exposures occurred via direct instruction: Each unfamiliar animal was presented in isolation and was labelled with a phrase like, “Remember the *torato*.” The link between the word and its referent is made explicit here, as is the goal of learning the word. Note that this type of instruction has been described in the research literature using different terms, including ostensive naming and explicit encoding. For the other participants, the exposures occurred via indirect exposure: The participant saw a familiar animal (e.g., an ant) alongside an unfamiliar animal and was asked, for example, “Are the antennae of the *blavid* pointing up?” Any word learning that occurs here is incidental; the participant is not directly told which one is the blavid and is not told to remember the word *blavid*. Note that this type of exposure has been labeled using different terms in the research literature including fast-mapping and referent selection (e.g., Carey & Bartlett, 1988; Horst & Sameulson, 2008). Given the inconsistency in the terminology and the tendency to use jargon in the literature, we use the terms direct instruction and indirect exposure for greater transparency.

When testing adult participants with TLD, both immediately after learning and one day later, Countache and Thompson-Schill (2014) found a large effect of training on word learning. On average, participants in the direct instruction condition correctly identified 80.7% of the referents when given the word form, while those in indirect exposure condition correctly identified only 56.2% of the referents. Note that these trials (where a participant is shown three novel objects from training and asked to identify the one that is named) are variously referred to as retention trials, declarative memory trials, or simply just test trials, because participants can only succeed if they successfully linked novel words to their intended referents during training. To distinguish between the multiple tests of learning we use (see below), we will refer to these as link recognition trials. McGregor, Eden et al. (2020) extended this protocol to adults with and without DLD. They similarly found a large improvement in link recognition and semantic category recognition following direct instruction compared to indirect exposure. While the DLD group performed more poorly overall than the TD group, there was no interaction between group and condition, suggesting that the DLD group and TLD group similarly benefitted from the identification of the intended referent and being prompted with the explicit goal of learning new words.

In the present study, we asked whether 6- to 8-year-old children with DLD would experience the same improvements in word learning from direct instruction as their peers with TLD. All participants were in the first grade, the point at which many children receive vocabulary interventions that involve direct instruction (Steele and Mills, 2011). We addressed this question using several aims, examining:

1. **The effect of direct instruction on multiple aspects of word learning, including learners' ability to link words to their referents (link recognition), encode**

**information about the semantic categories of novel words (semantic category recognition), and encode phonological representations of the word forms themselves (word form recognition).** Consistent with research involving adults (McGregor, Eden, et al., 2020), we expected the improvements in both link and semantic category recognition from direct instruction compared to indirect exposure would be similar for children with DLD and their peers with TLD. It is possible, however, that children with DLD would benefit more than their peers with TLD from direct instruction, because they may have a greater difficulty learning via indirect exposure given their lower extant linguistic knowledge. Although both training conditions are equated in the number of exposures to the word forms, there is a greater cognitive load in the indirect condition given the greater amount of visual information on the screen and the need to respond to a question with an answer based on inference. Therefore, we predicted that direct instruction would similarly boost word form recognition for both the DLD and TD groups.

2. **The effect of direct instruction on children's retention of this initial learning.** We focused on learners' ability to retain novel word forms over a 24- to 48-hour delay. Given consistent results in the extant literature (Leonard, Deevy et al., 2019; Leonard, Karpicke et al., 2019; McGregor, Licandro et al., 2013; McGregor, Gordon et al., 2017), we predicted that children with DLD and TLD would not differ in their retention and that retention would be similar following direct instruction and indirect exposure.
3. **The extent to which vocabulary, phonological memory, visuospatial memory, and sustained attention support each aspect of word learning and whether word**

**learning differences between children with DLD and TLD persist after controlling for variability in these cognitive factors**. As previously mentioned, each of these factors is associated with success in word learning for children with DLD (Jackson et al., 2019; 2021; Kan & Windsor, 2010; McGregor et al., 2022). We therefore predicted similar outcomes in word learning between groups (i.e., fail to reject the null hypothesis) after accounting for variability in learning attributed to individual differences in these cognitive measures.

## Method

### Ethics

This study was approved by the institutional review board of [removed for anonymous review], approval number 17-04-XP. Participants gave informed consent/assent before taking part. The data were collected between March of 2018 and October of 2020, during the first year of a four-year longitudinal study investigating changes in word learning for children with DLD (Research Registry 3425, 2017). Pilot data were collected between December of 2017 and February of 2018. Pilot participants were 9 children (2 female) all with TLD. Piloting was used to determine the feasibility for children in our age range to complete the word learning tasks. Pilot data was not analyzed, but checked to make sure that responses were being correctly logged by the software.

### Participants

Participants were 81 first graders (40 female) between 6 to 8 years of age (median = 86 months, range = 74 to 98 months), 36 children with Developmental Language Disorder (DLD) and 45 children with typical language development (TLD). Four additional children participated in data collection but were not included in the final sample, because they were the appropriate age but one grade ahead in school ($N$ = 1), were subsequently diagnosed with epilepsy ($N$ = 1), or could not conclusively be included in either group ($N$ = 2).[1] For the last 5 participants in our sample (all in the DLD group), data collection switched from in-person to online due to the COVID-19 pandemic. These children were unable to complete many of cognitive measures (see below) that did not have options for online administration at that time. In addition, 2 children in the TLD group did not complete the cognitive measure of sustained attention due to technical issues. Therefore, results for models including the cognitive measures as covariates were fit using a sample of 31 children with DLD and 43 children with TLD.

Children in the DLD group scored below the 15th percentile on a sentence recall screening task developed by Redmond (2005) and scored below a standard score of 92 on the *Test of Narrative Language*, first or second edition (TNL; Gillam & Pearson, 2004; 2017). The TNL assesses both receptive and expressive language, is normed nationwide, and exhibits minimal gender and racial bias. A cut-off of 92 has been demonstrated to have 92% sensitivity and specificity in identifying children with DLD (Gillam & Pearson, 2017). Children in the TLD group scored above a standard score of 92 on the TNL. Table 1 summarizes the demographic characteristics and test scores for children in each group.

All participants met the following inclusionary criteria: exposed primarily to English (fewer than 10 hours per week of another language), normal hearing (pass a pure-tone audiometric screening at 0.5, 1, 2, and 4 kHz at 25 dB bilaterally), no indication of intellectual disability (via parent report and a standard score of 70 or higher on the Matrices and Block Design subtests of the Wechsler Abbreviated Scales of Intelligence, 2nd Edition; Wechsler,

[1] One child scored within the TLD range on both language measures (see next paragraph), but was receiving services from a Speech Language Pathologist. The other child scored within the TLD range on one language measure (sentence recall), but within the DLD range on the second language measure (*Test of Narrative Language*).

2011), no diagnosis or suspected Autism Spectrum Disorder (score of 15 or below on the Social Communication Questionnaire; Rutter et al., 2003), and a health history report indicating no other neurological or developmental disorders aside from Attention Deficit Hyperactivity Disorder (ADHD is often co-morbid with DLD; Sciberras et al., 2014). Six children in the DLD sample (16.7%) and three children in the TLD sample (6.7%) had a diagnosis of ADHD. All analyses were repeated excluding children with ADHD and are available via Open Science Framework (https://osf.io/26djx/?view_only=3c133119e79144e4896439ab3227e0b1). We find a strikingly similar pattern of results both when including and excluding children with ADHD. This indicates that any observed group differences that are reported in our analyses below cannot be attributed to a greater proportion of children with ADHD in the DLD compared to TLD group.

## Procedure

Participation involved three visits. Each visit lasted approximately 1 hour with the second visit occurring 1 to 2 days following the first visit and the third visit occurring 3 to 5 days following the second visit. Forty-four children completed the second visit after 1 day (20 DLD; 24 TLD) and 37 completed the second visit after 2 days (16 DLD, 21 TLD). On average, the second visit occurred 1.45 days after the first visit and was similar for children with DLD and TLD (1.44 and 1.47 respectively). The order of the tasks for each visit is presented in Table 2. The tasks for the current research are described in greater detail below. Details and results for other tasks are reported elsewhere (McGregor et al., 2022; Smolak, McGregor, Arbisi-Kelm, & Eden, 2020).

## Cognitive Measures

### *Vocabulary*

Children completed the NIH Toolbox Picture Vocabulary Task (Gershon et al., 2013, 2014; Weintraub et al., 2013) to measure their receptive vocabulary. Using an iPad, children heard one word and saw four pictures on the screen. They were asked to touch the picture that best matched the meaning of the word they heard. Children completed 2 practice trials, which included feedback regarding accuracy. They then completed a maximum of 25 test trials without feedback. The administration of test trials is adaptive – children's accuracy on prior trials is used to select trials with moderate difficulty (i.e., 50% likelihood the child will answer correctly). Testing continues until children's performance reaches a cut-off (standard error less than 0.3). Children's performance was quantified using age-adjusted standard scores, which are normed to have a mean of 100 and a standard deviation of 15. Children in the DLD group had significantly lower vocabulary scores ($M = 92.4$, $SD = 14.6$, range = 75-125) than children in the TLD group ($M = 110.6$, $SD = 14.8$, range = 78-140), $b = 18.1$, $t = 5.3$, $p < .001$.

***Phonological Memory***

Children completed the nonword repetition test (Dollaghan & Campbell, 1998) to measure their phonological short-term memory. They were told that they would hear some made-up words and were asked to "repeat the words back in exactly the same way as you hear them." Children were tested on sixteen nonwords that varied in syllable length: four 1-syllable words (CVC), four 2-syllable words (CVCVC), four 3-syllable words (CVCVCVC), and four 4-syllable words (CVCVCVCVC). Children's responses were audio recorded and each phoneme (consonant or vowel) was scored as either correct or incorrect; substitutions and omissions were scored as incorrect, while additions were not scored as errors. Children's raw score was the number of correct phonemes produced with a maximum score of 96.

Children completed the backward digit recall test (Alloway et al., 2008) to measure their phonological working memory. Children were required to recall a sequence of spoken digits (between 1 and 9) in reverse order. They completed four practice trials: two trials with a 2-digit sequence and two trials with a 3-digit sequence. Children then completed up to 6 blocks of test trials, with each block increasing the length of the digit sequences to be recalled (starting with 2-digit sequences, ending with 7-digit sequences). Each block consisted of 6 trials and ceiling was reached when a child was unable to accurately recall 4 or more trials within a block. Children's raw score was the number of correct trials with a maximum score of 36.

Children's performance on the nonword repetition and backward digit recall tests were strongly correlated, $r = 0.56$, $t = 5.99$, $p < .001$. To avoid multi-collinearity in our models, we calculated a single composite phonological memory score for each child. This score was calculated by converting children's raw scores on each task into z-scores (dividing raw scores by the standard deviation for the entire sample) and averaging both z-scores. Children in the DLD group had significantly lower phonological memory z-scores ($M = -0.60$, $SD = 0.81$, range = -2.49 to 0.73) than children in the TLD group ($M = 0.48$, $SD = 0.60$, range = -0.98 to 2.09), $b = 1.09$, $t = 6.94$, $p < .001$.

