ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Inferring Word Class and Meaning From Spoken and Written Texts: A Comparison of Children With and Without Developmental Language Disorder

Karla K. McGregor,[a] Ron Pomper,[a] Nichole Eden,[a] Margo Appenzeller,[a] Timothy Arbisi-Kelm,[a] Elaina Polese,[a] and Deborah K. Reed[b]

[a] Boys Town National Research Hospital, Omaha, NE [b] Tennessee Reading Research Center, The University of Tennessee, Knoxville

ARTICLE INFO

Article History:
Received December 2, 2023
Revision received May 24, 2024
Accepted September 3, 2024

Editor-in-Chief: Julie A. Washington
Editor: Michael Hebert

https://doi.org/10.1044/2024_JSLHR-23-00743

ABSTRACT

Purpose: The aim of the study was to determine the ability of children with developmental language disorder (DLD) to infer word class and meaning from text and to document variations by word class (noun, verb, adjective) and modality (listening, reading). We also asked whether the children could integrate global cues across the entire passage as well as local cues from the immediate sentence frame to support inferences.
Method: Fourth graders with DLD (n = 28) and typical language development (TLD; n = 41) read and listened to expository texts and guessed the noun, verb, and adjective removed from each. Adults (n = 20) completed the task to establish a baseline of correct responses. We used latent semantic analysis (LSA) to determine the semantic fit of the responses to the texts and to determine whether global cues were more difficult for children with DLD than local cues.
Results: The DLD group was 24% less accurate than the TLD group. In both diagnostic groups, accuracy varied by word class (nouns > adjectives > verbs) but not modality (reading = listening). Word class errors were rare, and errors of semantic fit were frequent. LSA cosines were higher for correct responses relative to the passage as a whole than the immediate sentence frame, suggesting that both groups mined the more extensive information in the global cues to support inferences. Compared to the TLD group, the DLD group tended to make "worse" errors: repeating words from the sentence frame or coming up with no response at all. Accuracy in the DLD group, but not the TLD group, was related to vocabulary knowledge. When the two groups were collapsed, scores on verbal short-term/working memory and sustained attention also predicted performance, but weaknesses in these aspects of executive function on the part of individuals with DLD did not fully explain the difference between the performance of the DLD and TLD groups.
Conclusions: Whether listening or reading, fourth graders with DLD are less able to infer word meaning from texts than their age-mates. The problem reflects, in part, deficits in executive function and lexical semantic knowledge.

Children are seldom taught words directly. Instead, they use and integrate contextual information to draw inferences about word class and meaning (Bohn et al., 2021). The early lexicon comprises mainly concrete words, especially nouns with referents that may be inferred from physical and social contexts (Baldwin, 1991; Gillette et al., 1999; Smith & Yu, 2008; Snedeker, 2000). However, children who have established a nascent lexicon soon begin to use language to learn language. For example, children as young as 2 years use the difference between transitive and intransitive argument structures to infer whether a novel action is or is not performed on an object (Naigles, 1990). Three-year-olds use the grammatical morphemes surrounding a novel word to infer noun versus verb word classes (Brown, 1957).

Such linguistic bootstrapping is particularly important for the learning of predicates—verbs and adjectives—as these are less easily parsed from physical scenes

Correspondence to Karla K. McGregor: karla.mcgregor@boystown.org. *Disclosure:* *Karla K. McGregor is the chairperson of DLDandMe.org. The authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.*

(Gillette et al., 1999) and more constrained and defined by the linguistic contexts in which they appear (Booth & Waxman, 2009; Davies et al., 2023; Gentner, 1978). For example, the verb *throw* requires an object noun phrase like *the potato*, and the adjective *healthier* requires an adjectival complement like *than his brother*. The verb *assemble* takes on different meanings when combined with *bicycle* versus *group*, as does the adjective *firm* when modifying *handshake* versus *pillow*.