***Visuo-spatial memory***

Children completed the Corsi Block-Tapping Test (Farrell et al., 2006) to measure their visuo-spatial short-term memory. Children were presented with an array of nine wooden blocks. They watched as the experimenter pointed to some of the blocks in a certain order. Children were then asked to point to the blocks in the same order as the experimenter. They completed up to 9 sets, with each set increasing the tapping sequence by one additional block (starting with 1-block sequences, ending with 9-block sequences). Each set consisted of five trials. If a child correctly reproduced the first four trials within a block, the fifth trial was not administered and the child received full credit for that set (i.e., 5 correct trials). Children started on the 3rd set; if they did not answer all trials in this block correctly (i.e., establish a basal set), they completed the 2nd and 1st sets. Ceiling was reached when a child answered incorrectly on all five trials for a set. Children's raw score was the total number of correct trials with a maximum score of 45.

Children completed the Odd-One-Out Task (Henry, 2001) to measure their visuo-spatial working memory. They were shown images of three similar-looking figures displayed in a row on the computer screen; two of the figures were identical and the third differed slightly from the other two. Children were asked to tap the odd-one-out that is different from the others. The figures disappeared and were replaced with a row of three rectangular boxes. Children were then asked to tap the location of the odd-one-out figure. Children completed two practice trials: one with a 1-item length (i.e., identify one odd-one-out before recalling its position) and one with a 2-item length (i.e., identifying two odd-ones-out before recalling their positions). Children then completed up to 6 blocks of test trials with each block increasing the number of items to recall (starting with 1-item trials, ending with 6-item trials). Each block consisted of four odd-one-out sequences and four position recall trials. Children's responses on position recall trials were scored correct only if they correctly identified the positions for every odd-one-out figure in the sequence (e.g., all 6 positions in the 6th block).[2] Ceiling was reached when a child answered incorrectly on two or more position recall trials within a block. Children's raw score was the total number of position recall trials that were correct with a maximum score of 24.

Children's performance on the Corsi Block-Tapping and Odd-One-Out were strongly correlated, $r = 0.51$, $t = 5.09$, $p < .001$. To avoid multi-collinearity in our models, we calculated a

[2] When a child incorrectly identified which figure was the odd-one-out, this position was used as the target position on the position recall trial.

single composite visuo-spatial memory score for each child. This score was calculated by converting children's raw scores on each task into z-scores (dividing their performance by the standard deviation for the group) and averaging both z-scores. Children in the DLD group had significantly lower visuo-spatial memory z-scores ($M$ = -0.59, $SD$ = 0.73, range = -2.04 to 1.28) than children in the TLD group ($M$ = 0.41, $SD$ = 0.66, range = -0.74 to 2.23), $b$ = 1.00, $t$ = 6.19, $p$ < .001.

***Sustained Attention***

Children completed the Track-It task to measure their visual sustained attention (Erickson et al., 2015; Fisher et al., 2013). They were shown 4x4 grids of boxes with 9 of the boxes containing shapes. A target shape was identified using a red circle, the red circle disappeared, and the shapes randomly moved around the grid in a smooth path for 20 to 35 seconds. The shapes stopped moving, immediately disappeared, and children were asked to tap the last location of the target shape on the grid. For homogeneous trials, all the distractors were the same shape with only the target shape differing. For heterogeneous trials, all shapes were different. Children completed three training trials, six homogeneous trials, and six heterogeneous trials. Homogeneous and heterogenous trials were blocked and their order counter-balanced between children. Training trials were either homogeneous or heterogenous to match the first block of test trials. After every tracking trial, children completed a memory check – they were shown the target shape and three distractor shapes in a 2x2 grid and were asked to tap the shape they had been tracking.

Children's raw score was the proportion of heterogeneous trials correct, excluding tracking trials where they subsequently failed the memory check. This provides a measure of children's ability to endogenously sustain attention (children's performance on homogeneous trials results is also affected by the salience of the target shape which is more salient from the uniform distractors) that is not affected by failures in their ability to encode/retain the target shape. Previous research has demonstrated that this particular measure of sustained attention is correlated with individual differences in narrative language ability and cross-situational word learning for children with DLD (McGregor et al., 2022; Smolak et al., 2020). Children in the DLD group had significantly lower sustained attention scores ($M$ = 0.68, $SD$ = 0.30, range = 0 to 1) than children in the TLD group ($M$ = 0.86, $SD$ = 0.21, range = 0 to 1, $b$ = 0.17, $t$ = 2.88, $p$ < .01.

## Novel Word Learning

Children were taught the names of novel objects via either direct instruction or indirect exposure (between-subjects design). The methods for both training conditions were the same as prior research in which adults with DLD and TLD were taught different sets of words using direct instruction (referred to as Fast Mapping) and indirect exposure (referred to as Ostensive Naming; McGregor et al., 2020). Children were tested after a 5-minute and 24- to 48-hour delay to measure their success in both learning and retention.

### *Stimuli*

The entire stimulus set consisted of four sets of 12 novel words and 12 unfamiliar referents depicted in color photographs, for a total of 48 form-referent pairs. Each child was tasked with learning 12 form-referent pairs (the remaining sets were used during subsequent years in the longitudinal project).

The unfamiliar referents were mammals (e.g., a tenrec), insects (e.g., a giraffe-necked weevil), birds (e.g., a sunbittern), or fruits (e.g., a sapote). Each set of 12 unfamiliar referents consisted of three mammals, three insects, three birds, and three fruits. Two familiar referents were included as filler stimuli (dog and watermelon). Twelve familiar referents were included as distractors for the indirect exposure condition (giraffe, horse, bear, fly, ant, butterfly, duck, flamingo, parrot, coconut, banana, and strawberry).[3] Each child was randomly assigned to learn one of the four sets of novel words and referents such that each set of referents and words occurred equally often for each group (DLD, TLD) and each training condition (direct, indirect). Photographs of each referent were found online and edited using GNU Image Manipulation to be matched approximately in size and placed on a white background 400 by 400 pixels in size.

For each set of 12 novel words, half were monosyllabic and half were disyllabic and ranged in length from 3 to 6 phonemes. All disyllabic words contained first syllable stress patterns. Ten of the words had unique onsets and the remaining two words shared the same onset. All four sets of novel words were balanced in phoneme length, feature distribution (place, manner, and voicing), neighborhood density ($M = 3.85$ neighbors; Vitevitch & Luce, 2004), and phonotactic probability (positional segment frequency $M = 0.1913$; positional biphone frequency $M = 0.0108$; Kucera & Francis, 1967). Three different speakers (two female, one male) were recorded producing the novel and familiar words. Previous research has demonstrated that speaker variability facilitates the encoding of detailed phonological representations of new words (Creel et al., 2008; Richtsmeier et al., 2009; Rost & McMurray, 2009; 2010).

***Training***

For direct instruction, children were told, “You will see pictures on the computer screen. Your job is to remember what you see and hear.” On each trial, children were shown an image of an unfamiliar referent in isolation and heard a sentence labelling it (see Figure 1). For indirect exposure, children were told, “You are going to see two things on the computer screen, and we

[3] A test at the end of the second visit confirmed that children knew the names of all 12 familiar referents.

are going to ask you questions about one of them." On each trial, children were shown images of an unfamiliar referent and a familiar referent with a green thumbs up and a red thumbs down displayed beneath on the screen. They then heard a sentence with the label for the unfamiliar referent embedded in a yes/no question. Children responded by clicking/tapping the green thumbs up image to answer yes and the red thumbs down image to answer no. All questions focused on visual features of the unfamiliar referents. Each unfamiliar referent was paired with a familiar referent from the same semantic category that differed in the relevant visual feature. Using a between-subjects design, each child was randomly assigned to be in only one of the two training conditions.

Children completed a total of 70 training trials that were arranged into 5 blocks. Each block consisted of 14 trials: two trials with familiar referents (dog and watermelon) and twelve trials with the novel referents. Each unfamiliar referent was shown and labelled once per block (five times in total). For the indirect exposure condition, each unfamiliar referent occurred with the same familiar referent on all 5 trials. Within each block, trials were presented in random order with randomization varying across blocks (i.e., the order in which children encountered word-referent pairings varied across blocks). For the indirect exposure condition, the unfamiliar referent occurred equally often in the left and right position and the correct responses to the questions (yes vs. no) occurred equally often.

***Testing***

Children completed three tasks measuring different aspects of word learning across multiple visits (see Figure 2). These tasks measured children's receptive knowledge by quantifying their accuracy in identifying the target item (by tapping a touchscreen or clicking a mouse) from an array with two or three foils (i.e., 3- and 4-alternative forced choice measures). In the next paragraphs, we describe each task in greater detail.

Children completed a 3-AFC word-to-referent link recognition task. For each trial, children were shown images of three unfamiliar referents from training and heard a novel word labelling one of the referents. They were instructed to touch/click on the picture that went with the word. If children did not respond within 5 seconds, the trial ended and their accuracy for that trial was marked as incorrect (i.e., a timeout trial). This same time-out criterion is used in prior research (McGregor et al., 2020). Thus, a trial may be incorrect because the child consciously chose a foil (i.e., linked the wrong label to a referent from training), randomly chose a foil (i.e., did not form a link during and so guessed), or did not respond in time. Consciously choosing a foil likely reflects a different type of failure than randomly choosing or not responding. We therefore repeat our analyses excluding time-out trials (with the current methods, it is not possible to discriminate between conscious vs. random choices). Children completed 12 total trials. Item order was randomized for each child. The referents that occurred on each trial were chosen pseudorandomly such that a maximum of two items were from the same semantic category. Across trials, each unfamiliar referent occurred once as the target and twice as a foil, the target referent occurred equally often in each spatial location, and novel words were spoken equally often by each of the speakers from training. Children completed this once, which occurred 5 minutes after training.

Children completed a 4-AFC semantic category recognition task. For each trial, children were shown the same four silhouettes (eagle, beetle, cow, apple) representing four different semantic categories (bird, insect, mammal, fruit) and heard a novel word from training. Children were instructed to touch/click the picture that matched the kind of thing named by the word.

Given the large number of stimuli that were needed (48 unfamiliar referents), we chose referents from four, rather than three, semantic categories. The semantic category recognition task therefore had a 4-AFC format. To be consistent with prior research (Gordon et al., 2022; McGregor et al., 2020) we chose not to increase the number of foils for link and word-form recognition trials. The different number of foils complicates comparisons of children's link and word-form recognition accuracy with their semantic category recognition accuracy, but allows us to compare accuracy on all three tasks with prior research involving children and adults with DLD. The experimenter explained the task using two familiar referents: "If you heard the word *horse* you would touch the mammal picture, because a horse is a type of mammal. Horses, rabbits, and cats are all mammals. If you heard the word *grapes* you would touch the fruit picture, because grapes are a type of fruit. Apples and grapes are both fruit." Given the increased complexity of the task (generalizing referents to broader categories and the increased number of foils), we were uncertain how much extra time children would need to respond. Each trial therefore had an unlimited duration and only advanced after the child selected one of the images. The longer it takes children to respond, the less likely they are to remember the target word for that trial. We hypothesize that incorrect trials with longer latencies are therefore more likely to reflect random choices than incorrect trials with shorter latencies. We therefore repeat our analyses excluding trials with response latencies longer than 8 seconds. Children completed three practice trials with familiar words (duck, spider, dog) then 12 test trials. Across trials, each novel word from training occurred once, all four images occurred in the same fixed spatial locations, the target occurred equally often in each spatial location, and novel words were spoken equally often by each speaker. Children completed this once, which occurred 24 to 48 hours after training.