School speech-language pathologists and other education professionals realize that linguistic bootstrapping increases in importance after children learn to read. Written texts, devoid of the physical and social cues available in spoken communication, are the primary source of new word learning from third grade forward (Nagy et al., 1987). Vocabulary learning is a critical focus in the Common Core State Standards Initiative (2020), and if we want children to expand their vocabularies, they must be able to make inferences about the unfamiliar words they meet. In fact, vocabulary knowledge and language-based inferencing abilities positively correlated not only during childhood (Lynch et al., 2008; Oakhill & Cain, 2012; Schneider et al., 2023) but also during adulthood (Dixon et al., 1988), suggesting that language knowledge begets language learning across the life span.

In short, the ability to infer new word classes and meanings is integral to the development of the lexicon and to the comprehension of linguistic messages that contain any unfamiliar words. The primacy of language-based inferencing in language development is critical in any account of developmental language disorder (DLD). DLD is a neurodevelopmental condition that presents as clinically significant difficulties with language learning, comprehension, and expression. It follows that their ability to use language to learn language will be compromised, and this may explain, in part, why the language ability gap between children with DLD and their age-mates with typical language development (TLD) persists (Tomblin et al., 2003) or even widens (Rice & Hoffman, 2015) into the adult years. This article focuses on word learning via language-based inferencing among fourth graders with DLD. Our goal was to determine the extent and manifestation of the problem and its variation by word class (noun, verb, or adjective) and modality (listening or reading).

Vocabulary Development Among Individuals With DLD

Individuals with DLD tend to know fewer words than their peers at any given age, and the word meanings they know tend to be more superficial. For example, McGregor et al. (2013) analyzed the word definitions provided by children with and without DLD in second, fourth, eighth, and 10th grades by assigning them a zero (*no relevant information*), 1 (*a meaningful relationship to the target but failure to provide a minimal level of precision*), 2 (*conventional but minimal*), or 3 (*more than minimal*). The children with DLD scored more zeros than their peers and scored lower on the 1–3 scale for words they could define. The problem, relative to same-age peers, was equally evident at all grade levels.

To what extent do limitations on language-based inferencing contribute to the sparse lexicons of individuals with DLD? On average, children with DLD perform about 0.6 *SD* lower than age-mates with TLD on novel word-learning tasks (Kan & Windsor, 2010). Word-learning problems have been documented in many contexts where inferencing applies, including shared-book reading (Lavelli et al., 2019; Pile et al., 2010), independent reading (Steele & Watkins, 2010), and videotaped stories (Oetting et al., 1995; Rice et al., 2000). The extent of the problem lessens but does not entirely go away in direct instructional contexts where inferences are not required (Pomper et al., 2022).

Inference Making Among Individuals With DLD

Children with DLD have difficulties making language-based inferences to support their comprehension of narratives, including inferences that link ideas, bridge old and new information, and elaborate understanding by relating world knowledge to the text (Adams et al., 2009; Botting & Adams, 2005; Dodwell & Bavin, 2008; Norbury & Bishop, 2002). Some studies have been explicitly designed to evaluate the use of language-based inferences that support word learning. From these, we know, for example, that children with DLD are less able than same-age peers to use distributional statistics to parse new word forms from the speech stream (Evans et al., 2009) or syntactic cues to infer verb meanings (O'Hara & Johnston, 1997; Shulman & Guberman, 2007; van der Lely, 1994).

In a lab-based word learning study where unfamiliar nouns, adjectives, and verbs were presented in an animated story and the amount of exposure per word class was controlled, children with DLD had more difficulty learning adjectives than nouns and learning verbs than adjectives (Oetting et al., 1995). This finding accords with the thesis that there is a greater need for and more complicated application of linguistic bootstrapping for predicates as compared to nouns. Although some literature suggests that the challenge of learning predicates is no greater for children with DLD than for other children (Alt et al., 2004; Eyer et al., 2002; Leonard et al., 1982), in a meta-analysis of 28 studies, Kan and Windsor (2010) concluded that the gap between the DLD and

TLD groups is larger for verb learning than noun learning (adjectives were not considered).

Two mechanisms behind the language-based inferencing problem have been proposed. One possibility is that the limited linguistic knowledge base that characterizes DLD impedes inferencing (van der Lely, 1994), a reasonable hypothesis given that one must know the language structures and word meanings in the context around the new word to leverage that information. Consider the work of Evans (2002), who found that sentence comprehension strategies vary with the severity of the language impairment. Specifically, children with more severe language impairment tend to interpret the animate noun in a sentence to be the actor no matter the word order, whereas children with less severe impairment tend to interpret the first noun in the sentence to be the actor, as is typical of the canonical subject–verb–object sentence structure of English. The first strategy accords with the child's world knowledge (i.e., animate beings do things), whereas the second accords with the child's linguistic knowledge (i.e., subject nouns do things). The world knowledge approach will fail when the linguistic context involves noncanonical sentences or inanimate subjects.

A second possibility is that nonlinguistic processing limitations impede inferencing (O'Hara & Johnston, 1997). Three lines of data lend support. First, Shulman and Guberman (2007) concluded that a lack of linguistic knowledge could not explain the bootstrapping problems of their participants with DLD because younger peers with similar levels of language development were better able than those with DLD to use the bootstrapping cues provided. Second, in a previous publication involving the same sample of children studied here, we found that sustained attention was significantly correlated with overall language ability as measured by the Test of Narrative Language–Second Edition (TNL-2; Gillam & Pearson, 2017) among the children with DLD, $r(22) = .53$, but not among the children with TLD, $r(39) = .23$ (Smolak et al., 2020). Third, Blom and Boerma (2019) found that among children with DLD, working memory and attention shifting at 5 years of age predicted lexical development at 6 years of age even after controlling for lexical knowledge at the age of 5 years. This pattern did not hold for children with TLD, leading the authors to hypothesize that children with DLD rely more heavily on executive function than other children to learn words.

Thus, it may be that the children with DLD who have the most difficulty making language-based inferences are those for whom the attentional demands for noticing relevant linguistic cues or the memory demands of holding them in mind and integrating them while processing the spoken message are too great. Children with DLD often have weaknesses in sustained attention and working memory, as well as other key components of executive function as proposed by Friedman and Miyake (2017), namely, inhibition and shifting attention. Relative to age-mates with typical language development, children with DLD have reliable deficits: they average 1.27 *SD* lower on verbal short-term memory, a large effect (Estes et al., 2007); 0.63 *SD* lower on visuospatial working memory, a moderate effect (Vugs et al., 2013); 0.56 *SD* lower on inhibition, a moderate effect (Pauls & Archibald, 2016); and 0.27 *SD* shifting, a small effect (Pauls & Archibald, 2016). Of course, the linguistic knowledge and executive function hypotheses are not mutually exclusive.

Inferring in the Reading Modality

For children with DLD, reading might be a challenging modality for inferring new word meanings. A large proportion of children with DLD have reading deficits, with estimates ranging from 31% to 72% across studies (see summary in de Bree et al., 2022). Because reading develops on the foundation of spoken language skills, the vocabulary problems that characterize DLD may contribute to the reading problems. However, the reverse is also true; children who have difficulty reading will have fewer opportunities to encounter unfamiliar words, and when they do, they may be less able than their peers to comprehend the text well enough to infer the meanings of those words. Indeed, children who are poor at reading comprehension are known to have difficulty making inferences about new word meanings (Cain et al., 2003, 2004).

In Cain et al. (2003), 7- to 8-year-olds with stronger or weaker reading comprehension skills but similarly strong decoding skills read short stories in which a familiar word had been replaced with a novel word. They read to the end of the sentence containing the unfamiliar word and then defined it. Then, they read to the end of the story and defined the word again. Consider this excerpt from one of the stories:

> Bill was always very careful when riding his bike but the other day he fell off. When he looked round, he saw that the problem was a *gromp*. He phoned the council to complain. They sent a workman to mend the road and soon it was safe to ride along again. (Cain et al., 2003, p. 687)

Responses were classified as correct or as one of four error types. Here, *pothole* would be a correct response. Often, the children answered instead with a response that was semantically related to the story as a whole but not accurate, and these were classified as thematically appropriate errors. For this story, *flat tire* is an example. Other error types likely reflected strategies the children used when they could not make an inference. These included thematically inappropriate responses like *monster*, words that sounded like the unfamiliar word form like *grump* for *gromp,* repetitions of other words in the story, and “I don’t know” responses. The poorer comprehenders made significantly more errors than the stronger comprehenders, but the proportions of error by type were similar for the two groups. Also, both groups improved with their second guesses, and “I don’t know” responses declined; thus, the poor comprehenders could benefit from the extended cues in the full text. However, the farther away the cue from the novel word, the more difficulty they had. This finding speaks to the possibility that attention and working memory deficits impair inferencing not only from spoken language but also written language texts.