Children completed a 3-AFC word-form recognition task. For each trial, children heard a target novel word and two novel word foils. An image of a dot appeared on the screen simultaneously with the presentation of each word. Dots were arranged in a row and appeared from left to right. Children were instructed to touch/click on the dot that matched the word they just learned. If children did not respond within 5 seconds, the trial ended and children's accuracy for that trial was marked as incorrect (i.e., a timeout trial). As with link recognition, we repeat our analyses excluding time-out trials. For each trial, two phonological foils were created by changing one phoneme (always a consonant) from the target word; for monosyllabic words it was the offset and for disyllabic words it was the onset of the second syllable. For one foil, the modified phoneme differed from the target phoneme on one feature (place, manner, or voice); for the second foil, the modified phoneme differed on two features. Children first completed three practice trials with familiar words (e.g., "Imagine you just learned the word sparkle. You hear spartle, sparkle, sparfle. Touch the dot that matches your new word."). Afterwards children completed twelve test trials. Across trials, each novel word from training occurred once, the target occurred equally often in each spatial location, and novel words were spoken equally often by each speaker. Children completed this task twice – the first test occurred 5 minutes after training and the second test occurred 24 to 48 hours after training.

**Data Analyses**

The dependent variable was children's accuracy in selecting the target averaged across all 12 trials. Accuracy is centered on chance so that model intercepts indicate the extent to which accuracies were significantly greater than chance. Separate models were fit using children's accuracy on each type of test trial: link, semantic category, and word form recognition. For each model, children's accuracy was regressed on the between-subject effect of training condition (contrast coded as -0.5 for indirect and +0.5 for direct), the between-subject effect of diagnostic group (contrast coded as -0.5 for DLD and +0.5 for TLD), and the two-way interaction. All analyses were repeated using logistic mixed effects to analyze data at the individual trial level. We find a similar pattern of results. For ease of interpretability, we report here the linear regression analyses using accuracy averaged across trials. Results for the logistic mixed effects models using individual trial accuracies are available via OSF (https://osf.io/26djx/?view_only=3c133119e79144e4896439ab3227e0b1).

Recall that there are 12 trials for each test. For link recognition, children in the DLD group had on average more time-out trials (*M* = 1.1, *SD* = 2.3) than children in the TLD group (*M* = 0.40, *SD* = 0.61). Of the trials that were scored as incorrect (foil was selected or time-out), 13.8% (*SD* = 25.4%) were time-out trials in the DLD group and 7.1% (*SD* = 11.8%) were time-out trials in the TLD group. For form recognition, children in the DLD group also had on average more time-out trials (*M* = 1.31, *SD* = 1.51) than children in the TLD group (*M* = 1.04, *SD* = 1.83). Of the trials that were scored as incorrect (foil was selected or time-out), 15.8% (*SD* = 20.1%) were time-out trials in the DLD group and 23.6% (*SD* = 30.0%) were time-out trials in the TLD group. Since time-out trials were scored as incorrect, children with DLD may have lower accuracy in part because they had more time-out trials. For semantic category recognition there was no time limit, children in the DLD group, however, had on average more trials with response latencies longer than 8 seconds (*M* = 0.94, *SD* = 1.85) than children in the TLD group (*M* = 0.53, *SD* = 1.18). At longer intervals (e.g., one trial had a response latency of 80 seconds), children may no longer remember the target word and therefore respond randomly. Children with DLD may therefore have lower accuracy in part because they had more trials with long latencies.

All analyses were therefore repeated (link, semantic category, word form) with these trials excluded.

Children in the DLD group had on average lower vocabulary, phonological memory, visuo-spatial memory, and sustained attention than children in the TLD group. Because each of these cognitive factors has been shown to predict differences in children's success in learning new words, we refit our models (with all trials) to include the cognitive factors as covariates. Because nearly all of these measures are correlated (see OSF) there were high levels of multi-collinearity. The fixed effect for each cognitive factor therefore indicates the extent to which it accounts for *unique* variance in word learning (e.g., after removing the shared variance accounted for by the other cognitive factors).

Overall, we find a similar pattern of results across all three types of analyses (unadjusted, adjusted to exclude trials, adjusted to include covariates). This indicates that our observed effects are fairly robust. Model results for all analyses are included in the Supplementary Materials that are available via OSF and Table 3 provides comparisons of *beta* estimates for all significant effects in every model. For brevity, we report in detail the results for the models that were unadjusted (i.e., did not exclude trials or include the cognitive measures as covariates). We then highlight any changes that occur when the covariates were added. This two-step approach allows us to determine the size of the group differences between children with DLD and TLD and then examine the extent to which group differences are accounted for by the cognitive measures. This approach is similar, but not equivalent to a mediation analysis – testing whether the addition of a third variable (cognitive measure) significantly decreases the strength of the original correlation (between language group and word learning success).

Given the between-subject design and potential for heterogeneous variance between groups, we used linear regression models fit using Generalized Least Squares (GLS) and restricted maximum likelihood approach (REML). Each model was fit twice – once assuming homogeneous variance between groups and once assuming heterogeneous variance between groups. AIC values were compared between model fits to identify the most parsimonious model (i.e., a value more than 2 points lower than the other model). For link recognition, models that assumed homogeneous variance were more parsimonious (indicating that variability amongst children in identifying the correct referents of novel words was similar for the DLD and TLD groups). For semantic category recognition and word form recognition, models that assumed heterogeneous variance were more parsimonious (indicating that variability in identifying the correct semantic categories and word forms was not the same for the DLD and TLD groups). Analyses were completed using R (R Core Team, 2022; version 4.1.1) in RStudio (RStudio Team, 2020; version 1.4.1717) and the nlme package (Pinheiro et al., 2022; version 3.1-155). Data manipulation and plotting were completed using the tidyr (Wickham & Girlich, 2022; version 1.2.0) and ggplot2 (Wickham, 2016; version 3.3.5) packages. The raw data and R code are available in OSF (https://osf.io/26djx/?view_only=3c133119e79144e4896439ab3227e0b1).

## Results

The complete results are summarized in Figure 3 and Tables 3-6. Below, we consider the results in more detail for each aspect of word learning that we measured: link recognition, semantic category recognition, word form recognition, and word form retention.

### Link recognition

Full model results for the unadjusted analysis (including all trials, no cognitive covariates) are available in Table 4. Recall, that children's accuracy in identifying a referent after it was named was tested 5 minutes after training. Overall, children's accuracy in linking novel words to their pictured referents (*M* = 49.79%, *SD* = 23.53%) was significantly greater than chance (33%).

There was a significant effect of group; children in the TLD group were more accurate in recognizing the link between word and referent (*M* = 55.74%, *SD* = 23.49%) than children in the DLD group (*M* = 42.36%, *SD* = 21.67%) There was a significant effect of training condition; children were more accurate after direct instruction (*M* = 59.76%, *SD* = 23.34%) than indirect exposure (*M* = 39.58%, *SD* = 19.13%). The interaction between group and training condition was not statistically significant, indicating that the effect of training was similar for both groups.

Although there is no need to further explore this interaction, we report the effect of training condition separately for each group for full transparency. For children with TLD there was a significant effect of training condition; they were more accurate in identifying the correct referents of novel words learned via direct instruction (*M* = 65.91%, *SD* = 23.7%) than indirect exposure (*M* = 46.01%, *SD* = 19.11%). Their accuracy in both training conditions was significantly greater than chance. For children with DLD there was also a significant effect of training condition; they were more accurate in identifying the correct referents of novel words learned via direct instruction (*M* = 52.63%, *SD* = 21.35%) than indirect exposure (*M* = 30.88%, *SD* = 15.8%). Their accuracy in the direct condition, but not the indirect condition, was significantly greater than chance.

After adding covariates into the model, the effect of group was no longer significant, $b = 0.111$, $t(66) = 1.731$, $p = 0.088$. All other fixed effects remain unchanged (see Supplementary Materials). This suggests that group differences in word-referent mapping (TLD > DLD) are accounted for, in part, by differences in cognitive factors (vocabulary, phonological memory, visuospatial memory, and sustained attention) between the groups. Put another way, the variance accounted for by diagnostic group is shared with the variance accounted for by the cognitive factors and the remaining unique variance accounted for by diagnostic group is not statistically significant. Of the four cognitive factors, only children's phonological memory (combination of Non Word Repetition and Backwards Digits Tasks) was a significant predictor of their success in identifying the correct referents of novel words, $b = 0.089$, $t(66) = 2.599$, $p = 0.012$. For each 1 SD increase in children's phonological memory, their accuracy in identifying the correct referents of novel words increased by 8.9%.

**Semantic category recognition**

Full model results for the unadjusted analysis (including all trials, no cognitive covariates) are available in Table 5. Recall, that children's accuracy in identifying the semantic category (bird, insect, mammal, fruit) for each word was tested 24 to 48 hours after training. Overall, children's accuracy in identifying the correct semantic categories of novel words ($M$ = 39.3%, $SD$ = 19.38%) was significantly greater than chance (25%).

There was a significant effect of group; children in the TLD group ($M$ = 46.48%, $SD$ = 21.06%) were more accurate in identifying the correct semantic categories of novel words than children in the DLD group ($M$ = 30.32%, $SD$ = 12.3%). There was also a significant effect of training condition; children were more accurate in identifying the correct semantic categories of novel words learned via direct instruction ($M$ = 44.92%, $SD$ = 21.48%) than indirect exposure ($M$ = 33.54%, $SD$ = 15.15%). The interaction between group and training condition was not statistically significant in the unadjusted analyses, but was significant for the analyses excluding trials with long latencies, $b = 0.157$, $t(77) = 2.146$, $p = 0.035$.

For children with TLD there was a significant effect of training condition; they were more accurate in identifying the correct semantic categories of novel words learned via direct instruction ($M$ = 55.68%, $SD$ = 14.41%) than indirect exposure ($M$ = 37.68%, $SD$ = 9.29%). Their accuracy in both training conditions was significantly greater than chance. For children with DLD, however, there was not a significant effect of training condition; they were similarly accurate in identifying the correct semantic categories of novel words learned via direct instruction ($M$ = 32.46%, $SD$ = 14.41%) and indirect exposure ($M$ = 27.94%, $SD$ = 9.29%). Their accuracy in the direct, but not the indirect, condition was significantly greater than chance.