Two articles demonstrate that learners with trouble in one modality will likely have trouble in the other. Cain et al. (2001) evaluated the ability of poor readers to draw inferences while listening. Relative to peers, they demonstrated difficulties with various inference types: inferring the meaning of a simile, inferring links between utterances that gave the story coherence, and inferring links between utterances that enriched their representation of the story. The authors did not test the children’s ability to infer the class or meaning of unfamiliar words.

Steele and Watkins (2010) evaluated the ability of 9- to 11-year-olds with and without DLD to draw inferences about word meaning while reading. They replaced five English words with novel words in each of four reading passages. Half were nouns, and half were verbs. After reading, the children defined the words and selected their meanings from a four-alternative forced-choice list. The children with DLD performed lower on both measures than their age-mates with TLD. The difference in noun and verb learning was not significant. This study demonstrates that learning words while reading is difficult for children with DLD. However, it does not tell us whether it is more or less difficult than learning words while listening. Moreover, because the word learning outcomes were measured by definitions and a recognition task after the readings were completed, we cannot parse any problem with inference from problems with longer term memory or the formulation of definitions.

In the current study, we asked whether children with DLD face challenges when inferring word class and meaning and whether those challenges are more significant with written than spoken texts. We sought to isolate the process of inferring from the challenge of long-term memory and word definitions by removing a word from a given text and simply asking the children to guess the missing word. Moreover, because of disparate results in the literature concerning the relative challenge of various word classes, we compared inferences for nouns, verbs, and adjectives. Specifically, we tested these predictions:

1. Given that linguistic deficits define DLD, children with DLD will be less accurate in inferring a missing word from a grade-appropriate text than their age-mates with TLD. Accuracy was operationalized in two ways: as a binary correct or incorrect (with correct answers being any in the original text or given by a sample of 20 adults) and as a continuum of semantic similarity relative to the sentence and passage context as determined by latent semantic analysis (LSA). In both cases, we compared the accuracy of the children with DLD to that of their same-age peers with TLD, our logic being that their ability to make inferences is best judged against appropriate developmental expectations.
2. Given the more complicated linguistic constraints on predicates relative to nouns, children with DLD will have particular problems inferring the word class and meaning of verbs and adjectives.
3. Given the breadth of the language impairment associated with DLD, children with DLD will demonstrate difficulties when listening and reading. Alternatively, reading, especially in relatively novice readers with DLD, may add cognitive demands that render inferencing while reading even more problematic than listening.
4. Given that attention and memory deficits often accompany DLD and these are a necessary contributor to inference making,
    a. Participants with DLD and TLD will use both local cues (i.e., words in the immediate sentence frame) and global cues (i.e., words in the passage as a whole) as a basis for inference; however, participants with DLD will have more difficulty with global cues because attending to, remembering, and integrating them across the passage is problematic.
    b. Variation in inferencing skills will relate to independent measures of attention, verbal short term memory, and verbal working memory.

Method

Registration and Ethics

The current study was part of the Dynamics of Word Learning, a longitudinal study of word learning in Grades 1–4. Previous results appear in McGregor et al.

(2021, 2022, 2023), Pomper et al. (2022), and Smolak et al. (2020).

The project as a whole was preregistered (McGregor, 2023); however, this study was deemed exploratory, so no hypotheses were stated in the registration. We planned to administer the tasks of the current study twice to examine changes in performance from third to fourth grade, but the COVID-19 pandemic prevented data collection in Grade 3. Thus, the current study included only fourth graders and compared participants with DLD and TLD rather than change over time. Before participating, the children and their parents consented to the protocol approved by the institutional review board at Boys Town National Research Hospital.

Participants

Adults

Because more than one word could correctly fill any blank, we collected inferences from 20 adults and used the resulting responses as a baseline against which to judge accuracy. That is, if a child's response was adult-like, we took it to be correct. Furthermore, the variability of the adults' responses gave us insights into the extent to which the linguistic contexts motivated precise inferences. The adults were recruited via the online recruitment portal Prolific, where they reported current residence in the United States and exposure to English from birth and then completed the task by reading the passages and typing in their answers.

Children

The primary participants were 28 children with DLD (11 girls, 17 boys) and 41 children with TLD (23 girls, 18 boys). The DLD group averaged 10.18 years of age (SD = 0.45), and the TLD group averaged 10.24 (SD = 0.35), $t(63) = -0.61$, $p = .55$. All were in the fourth grade, and none had been retained for a grade. Fifty-five children were White, five were Black, one was Asian, and eight were more than one race. One was Hispanic, and five did not report their ethnicity. All were exposed to English from birth and lived in the United States. The level of caregiver education (total years of formal education based on the higher of the two caregivers) was significantly lower for the DLD group (M = 14.14 years, SD = 2.41) than the TLD group (M = 17.10 years, SD = 2.28), $t(63) = 5.16$, $p < .001$. To ensure that the amount of reading instruction they had received was similar, we compared the DLD and TLD groups by the month of Grade 4 in which they were tested: DLD, M = 7.2, SD = 3.2 and TLD, M = 8.1, SD = 2.9; $t(63) = -1.20$, $p = -.23$.

Children qualified for the DLD group if they scored below the 15th percentile on the Redmond Sentence Repetition Task (Redmond, 2005) and below a standard score of 92 on the TNL-2 (Gillam & Pearson, 2017), 92 being the cut-point that maximizes the sensitivity of identification of DLD on this particular test (Gillam & Pearson, 2017). Children qualified for the TLD group if they scored above these cutoffs. Children in both groups also passed a pure-tone audiometric screening administered per guidelines of the American Speech-Language-Hearing Association (n.d.) and scored better than a standard score of 70 on the Wechsler Abbreviated Scale of Intelligence–Second Edition Matrices subtest (Wechsler, 2011). This test battery was administered upon the children's enrollment in the Dynamics of Word Learning Project in first grade.

At the time of the current study, administration of the TNL-2, the Peabody Picture Vocabulary Test–Fifth Edition (PPVT-5; Dunn, 2019), and the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner et al., 2010) confirmed that the linguistic abilities of the DLD group remained significantly lower than those of the TLD group (see Table 1). One child with DLD did not complete the TOSREC. Still, we do have an indication that her ability to decode written words was low, given a standard score of 0 on the Nonword Reading subtest of the Test of Integrated Language and Literacy Skills (Nelson et al., 2016). Note that, by Grade 4, two children with DLD scored above the specified cutoff on the TNL-2, and one in the TLD group scored below. Some movement between the artificial categories that language researchers place on the continuous distribution of language scores is common in longitudinal studies of DLD (Tomblin & Nippold, 2014). We chose to retain the original classification, a conservative decision that could work against finding group differences in the current study.

We also administered the Track-It (Fisher et al., 2013), a measure of sustained visual attention; the backward digit span task from the Automated Working Memory Assessment (Alloway, 2007), a measure of verbal working memory; and the nonword repetition (NWR) task (Dollaghan & Campbell, 1998), a measure of verbal short-term memory. Although the sustained attention scores did not differ by diagnostic group, the children with DLD had lower backward digit spans and recalled fewer digits in total as well as fewer phonemes on the NWR task than their peers (see Table 1).

Three participants in the DLD group could not read well enough to complete the task in the written modality; the comparisons presented in Table 1 remained the same when these children were removed from the sample. These children were removed in the primary statistical analyses below; however, we included their responses in the listening modality when determining error types.

Table 1. Test scores of the participant groups in Grade 4.