When adding covariates to the model, the fixed effects remain unchanged. Of the four cognitive factors, only children's sustained attention (TrackIt task) was a significant predictor of their success in identifying the correct semantic categories of novel words, $b = -0.185$, $t(66) = -2.53$, $p = 0.014$. A child with the highest sustained attention (i.e., 100% correct) was 18.5% *less* accurate than a child with the lowest sustained attention (i.e., 0% correct). This effect is contrary to our prediction and should be interpreted with caution since the effect is less robust (i.e., when excluding children with ADHD the effect is marginally significant, $b = -0.16$, $t(59) = -1.92$, $p = 0.06$).

**Word form recognition**

Full model results for the unadjusted analysis (including all trials, no cognitive covariates) are available in Table 6. Recall, that children's accuracy in identifying the trained novel word from two phonological foils was tested both 5 minutes and 24 to 48 hours after training. We report here the results after the 5-minute delay and focus on changes in accuracy between tests in the next section. Overall, children's accuracy in identifying the correct forms of novel words ($M$ = 45.27%, $SD$ = 23.82%) was significantly greater than chance (33%).

There was a significant effect of group; children in the TLD group (*M* = 56.11%, *SD* = 24.52%) were more accurate in identifying the correct forms of novel words than children in the DLD group (*M* = 31.71%, *SD* = 14.2%). There was a significant effect of training condition; children were more accurate in identifying the correct forms of novel words learned via direct instruction (*M* = 50.81%, *SD* = 26.27%) than indirect exposure (*M* = 39.58%, *SD* = 19.77%). The interaction between group and training condition was statistically significant, indicating that the effect of training varied between groups.

For children with TLD there was a significant effect of training condition; they were more accurate in identifying the correct forms of novel words learned via direct instruction (*M* = 68.94%, *SD* = 18.93%) than indirect exposure (*M* = 43.84%, *SD* = 23.19%). Their accuracy in both training conditions was significantly greater than chance. For children with DLD, however, there was not a significant effect of training condition; they were similarly accurate in identifying the correct forms of novel words learned via direct instruction (*M* = 29.82%, *SD* = 15.79%) and indirect exposure (*M* = 33.82%, *SD* = 12.31%) and in neither condition did they perform above chance.

When covariates were added to the model, the effect of group was no longer significant, $b = 0.083$, $t(66) = 1.485$, $p = 0.142$. Children with TLD performed higher than chance in the direct condition only while the children with DLD performed higher than chance in the indirect condition only. All other fixed effects remain unchanged (see Supplementary Materials). This suggests that group differences in word form learning (TLD > DLD) are accounted for by differences in cognitive factors (vocabulary, phonological memory, visuospatial memory, and sustained attention) between the groups. Of the four cognitive factors, only children’s phonological memory was a significant predictor of their success in identifying the correct forms of novel words, *b* = 0.139, *t*(66) = 4.679, *p* = <.001. For each 1 SD increase in children's phonological memory, their accuracy in recognizing the forms of the novel words increased by 13.9%.

**Word form retention**

The dependent variable in these analyses is the change in children's accuracy when tested at the 5-minute and 24- to 48-hour delays. Positive values indicate an increase in children's accuracy over time. Children's accuracy in recognizing the forms of novel words was significantly greater when tested after a 24- to 48-hour delay compared to the 5-minute delay (*M* gain = 14.92%, *SD* = 19.66%), *b* = 0.147, *t*(77) = 6.84, *p* = <.001.

There was not a significant effect of group, *b* = 0.046, *t*(77) = 1.063, *p* = 0.291; the gain in the accuracy of form recognition over the retention interval was similar for children in the TLD group (*M* gain= 17.04%, *SD* = 21.54%) and DLD group (*M* gain= 12.27%, *SD* = 16.96%). There was not a significant effect of training condition, *b* =-0.046, *t*(77) = -1.074, *p* = 0.286; the gain in the accuracy of form recognition over the retention interval was similar for words learned via direct instruction (*M* gain = 12.6%, *SD* = 17.98%) and indirect exposure (*M* gain = 17.29%, *SD* = 21.22%). The interaction between group and training condition was not statistically significant, indicating that the effect of training condition on the size of the gain over the retention interval was the same for both groups, *b* = 0.019, *t*(77) = 0.216, *p* = 0.83.

For children with TLD there was not a significant effect of training condition, *b* = -0.037, *t*(77) = -0.57, *p* = 0.57; the gain in the accuracy of form recognition over the retention interval was similar for words learned via direct instruction (*M* gain= 15.15%, *SD* = 17.94%) and indirect exposure (*M* gain = 18.84%, *SD* = 24.77%). The gain in accuracy was statistically significant both for the direct [*b* = 0.096, *t*(77) = 2.478, *p* = 0.015] and indirect condition [*b* =

0.152, $t(77) = 3.692$, $p = <.001$]. For children with DLD there also was not a significant effect of training condition, $b = -0.055$, $t(77) = -0.979$, $p = 0.331$; the gain in the accuracy of form recognition over the retention interval was similar for words learned via direct instruction (*M* gain= 9.65%, *SD* = 18.06%) and indirect exposure (*M* gain= 15.2%, *SD* = 15.66%). The gain in the accuracy of form recognition over the retention interval was statistically significant in both for the direct [$b = 0.096$, $t(77) = 2.478$, $p = 0.015$] and indirect condition [$b = 0.152$, $t(77) = 3.692$, $p = <.001$].

When covariates were added to the model, children's vocabulary size, phonological memory, visuospatial memory, and sustained attention did not predict variability in how much their accuracy in recognizing the forms of the novel words changed over the delay. All other fixed effects remain unchanged.

# Discussion

In this study, we found that isolating the intended referent and explicitly identifying the goal of learning new words improved learning for children with DLD. These improvements resulting from direct instruction compared to indirect exposure were observed for most aspects of word learning and were similar in magnitude to their peers with TLD (Aim 1). Children were able to retain detailed phonological representations of the new word forms over a 24- to 48-hour delay; this ability was similar for children with DLD and TLD and was unaffected by direct instruction (Aim 2). Finally, individual differences in children's phonological memory accounted for heterogeneity amongst children in most aspects of word learning and accounted for the greater success in word learning by children with TLD than children with DLD (Aim 3). We examine each of these aims in greater detail for each aspect of word learning.

## Learning word-referent links

In laboratory settings, word learning outcomes are commonly measured as the ability to identify a referent when hearing its label. We found lower accuracy on this measure of word-to-referent linking for learners with DLD than for learners with TLD. These group differences were accounted for by individual differences in our cognitive measures, in particular phonological memory. The size of the group difference observed here between children (~13%) is similar to the difference between adults (~18%) observed by McGregor, Eden, and colleagues (2020). These findings reveal continuity in word learning difficulties, which persist throughout development (learners with DLD lag behind their peers with TLD both as children and adults) and across learning environments (learners with DLD lag behind their peers with TLD following both direct instruction and indirect exposure).

We also found higher accuracy for word-to-referent links learned via direct instruction than indirect exposure and the benefit of direct instruction held for both groups. This matches the pattern of results observed in prior research involving adults with and without DLD (Coutanche & Thompson-Schill, 2014; McGregor et al., 2020). There were several differences between our training conditions which may have affected children's success in linking words to referents. First, indirect exposure increased the number of images presented on the screen (from 1 to 2), which increased the processing/cognitive load for each trial. Second, the presence of a second referent (although familiar) increased competition by requiring children to identify which image was the intended referent (e.g., Halberda, 2006; Markman & Wachtel, 1988). Third, the instructions provided to the child changed task demands from primarily memory (direct instruction) to attentional (indirect exposure), which affects children's behavior (e.g., Csibra & Gergely, 2009). With the current methods, it is not possible to determine the extent to which each of these factors contributed to children's success in word learning. To the extent that these factors are dissociable (i.e., do not frequently co-occur in natural language), identifying the relative contribution of each factor is an important direction for future research.

While it may not be surprising that direct instruction is more effective than indirect exposure, these results are nevertheless important because they support language interventions using direct instruction to help children with DLD learn words. It is important to note that this does not mean that direct instruction is universally better. Learning via indirect exposure may be a slower, yet crucial, aspect of word learning (McMurray, Horst, & Samuelson, 2012). In fact, the added complexity from indirect exposure (e.g., the need to reject familiar objects as potential referents) sometimes improves learning outcomes (Zosh, Brinster, & Halberda, 2013). Increased competition, but not too much competition, may improve learning by creating the ideal balance in learning difficulty – not too easy so as to be boring, but not too hard so as to be overwhelming (Horst, Scott, & Pollard, 2010; Kidd, Piantadosi, & Aslin, 2012). That said, "learning" a new word involves much more than just identifying a referent when it is named. We turn next to our results investigating how children form more detailed representations of both the referent and word form.

**Learning semantic categories**

In addition to associating a specific referent with a novel word, learners may be encoding information about the referent itself. Here we focused on the extent to which children encoded the superordinate categories (e.g., bird, mammal, insect, fruit) for referents and linked this information to the novel words. Similar to link recognition, we found better accuracy for learners with TLD than DLD, but in contrast, these group differences could not be accounted for by our constellation of cognitive measures. We again found better accuracy for words learned via direct instruction than indirect exposure. The comparison of the effect of training between groups, however, yielded mixed results. For children with TLD, direct instruction unambiguously improved their success in learning superordinate categories compared to indirect exposure. Moreover, the size of this improvement (~18%) was similar to the improvement in link recognition (~19%). For children with DLD, the improvement in semantic category learning (~9%) was smaller and not statistically significant. Children with DLD, however, learned superordinate categories only from direct instruction, but not indirect exposure. Taken together, these results highlight the importance of including additional supports in vocabulary interventions to help children with DLD learn semantic information, like providing explicit definitions for words (e.g., Nash & Donaldson, 2005).

Superordinate categories are just one of the many types of semantic information children must learn when they encounter new words. For instance, apples are fruits, but they are also edible, typically red in color, grow on trees, etc. Moreover, it is not clear to what extent children's ability to make post hoc judgements about superordinate category membership in our task is associated with children's ability to embed this information into semantic networks. For instance, a newly-learned word for an insect could semantically prime lexical recognition of the word "ant" (e.g., Coutanche and Thompson-Schill, 2014). Additionally, children may associate category-specific knowledge (e.g., insects lay eggs) with the newly learned word (e.g., Gelman & O'Reilly, 1988). Tests of semantic knowledge are challenging to create and can be difficult to replicate (e.g., McGregor, Eden et al., 2020). This aspect of word learning is often overlooked and therefore an important direction for future research.

**Sustained Attention**

We found that individual differences in children's sustained attention predicted their accuracy in identifying the semantic categories of the novel words. This correlation, however, was opposite our prediction and contrary to prior research (McGregor et al., 2022). We found that children with better sustained attention were worse at identifying semantic categories. This relation, however, was only marginally significant when children with ADHD were removed from our sample. These results should be interpreted with caution because measures of sustained attention are not often included in word learning research and children with ADHD are often excluded from DLD research. In other words, replication is needed to be certain that this correlation is not spurious and extended research is needed to better elucidate the similarities and differences between children with DLD only and those whose DLD occurs alongside other neurodevelopmental challenges.