Measure		DLD (n = 28[a])	TLD (n = 41)	t	p	Effect size d
TNL-2 (standard score)	*M* (*SD*)	83.44 (10.14)	111.2 (10.38)	10.89	< .001	−2.698
	Min–max	61–116	86–130			
PPVT-5 (standard score)	*M* (*SD*)	90.14 (9.05)	116.15 (15.08)	8.165	< .001	−2.002
	Min–max	68–113	91–141			
WASI-II Matrices (standard score)	*M* (*SD*)	91.5 (12.07)	112.46 (12.67)	6.72	< .001	−1.685
	Min–max	66–112	81–139			
TOSREC (standard score)	*M* (*SD*)	76.96 (14.22)	104.63 (15.1)	7.474	< .001	−1.874
	Min–max	54–99	78–142			
Track-It Attention (% correct[b])	*M* (*SD*)	0.82 (0.29)	0.91 (0.14)	1.523	.133	−0.386
	Min–max	0–1	0.33–1			
Backward Digit Span	*M* (*SD*)	2.82 (0.82)	3.80 (0.90)	4.618	< .001	−1.132
	Min–max	2–5	2–6			
Backward Digit total correct	*M* (*SD*)	11.79 (4.21)	17.29 (5.48)	4.487	< .001	−1.100
	Min–max	6–22	9–31			
NWR (no. of phonemes correct)	*M* (*SD*)	75.93 (8.76)	86.61 (5.42)	6.218	< .001	−1.541
	Min–max	53–89	74–96			

Note. DLD = developmental language disorder group; TLD = typical language development group; TNL-2 = Test of Narrative Language–Second Edition; PPVT-5 = Peabody Picture Vocabulary Test–Fifth Edition; WASI-II = Wechsler Abbreviated Scale of Intelligence–Second Edition; TOSREC = Test of Silent Reading Efficiency and Comprehension; NWR = nonword repetition.

[a]For the TOSREC, n = 27. [b]Track-It scores were based on correct memory trials in the heterogeneous condition.

Stimuli

The stimuli were 10 expository texts from the Grade 4 Stories and Reading Worksheets at K5 Learning (2023). In each, a noun, a verb, and an adjective were removed. The median age of acquisition (AoA; Kuperman et al., 2012) of the removed words was 6.5 years, with a range of 3.4–9.9. AoA did not differ by word class: Noun AoA: M = 6.6, SD = 2.2; Verb AoA: M = 6.6, SD = 1.2; Adjective AoA: M = 5.6, SD = 1.8; Noun to Verb: t = −0.05, p = .96; Noun to Adjective: t = −1.19, p = .25; and Verb to Adjective: t = −1.44, p = .17. The texts were grouped in two sets and modified slightly to better equate the sets for length, cohesion, syntactic complexity, word frequency, and Flesch–Kincaid grade level (see Table 2) as determined by the automated text analysis system Coh-Metrix (Graesser et al., 2011). Note that the validity of the Flesch–Kincaid grade level score improves if the text has more than 200 words (Graesser et al., 2004). Our texts did not; however, they were all excerpts from Grade 4 reading materials, so we can be confident that they are reasonable for a child who reads at a fourth-grade level. Assignment of sets to the reading or listening modality and the order of administration were counterbalanced across children.

The texts were presented in a PowerPoint deck. For each text, there was a single picture that illustrated something about the content of the text. For example, there was a picture of a mummy alongside a text about a farmer discovering a mummy when working the land. These were included to mirror common practice in children's readers. None of the pictures directly cued a target word (e.g., *mummy* was not one of the missing nouns). In the listening task, all texts were prerecorded, and the deck included the picture only, while the text was presented to the children via audio. In the reading task, the printed words and the picture were visible, and there was no audio.

Procedure

Adults

The adults completed the task in the reading modality only via a REDCap survey. For each passage, there was one page with the written text and picture, followed by one page with three sentence frames in which they typed their answers to fill in the blanks. Their instructions were:

> You will read 10 paragraphs. In each, there will be three blanks for you to fill. For example, if you read: "It's my mom's birthday. She hopes she gets lots of _______s." You might write **present**s.
>
> Please follow these rules:

- Skim the entire paragraph before you fill in the missing words.
- Write the first word that comes to mind.
- Write only a single word. Although "exciting present" would fit well in the blank above, go with a single word only.