**Learning word forms**

As in most research on word learning, the novel words in the current study were intentionally chosen to be phonologically distinct. As a consequence, children did not need detailed phonological representations of words to succeed on link and semantic category trials. For example, children do not have to remember the exact combination of phonemes to correctly identify the target *kaktub* when the foil referents are *melig* and *zimp*. We therefore included trials that measured children's ability to discriminate trained words like *kaktub* from foils that were phonological neighbors like *kakpub* and *kakfub.* We found better accuracy for learners with TLD than DLD, which was accounted for by individual differences in our cognitive measures, in particular phonological memory. The gap in average accuracy between groups (TLD > DLD) is larger for word form recognition (~23%) than both link recognition (~13%) and semantic category recognition (~15%). Prior research indicates that novel word learning involves separate phonological and semantic factors (Gray et al., 2020) and that, at the early stages of learning, encoding phonological information is more challenging than encoding semantic information for children and adults with DLD (Gray, 2004; Jackson et al., 2021; McGregor, Arbisi-Kelm, et al., 2020; McGregor, Licandro et al., 2013). As previously discussed, semantic knowledge entails much more than recognizing the referent (and its superordinate category), therefore, we might find that individuals with DLD struggle more with semantic learning during the latter stages of word learning (McGregor, Oleson, et al., 2013).

In tests of nonword repetition, the deficit between children with DLD and their peers with TLD is larger for longer words (three to four syllables) compared to shorter words (one to two syllables; e.g., Graf Estes, Evans, & Else-Quest, 2007). The novel words in the current experiment were relatively short, consisting of one or two syllables. Nevertheless, we observed that children with DLD were less successful in encoding detailed phonological representations of words than their peers with TLD. With longer novel words, we expect that the (already large) gap in performance would widen further.

We found better accuracy for word forms learned via direct instruction than indirect exposure. This benefit from direct instruction only occurred for children with TLD and not for children with DLD. In fact, children with DLD were unable to discriminate trained words from close phonological foils after a 5-minute delay even with direct instruction. This failure is particularly striking, because it illustrates how the ability to succeed on several metrics of word learning (i.e., link and semantic category recognition) despite impoverished phonological representations of word forms, can mask the need for further intervention for children with DLD.

Children with DLD may struggle to identify the correct forms of novel words from close phonological foils for a variety of reasons. Failure in our task could result from difficulties in perception (e.g., discriminating the subtle differences between foils), in encoding (e.g., identifying the phonemes that combine to form the word), and in retention (e.g., maintaining the phonological representations over the 5-minute delay). Without additional measures (e.g., a same-different task with familiar words), it is difficult to discern to what extent each of these factors contributed to their failures. An extensive body of research investigating why children with DLD struggle in non word repetition tasks, however, suggests that both perception and encoding may play a role (see Coady & Evans, 2008 for review). Children with DLD are frequently reported to have deficits in auditory discrimination and speech perception (e.g., Brosseau-Lapré et al., 2020; Kujala & Leminen, 2017; Quam et al., 2021; Ziegler et al., 2011). The improvements in accuracy that we observed over the 24- to 48-hour delay (see next section) suggest that retention is not a problem for children with DLD. In fact, their accuracy after the delay was significantly greater than chance (see Supplementary Materials available via OSF). These findings are particularly striking for several reasons. First, they reveal that children with DLD can succeed in our word form discrimination task. Second, they indicate that the lack of an effect of training (direct = indirect) was not due to floor effects. Third, they serve as a reminder to interpret chance performances with caution. Without any intervening exposure, children with DLD could only succeed in identifying the correct forms of novel words after a 24- to 48-hour delay if they had learned something during training. This learning, however, was not evident when they were tested after a 5-minute delay. Especially when dealing with disordered populations, we may be too quick to interpret null results as a failure to learn. As the results here demonstrate, this is not always the case. With the consolidation of memory enabled by time and sleep, learning may become evident (e.g., Dumay & Gaskell, 2007).

**Retaining word forms**

We found that children with DLD were more accurate in identifying the correct forms of novel words when tested after a 24- to 48-hour delay compared to 5-minute delay. In some cases, performance after a retention interval is similar to performance immediately after training (e.g., McGregor, Gordon et al., 2017, study 2). In some cases, like the current one, it improves and, in other cases still, it declines (e.g., Jackson et al., 2021). These differences may be attributable to a variety of methodological differences in both training (i.e., number of words to be learned and the number, timing, and spacing of exposures during training) and testing (i.e., number of phonological foils, how phonological foils differed from the target). Regardless of whether accuracy improves, decreases, or remains the same, an emerging body of research has consistently demonstrated that both children and adults with DLD are just as successful as their peers with TLD in retention (Bishop & Hsu, 2015; Gordon et al., 2021; Haebig et al., 2019; Leonard et al., 2019; McGregor, Licandro et al., 2013; McGregor, Gordon et al., 2017; McGregor, Arbisi-Kelm et al., 2017; Nash & Donaldson, 2005). Here too we found equal rates of retention for children with DLD and TLD. We also found equal rates of retention for words learned via direct instruction and indirect exposure. These results are also consistent with the aforementioned prior research in that the factors which affect children's success in encoding during word learning do *not* affect their success in retention. This suggests distinct cognitive mechanisms support these different stages of word learning.

**Clinical Implications**

Children with TLD readily learn words from indirect exposure. Many children with DLD need language interventions that involve direct instruction. In exploratory analyses (see Supplementary Materials), we compared word learning outcomes for children with DLD who received direct instruction and children with TLD who received indirect exposure These cross-condition analyses revealed that the additional supports provided by direct instruction (i.e., isolating the intended referent and explicitly identifying the goal of learning a new), were sufficient for children with DLD to achieve similar levels of success in link and semantic category recognition as their peers with TLD who only received indirect exposure. Although direct instruction in the current experiment did not include information to help children encode (or even encourage them to attend to) superordinate categories, it nevertheless reduced learning demands by decreasing the amount of visual information presented and eliminating the need for children to respond to answer a question using inference.

The cross-condition analyses indicated that direct instruction was insufficient to help children with DLD achieve similar levels of success in word form encoding compared to their peers with TLD who only received indirect exposure. In fact, even with direct instruction, children with DLD struggled to form detailed phonological representations of the novel word forms. Difficulty in developing detailed and stable phonological representations of words has broader, cascading impacts on development. Computational work demonstrates how deficits in phonological representations impact other cognitive skills, including working memory capacity (Jones & Westermann, 2022) and reading (Harm & Seidenberg, 2004). Children's language ability, measured in part by how many words they know, predicts academic achievement (Pace et al., 2019). These consequences highlight the importance of helping students succeed in learning the forms of new words.

Prior research has shown that variability in word learning success between children with DLD is associated with differences in their performance across a variety of cognitive measures, including vocabulary, phonological memory, visuospatial memory, and sustained attention (Jackson et al., 2019; 2021; Kan & Windsor, 2010; McGregor et al., 2022). Children's performance is often correlated across these different measures indicating underlying constructs that support their general success in these tasks. By including all four measures in our models, we were able to identify the extent to which each one accounts for *unique* variance in word learning success. We found that only children's phonological memory predicted how accurate they were at both linking words to their referents and forming detailed phonological representations of the words. These findings have important implications for vocabulary interventions. First, they suggest that the ability to actively maintain representations of newly heard words is a primary limiting factor of children's success in word learning. Thus, an important goal for vocabulary interventions is to scaffold the learning environment to help all children succeed in encoding this information. Second, these findings indicate that not all children with DLD will struggle with all aspects of word learning. Children with DLD who have strong phonological memory may require less support with learning detailed phonological representations of new words, while still requiring support with building rich semantic representations.

Based upon the strength of the prior literature and the current study, there is little doubt that during the early stages of learning many children with DLD (particularly those with poor phonological memory) will need significantly more support to learn the forms of new words. Vocabulary interventions, however, are often not tailored to meet this specific need for children with DLD. Recent surveys of Speech Language Pathologists and recordings of their sessions reveal that they most commonly use techniques that are focused on teaching children the meanings of words, but seldom use techniques focused on phonology or orthography (Justice et al., 2014; Steele, 2020). Other techniques, including increasing the number of exposures (McGregor, Arbisi-Kelm, Eden, & Oleson, 2020), testing learners' ability to recall the names of referents throughout learning (i.e., repeated spaced retrieval; Haebig et al., 2019; Leonard et al., 2019a; 2019b; 2020; McGregor, Gordon et al., 2017), and explicitly asking learners to monitor words for the presence of specific sounds (McGregor, Arbisi-Kelm et al., 2017), have all been shown to improve word form learning for children and adults with DLD.

**Future Directions**

In the current study, we systematically investigated the extent to which direct instruction facilitates different aspects of word learning. As in most research on word learning, we focused on children's ability to learn nouns that label concrete objects. Learning words for abstract concepts (e.g., emotions, thoughts) is more difficult than learning words for concrete objects (e.g., de Groot & Keijzer, 2000). This concrete vs. abstract gap is larger for children with DLD than TLD (McGregor et al., 2012). The components of direct instruction that were the focus of the current project – isolating the referent and explicitly identifying the goal of learning – are unlikely to help children learn abstract words. Other strategies, like explicit definitions, however, may help (e.g., Nash & Donaldson, 2004). Given the transition throughout elementary school from perceptually- to linguistically-acquired word meanings (Wauters et al., 2003), it is critical that future research investigate ways in which vocabulary interventions can help children with DLD learn abstract words.

**Conclusions**

For words labeling concrete objects, we found that isolating the referent and explicitly identifying the goal of learning were sufficient to help children with DLD in multiple aspects of word learning; it unequivocally improved their ability to link words to their referents and, to some extent, also improved their ability to generalize words to their superordinate semantic categories. The results support the use of direct instruction in vocabulary interventions for children with DLD. Nevertheless, the children with DLD struggled to learn the forms of the new words and direct instruction was no more effective than indirect exposure for that aspect of word learning, at least with the number of exposures provided here. Additional exposures and, perhaps, supplemental ways to emphasize word forms and practice their productions will be required. For many children with DLD, word forms do not come along 'for free.' Although isolating the referent and specifying the learning goal are enough to help children with DLD learn referents, this stripped-down version of direct instruction is not enough to support their word form learning.

## Acknowledgements

We would like to thank the participants and their families for the generosity. We would also like to thank Katie Gordon for her feedback on the manuscript and Jake Oleson for his advice on statistical analyses. This research was supported by the National Institutes of Health Grant R01 DC011742-07 awarded to the second author. Some of the data in this project were presented as a flash talk at the biennial meeting of the Society for Research in Child Development in 2021.

## Data Availability Statement

The analysis code, raw data, model specifications, and full statistical results are available via Open Science Framework (https://osf.io/26djx/?view_only=3c133119e79144e4896439ab3227e0b1).

## References

Alloway, T. P. (2007). *Automated Working Memory Assessment*. Pearson Assessment.

Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review*, *14*(5), 919–924. https://doi.org/10.3758/BF03194122

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). Guilford Press.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*(3), 498–520. https://doi.org/10.1037/0022-0663.93.3.498

Brosseau-Lapré, F., Schumaker, J., & Kluender, K. R. (2020). Perception of Medial Consonants by Children With and Without Speech and Language Disorders: A Preliminary Study. *American Journal of Speech-Language Pathology*, *29*(2), 883–889. https://doi.org/10.1044/2020_AJSLP-19-00062

Cable, A. L., & Domsch, C. (2011). Systematic review of the literature on the treatment of children with late language emergence. *International Journal of Language & Communication Disorders*, *46*(2), 138-154. https://doi.org/10.3109/13682822.2010.487883

Carey, S. (2010). Beyond Fast Mapping. *Language Learning and Development*, *6*(3), 184–205. https://doi.org/10.1080/15475441.2010.484379

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, *43*(1), 1–40. https://doi.org/10.1080/13682820601116485

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, *143*(6), 2296–2303. https://doi.org/10.1037/xge0000020

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching Vocabulary During Shared Storybook Readings: An Examination of Differential Effects. *Exceptionality*, *12*(3), 145–162. https://doi.org/10.1207/s15327035ex1203_3

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945. https://doi.org/10.1037/0012-1649.33.6.934

de Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

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

Dockrell, J. E., Lindsay, G., Connelly, V., & Mackie, C. (2007). Constraints in the Production of Written Text in Children with Specific Language Impairments. *Exceptional Children*, *73*(2), 147–164. https://doi.org/10.1177/001440290707300202

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-

Analysis. *Review of Educational Research*, *71*(3), 393–447. https://doi.org/10.3102/00346543071003393

Erickson, L. C., Thiessen, E. D., Godwin, K. E., Dickerson, J. P., & Fisher, A. v. (2015). Endogenously and exogenously driven selective sustained attention: Contributions to learning in kindergarten children. *Journal of Experimental Child Psychology*, *138*, 126–134. https://doi.org/10.1016/j.jecp.2015.04.011

Ezell, H. K., & Justice, L. M. (2005). Shared storybook reading: Building young children's language and emergent literacy skills. Baltimore, MD: Paul H. Brookes.

Farrell Pagulayan, K., Busch, R., Medina, K., Bartok, J., & Krikorian, R. (2006). Developmental normative data for the Corsi Block-Tapping task. *Journal of Clinical and Experimental Neuropsychology*, *28*(6), 1043–1052. https://doi.org/10.1080/13803390500350977

Fisher, A., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention in 3- to 5-year-old children: Evidence from a new paradigm. *Journal of Experimental Child Psychology*, *114*(2), 275–294. https://doi.org/10.1016/j.jecp.2012.07.006

Gelman, S. A., & O'Reilly, A. W. (1988). Children's inductive inferences within superordinate categories: The role of language and category structure. *Child Development*, *59*(4), 876–887.

Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, *20*(6), 642–651. https://doi.org/10.1017/S1355617714000411

Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., Wallner-

Allen, K., Golinkoff, R. M., Gleason, J. B., Hirsh-Pasek, K., Adams, M. J., & Weintraub, S. (2013). IV. NIH TOOLBOX COGNITION BATTERY (CB): MEASURING LANGUAGE (VOCABULARY COMPREHENSION AND READING DECODING). *Monographs of the Society for Research in Child Development*, *78*(4), 49–69. https://doi.org/10.1111/mono.12034

Gillam, R. B., & Pearons, N. A. (2004). *TNL: Test of Narrative Language*. Pro-ed.

Gillam, R. B., & Pearson, N. A. (2017). *TNL-2: Test of Narrative Language*. Pro-ed.

Girolametto, L., Pearce, P. S., & Weitzman, E. (1996). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1274-1283. https://doi.org/10.1044/jshr.3906.1274

Gordon, K. R., Storkel, H. L., Lowry, S. L., & Ohlmann, N. B. (2021). Word Learning by Preschool-Age Children With Developmental Language Disorder: Impaired Encoding and Robust Consolidation During Slow Mapping. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4250–4270. https://doi.org/10.1044/2021_JSLHR-21-00046

Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the Nonword Repetition Performance of Children With and Without Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Gray, S., Lancaster, H., Alt, M., Hogan, T. P., Green, S., Levy, R., & Cowan, N. (2020). The Structure of Word Learning in Young School-Age Children. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1446–1466. https://doi.org/10.1044/2020_JSLHR-19-00186

Gray, S. I., Levy, R., Alt, M., Hogan, T. P., & Cowan, N. (2022). Working Memory Predicts

New Word Learning Over and Above Existing Vocabulary and Nonverbal IQ. *Journal of Speech, Language, and Hearing Research*, *65*(3), 1044–1069. https://doi.org/10.1044/2021_JSLHR-21-00397

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., Kueser, J. B., Souto, S., Krok, W., & Weber, C. (2019). Retrieval-Based Word Learning in Young Typically Developing Children and Children With Development Language Disorder II: A Comparison of Retrieval Schedules. *Journal of Speech, Language, and Hearing Research*, *62*(4), 944–964. https://doi.org/10.1044/2018_JSLHR-L-18-0071

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: \ Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review*, *111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Henry, L. A. (2001). How does the severity of a learning disability affect working memory performance? *Memory*, *9*(4–6), 233–247. https://doi.org/10.1080/09658210042000085

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language*, *63*(1), 100–116. https://doi.org/10.1016/j.jml.2010.02.003

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/10.1080/15250000701795598

Horst, J. S., Scott, E. J., & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science*, *13*(5), 706–713. https://doi.org/10.1111/j.1467-7687.2009.00926.x

Jackson, E., Leitao, S., Claessen, M., & Boyes, M. (2019). Fast mapping short and long words: Examining the influence of phonological short-term memory and receptive vocabulary in children with developmental language disorder. *Journal of Communication Disorders*, *79*, 11–23. https://doi.org/10.1016/j.jcomdis.2019.02.001

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

Jones, S. D., & Westermann, G. (2022). Under-resourced or overloaded? Rethinking working memory deficits in developmental language disorder. *Psychological Review*. https://doi.org/10.1037/rev0000338

Justice, L. M., Schmitt, M. B., Murphy, K. A., Pratt, A., & Biancone, T. (2014). The 'robustness' of vocabulary intervention in the public schools: targets and techniques employed in speech-language therapy. *International Journal of Language & Communication Disorders*, *49*(3), 288–303. https://doi.org/10.1111/1460-6984.12072

Kan, P. F., & Windsor, J. (2010). Word Learning in Children With Primary Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *53*(3), 739–756. https://doi.org/10.1044/1092-4388(2009/08-0248)

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kujala, T., & Leminen, M. (2017). Low-level neural auditory discrimination dysfunctions in specific language impairment—A review on mismatch negativity findings. *Developmental Cognitive Neuroscience*, *28*, 65–75. https://doi.org/10.1016/j.dcn.2017.10.005

Lancaster, H. S., & Camarata, S. (2019). Reconceptualizing developmental language disorder as a spectrum disorder: issues and evidence. *International Journal of Language & Communication Disorders*, *54*(1), 79–94. https://doi.org/10.1111/1460-6984.12433

Leach, L., & Samuel, A. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

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

Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S., Weber, C., Kueser, J. B., & Haebig, E. (2019). Adjective Learning in Young Typically Developing Children and Children With Developmental Language Disorder: A Retrieval-Based Approach. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4433–4449. https://doi.org/10.1044/2019_JSLHR-L-19-0221

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., Souto, S., Kueser, J. B., & Krok, W. (2019). Retrieval-Based Word Learning in Young Typically Developing Children and Children With Developmental Language Disorder I: The Benefits of Repeated Retrieval. *Journal of Speech, Language, and Hearing Research*, *62*(4), 932–

943. https://doi.org/10.1044/2018_JSLHR-L-18-0070

Lund, E., & Douglas, W. M. (2016). Teaching Vocabulary to Preschool Children With Hearing Loss. *Exceptional Children*, *83*(1), 26–41. https://doi.org/10.1177/0014402916651848

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

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

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

McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Oleson, J. (2020). The word learning profile of adults with developmental language disorder. *Autism & Developmental Language Impairments*, *5*, 239694151989931. https://doi.org/10.1177/2396941519899311

McGregor, K. K., Berns, A. J., Owen, A. J., Michels, S. A., Duff, D., Bahnsen, A. J., & Lloyd, M. (2012). Associations Between Syntax and the Lexicon Among Children With or Without ASD and Language Impairment. *Journal of Autism and Developmental Disorders*, *42*(1), 35–47. https://doi.org/10.1007/s10803-011-1210-4

McGregor, K., & Duff, D. (2015). Promoting Diverse and Deep Vocabulary Development. In T. Ukrainetz (Ed.), *School-age language intervention: Evidence-based practices* (1st ed., pp. 247–278). Pro Ed.

McGregor, K. K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2020). The Fast-Mapping Abilities of Adults With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(9), 3117–3129. https://doi.org/10.1044/2020_JSLHR-19-00418

McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding Deficits Impede Word Learning and Memory in Adults With Developmental Language Disorders. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2891–2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031

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), 1845–1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

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

McGregor, K. K., Smolak, E., Jones, M., Oleson, J., Eden, N., Arbisi-Kelm, T., & Pomper, R. (2022). What Children with Developmental Language Disorder Teach Us About Cross-Situational Word Learning. *Cognitive Science*, *46*(2). https://doi.org/10.1111/cogs.13094

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Nash, M., & Donaldson, M. L. (2005). Word Learning in Children With Vocabulary Deficits.

*Journal of Speech, Language, and Hearing Research*, *48*(2), 439–458. https://doi.org/10.1044/1092-4388(2005/030)

Noble, C., Sala, G., Peter, M., Lingwood, J., Rowland, C., Gobet, F., & Pine, J. (2019). The impact of shared book reading on children's language skills: A meta-analysis. *Educational Research Review*, *28*, 100290. https://doi.org/10.1016/j.edurev.2019.100290

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

Pinheiro, J., Bates, D., DebRoy Saikat, Sarkar, D., & R Core Team. (2022). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-155. Retreived from https://CRAN.R-project.org/package=nlme

Pollard-Durodola, S. D., Gonzalez, J. E., Simmons, D. C., Kwok, O., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children*, *77*(2), 161-183. https://doi.org/10.1177/001440291107700202

Quam, C., Cardinal, H., Gallegos, C., & Bodner, T. (2021). Sound discrimination and explicit mapping of sounds to meanings in preschoolers with and without developmental language disorder. *International Journal of Speech-Language Pathology*, *23*(1), 26–37. https://doi.org/10.1080/17549507.2020.1750701

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Redmond, S. M. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics*, *19*(2), 109–127. https://doi.org/10.1080/02699200410001669870

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the Signal by Adding Noise: The Role of Noncontrastive Phonetic Variability in Early Word Learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

RStudio Team (2020). *RStudio: Integrated Development for R.* RStudio, PBC, Boston, MA. Retrived from http://www.rstudio.com

Sciberras, E., Mueller, K. L., Efron, D., Bisset, M., Anderson, V., Schilpzand, E. J., Jongeling, B., & Nicholson, J. M. (2014). Language Problems in Children With ADHD: A Community-Based Study. *Pediatrics*, *133*(5), 793–800. https://doi.org/10.1542/peds.2013-3355

Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2020). Sustained Attention in Developmental Language Disorder and Its Relation to Working Memory and Language. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4096–4108. https://doi.org/10.1044/2020_JSLHR-20-00265

Steele, S. C. (2020). Vocabulary Intervention: A National Survey of School-Based Speech–Language Pathologists. *Communication Disorders Quarterly*, *41*(3), 151–161. https://doi.org/10.1177/1525740119827008

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation.

*Behavioral and Brain Sciences*, *28*(1), 51–64. https://doi.org/10.1017/S0140525X05000026

Wauters, L. N., Tellings, A. E. J. M., van Bon, W. H. J., & van Haaften, A. W. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, *24*(3), 385–406. https://doi.org/10.1017/S0142716403000201

Wechsler, D. (2011). *WASI-II: Wechsler Abbreviated Scales of Intelligence*. The Psychological Corporation.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New Work. ISBN 978-3-319-24277-4. Retrieved from https://ggplot2.tidyverse.org

Wickham, H., & Girlich, M. (2022). *tidyr:Tidy Messy Data.* Retrieved from https://tidyr.tidyverse.org, https://github.com/tidyverse/tidyr

Wojcik, E. H., Zettersten, M., & Benitez, V. L. (2022). The map trap: Why and how word learning research should move beyond mapping. *WIREs Cognitive Science*. https://doi.org/10.1002/wcs.1596

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2011). Noise on, voicing off: Speech perception deficits in children with specific language impairment. *Journal of Experimental Child Psychology*, *110*(3), 362–372. https://doi.org/10.1016/j.jecp.2011.05.001

Zosh, J. M., Brinster, M., & Halberda, J. (2013). Optimal Contrast: Competition Between Two Referents Improves Word Learning. *Applied Developmental Science*, *17*(1), 20–28. https://doi.org/10.1080/10888691.2013.748420

## Tables and Figures

**Table 1**

*Comparisons of test scores and demographic information between diagnostic groups.*

| | | DLD (*n* = 36) | | TLD (*n* = 45) | | |
|---|---|---|---|---|---|---|
| **Domain** | **Measure** | **Mean (SD)** | **Range** | **Mean (SD)** | **Range** | **Sig.** |
| Narrative | TNL | 81.06 (8.1) | 61-91 | 111.33 (8.98) | 94-127 | * |
| Nonverbal IQ | WASI | 89.03 (10.71) | 71-116 | 107.47 (10.41) | 86-130 | * |
| Receptive Vocabulary | PVT | 92.42 (14.55) | 75-125 | 110.56 (14.82) | 78-140 | * |
| Phonological STM | NWR | 67.69 (13.15) | 29-88 | 81.38 (8.58) | 54-96 | * |
| Phonological WM | BDT | 5.5 (4.07) | 0-15 | 10.64 (3.81) | 6-24 | * |
| Visuospatial STM | Corsi | 16.9 (4.5) | 6-28 | 21.27 (2.92) | 15-28 | * |
| Visuospatial WM | OOO | 5.32 (2.45) | 1-13 | 9.38 (4.02) | 5-20 | * |
| Sustained Attention | Track-It | 0.68 (0.3) | 0-1 | 0.86 (0.21) | 0-1 | * |
| Age | in months | 86.81 (5.64) | 74-96 | 86.6 (4.58) | 76-98 | |
| Maternal education | in years | 14.36 (2.67) | 10-20 | 16.87 (2.18) | 12-22 | * |
| **Gender** | | **N** | **%** | **N** | **%** | |
| Male | | 21 | 58.3 | 20 | 44.4 | |
| Female | | 15 | 41.7 | 25 | 55.6 | |
| **Race** | | **N** | **%** | **N** | **%** | |
| Black or African American | | 5 | 13.9 | 1 | 2.2 | |
| More than one race | | 4 | 11.1 | 6 | 13.3 | |
| White | | 26 | 72.2 | 38 | 84.4 | |
| Did not reply | | 1 | 2.8 | 0 | 0.0 | |
| **Ethnicity** | | **N** | **%** | **N** | **%** | |
| Hispanic or Latino | | 1 | 2.8 | 0 | 0.0 | |
| Not Hispanic or Latino | | 28 | 77.8 | 39 | 86.7 | |
| Did not reply | | 7 | 19.4 | 6 | 13.3 | |

Note: scores on the Test of Narrative Language (TNL; Gillam & Pearson 2004; 2017), Wechsler Abbreviated Scales of Intelligence, 2nd edition (WASI; Wechsler, 2011), and Picture Vocabulary

Test from the NIH Toolbox (PVT; Gershon et al., 2013) are standard scores with a normative mean of 100 and a standard deviation of 15. Scores on the Nonword Repetition task (NWR; Dollaghan & Campbell, 1998) are the number of phonemes children correctly produced (maximum of 96). Scores on the Backwards Digit Test (BDT; Alloway et al., 2007) are the number of sequences children correctly produced (maximum of 36). Scores on the Corsi Block-Tapping Task (Corsi; Farrell et al., 2006) are the number of correct sequences children correctly recalled (maximum of 45). Scores on the Odd-One-Out task (OOO; Henry, 2001) are the number of sequences children correctly recalled (maximum of 24). Scores on the Track-It test (Erickson et al., 2015; Fisher et al., 2013) are the proportion of heterogenous trials correct after excluding trials for which children failed the memory check. All between-group differences are statistically significant ($t$'s $> 2.8$, $p$'s $< .01$) except for age.

**Table 2**

*Protocol schedule*

| **Visit 1** | **Visit 2** | **Visit 3** |
|---|---|---|
| Parent Consent, Child Assent, HIPAA Forms | Form recognition (3AFC) | Cross-Situational (CS) learning |
| Novel word training | Category recognition (4AFC) | 5-min break |
| 5-min break | Track-It (A or B) | CS Form recognition (3AFC) |
| Form recognition (3AFC) | 2-min break | CS Link recognition (3AFC) |
| Link recognition (3AFC) | Track-It (A or B) | Odd One Out |
| Same-Not Same (NIH Toolbox) | Corsi Blocks | Nonword Repetition |
| Backward Digit | Picture Vocabulary Test (NIH Toolbox) | |

Note: tasks included in the current project are listed in black; tasks that are reported elsewhere are listed in gray.

**Table 3**

*Comparison of fixed effects across different model criteria*

| | link | | | semantic category | | | word form | | |
|---|---|---|---|---|---|---|---|---|---|
| | no covariates | | covariates | no covariates | | covariates | no covariates | | covariates |
| **fixed effect** | **all trials** | **exclusions** | **all trials** | **all trials** | **exclusions** | **all trials** | **all trials** | **exclusions** | **all trials** |
| Intercept | 0.16 | 0.18 | 0.16 | 0.13 | 0.14 | 0.14 | 0.11 | 0.16 | 0.12 |
| DLD Direct | 0.2 | 0.23 | 0.22 | 0.08 | 0.07 | 0.11 | | | |
| DLD Indirect | | | | | | | | | 0.1 |
| TLD Direct | 0.33 | 0.35 | 0.31 | 0.31 | 0.32 | 0.3 | 0.36 | 0.42 | 0.27 |
| TLD Indirect | 0.13 | 0.14 | 0.13 | 0.13 | 0.13 | 0.12 | 0.11 | 0.16 | |
| Group | 0.14 | 0.13 | | 0.16 | 0.17 | 0.15 | 0.25 | 0.25 | |
| Training | 0.21 | 0.21 | 0.21 | 0.11 | 0.11 | 0.13 | 0.11 | 0.1 | 0.08 |
| DLD | 0.22 | 0.22 | 0.24 | | | | | | |
| TLD | 0.2 | 0.21 | 0.19 | 0.18 | 0.19 | 0.18 | 0.25 | 0.26 | 0.21 |
| Group:Training | | | | | 0.16 | | 0.29 | 0.31 | 0.26 |
| Vocab | | | | | | | | | |
| Phono. Memory | | | 0.09 | | | | | | 0.14 |
| Visuo. Memory | n/a | | | n/a | | | n/a | | |
| Sustained Attention | | | | | | -0.18 | | | |

Note: Models varied based on whether trials were (exclusions) or were not (all trials) filtered based on response latency criteria and whether children's performance on the cognitive measures was (covariates) or was not (no covariates) included as fixed effects.

**Table 4**

*Evaluation of factors determining accuracy in linking words to referents*

| **fixed effect** | **b** | **se** | **95% CI** | | **t** | **p** | **sig** |
|---|---|---|---|---|---|---|---|
| | | | ***LL*** | ***UL*** | | | |
| Intercept | 0.159 | 0.023 | 0.113 | 0.204 | 6.949 | <.001 | * |
| TLD Direct | 0.329 | 0.043 | 0.243 | 0.416 | 7.569 | <.001 | * |
| TLD Indirect | 0.13 | 0.043 | 0.045 | 0.215 | 3.06 | 0.003 | * |
| DLD Direct | 0.196 | 0.047 | 0.103 | 0.289 | 4.196 | <.001 | * |
| DLD Indirect | -0.021 | 0.049 | -0.12 | 0.077 | -0.428 | 0.67 | |
| Group | 0.142 | 0.046 | 0.051 | 0.233 | 3.112 | 0.003 | * |
| Training | 0.208 | 0.046 | 0.117 | 0.299 | 4.561 | <.001 | * |
| TLD | 0.199 | 0.061 | 0.078 | 0.32 | 3.271 | 0.002 | * |
| DLD | 0.217 | 0.068 | 0.082 | 0.353 | 3.194 | 0.002 | * |
| Group:Training | -0.019 | 0.091 | -0.2 | 0.163 | -0.203 | 0.84 | |

Residual Standard Error: 0.204
Degrees of freedom: 81 total; 77 residual

**Table 5**

*Evaluation of factors determining accuracy in linking words to semantic categories*

| fixed effect | b | se | 95% CI | | t | p | sig |
|---|---|---|---|---|---|---|---|
| | | | *LL* | *UL* | | | |
| Intercept | 0.159 | 0.023 | 0.113 | 0.204 | 6.949 | <.001 | * |
| TLD Direct | 0.307 | 0.041 | 0.225 | 0.388 | 7.491 | <.001 | * |
| TLD Indirect | 0.127 | 0.04 | 0.047 | 0.207 | 3.166 | 0.002 | * |
| DLD Direct | 0.075 | 0.028 | 0.019 | 0.131 | 2.649 | 0.01 | * |
| DLD Indirect | 0.029 | 0.03 | -0.03 | 0.089 | 0.989 | 0.326 | |
| Group | 0.142 | 0.046 | 0.051 | 0.233 | 3.112 | 0.003 | * |
| Training | 0.208 | 0.046 | 0.117 | 0.299 | 4.561 | <.001 | * |
| TLD | 0.18 | 0.057 | 0.066 | 0.294 | 3.142 | 0.002 | * |
| DLD | 0.045 | 0.041 | -0.036 | 0.127 | 1.102 | 0.274 | |
| Group:Training | -0.019 | 0.091 | -0.2 | 0.163 | -0.203 | 0.84 | |

Residual Standard Error: 0.123
Degrees of freedom: 81 total; 77 residual

**Table 6**

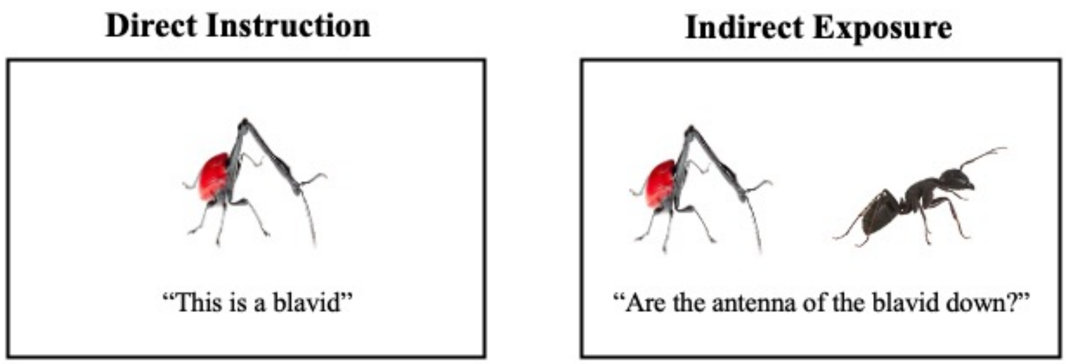
*Evaluation of factors determining accuracy in identifying the correct forms of novel words*

| **fixed effect** | **b** | **se** | **95% CI** | | **t** | **p** | **sig** |
|---|---|---|---|---|---|---|---|
| | | | ***LL*** | ***UL*** | | | |
| Intercept | 0.159 | 0.023 | 0.113 | 0.204 | 6.949 | <.001 | * |
| TLD Direct | 0.359 | 0.045 | 0.269 | 0.449 | 7.944 | <.001 | * |
| TLD Indirect | 0.108 | 0.044 | 0.02 | 0.197 | 2.45 | 0.017 | * |
| DLD Direct | -0.032 | 0.033 | -0.097 | 0.033 | -0.971 | 0.335 | |
| DLD Indirect | 0.008 | 0.035 | -0.061 | 0.077 | 0.238 | 0.812 | |
| Group | 0.142 | 0.046 | 0.051 | 0.233 | 3.112 | 0.003 | * |
| Training | 0.208 | 0.046 | 0.117 | 0.299 | 4.561 | <.001 | * |
| TLD | 0.251 | 0.063 | 0.125 | 0.377 | 3.966 | <.001 | * |
| DLD | -0.04 | 0.048 | -0.135 | 0.055 | -0.84 | 0.404 | |
| Group:Training | -0.019 | 0.091 | -0.2 | 0.163 | -0.203 | 0.84 | |

Residual Standard Error: 0.143
Degrees of freedom: 81 total; 77 residual

**Figure 1**

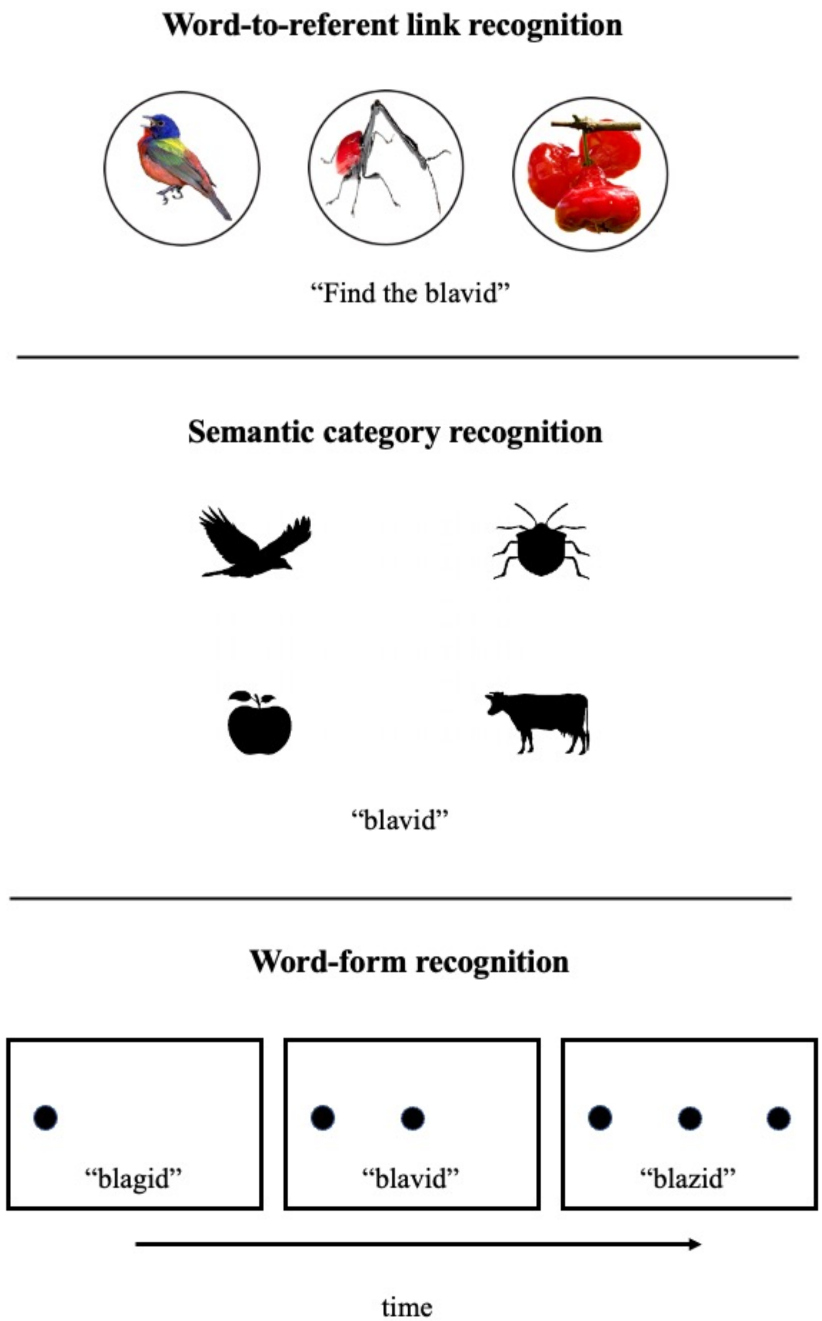
*Example training trials*




Note: children listened to speakers reading the sentences and did not see the written text in the experiment

**Figure 2**

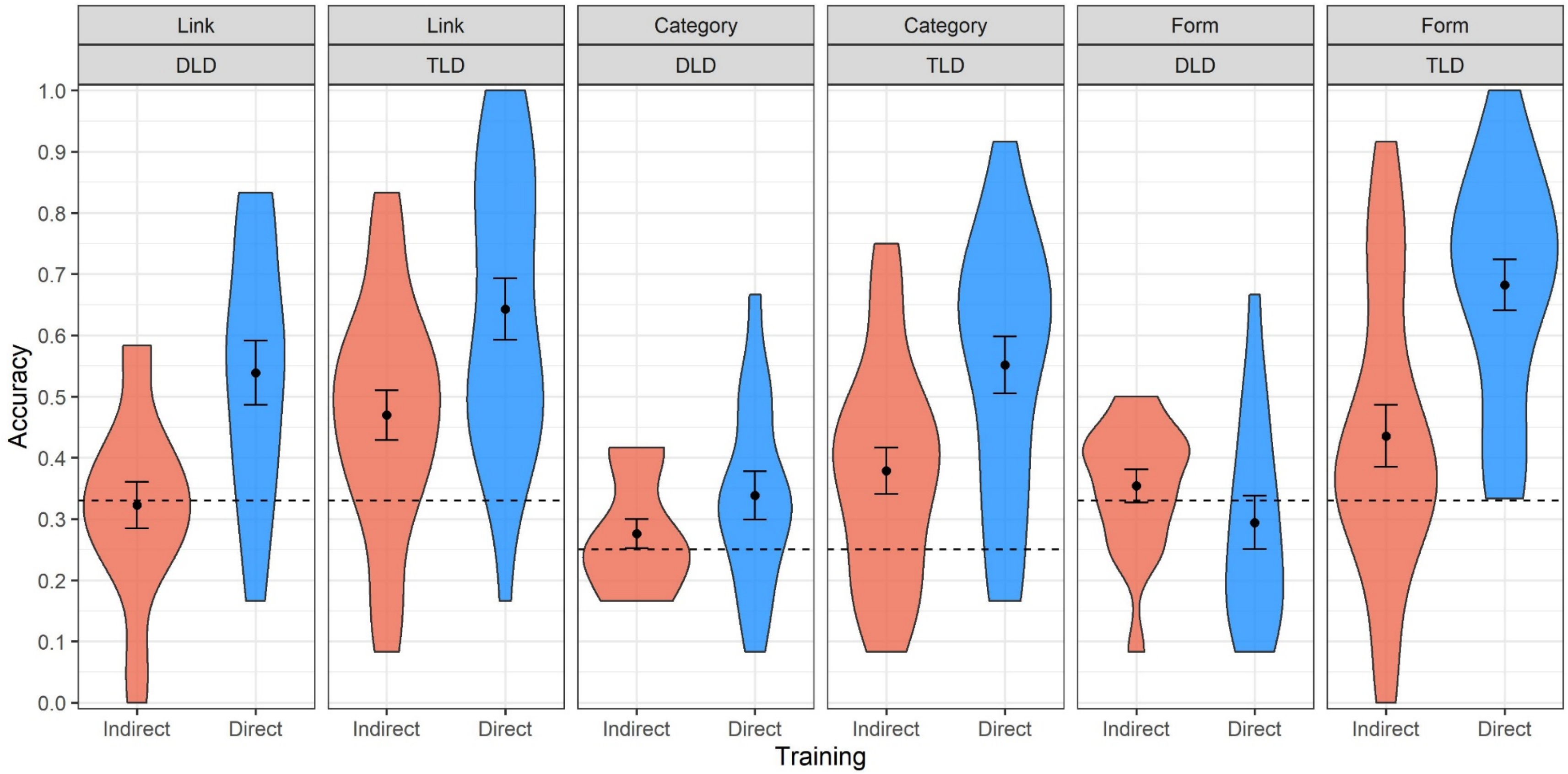
*Example test trials*




Note: children listened to speakers reading the sentences and did not see the written text in the experiment.

**Figure 3**

*Children's accuracy on test trials*




Note: data points represent the average across children and error bars +/- 1 SE. Violins show the distribution of accuracies across children. The dashed horizontal line represents chance performance (i.e., equal likelihood of choosing the target vs. the foils).