Table 2. Characteristics of text stimuli.

Set	Text topics	Length in sentences	Length in words	Cohesion	Syntactic complexity	Word frequency	Grade level
A	Eyes	14	157	62.17	3.286	2.280	3.970
	Firefighters	12	163	84.13	2.667	2.728	5.413
	Camouflage	14	156	79.39	2.000	2.193	4.639
	Mummies	13	115	54.78	2.231	2.256	3.967
	Pie Town	15	170	51.99	2.600	2.205	4.512
	Set A mean[a] (*SD*)	13.6 (1.1)	152.2 (21.5)	66.49 (14.5)	2.56 (0.49)	2.33 (0.22)	4.50 (0.60)
B	Otters	16	143	89.25	1.813	1.975	4.569
	Zebras	13	146	76.11	2.615	2.746	4.708
	Sloths	12	130	47.21	2.667	1.922	3.798
	Washington	13	128	64.80	2.538	2.640	4.381
	Astronauts	15	170	90.82	2.867	2.145	5.137
	Set B mean[a] (*SD*)	13.8 (1.6)	143.4 (16.8)	73.64 (18.18)	2.50 (0.40)	2.29 (0.38)	4.52 (0.49)

Note. Length in sentences = number of sentences in the text; Length in words = number of words in the text; Cohesion = the degree to which the text contains causal and intentional connectives (expressed as a percentile); Syntactic complexity = the average number of words that precede the verb in each sentence; Word frequency = the average log frequency with which the content words in the text appear per one million words in corpora of written English; Grade level (Flesch–Kincaid) = grade at which most students can read the text accurately.
[a]When comparing means for Sets A and B via t tests, all ps $\geq .49$.

- Do not worry about spelling. As long as you are close, we will figure it out. We would rather that you give your first guess than write a different word that is easier to spell!

Children

The children completed the task via a Zoom meeting with the examiner. They were told that they would listen to five stories and read five stories, and three words would be missing in each. First, they heard or read the stories without filling in the missing words. Next, they heard or read the three sentence frames containing the missing words. After each frame, they were to guess the word that best fit in the blank. This practice example was shared:

"If you read, 'it's my mom's birthday and she hopes she gets lots of BLANKs,' you might guess … presents." They were explicitly instructed to use one-word answers only, and if they responded with phrases instead, they were reminded of the one-word rule.

The assignment of text set to modality condition and the order of modality conditions were counterbalanced across participants.

Data Analysis

Scoring and Error Classification

All responses that were identical to the original words in the texts or to any of the adults' answers were scored as correct. The adults' answers appear in the Stories section of the supplemental material (see https://osf.io/9gubc). As a preliminary step, we identified the children's errors that were not of primary interest, including grammatical errors (problems with tense, number or person agreement, or argument structure), word form errors (a mispronunciation of a target word), repetitions (repetitions of a word in the sentence frame or story as a whole), phrases (supplying more than a single word response), and "I don't know." Next, the first and third authors identified the word class errors by consensus. These were any violations of the target word class that rendered the sentence ungrammatical:

- Noun target: You would have a hard time finding a city or <u>state</u> on a map if it had no name. The child answered, "*harder*."
- Verb target: People <u>discover</u> mummies in hot deserts and wet places. The child answered, "*and*."
- Adjective target: Because many people think that mummies are <u>interesting</u>, you can often find them in museums. The child answered, "*history*."

The remaining responses tended to bear a semantic relationship to the target but lacked precision or relevance. It was readily apparent that some were closer and others farther from the target meaning. Consider these two responses:

- Target: Animals use sight to hunt <u>prey</u> and to avoid predators. The child answered, "*mice*."
- Target: A farmer was working on his land when his <u>shovel</u> hit something hard. The child answered, "*book*."

The first example is closer to the target and a better fit to the sentence because mice are common prey. The second example is farther from the target and a worse fit to the sentence because books are not strongly associated with shovels. To get at these distinctions objectively, we used LSA (Landauer, 2007) to quantify the semantic similarity between the answer and the sentence frame (local cues) and between the answer and the passage as a whole (global cues).

LSA determines the degree to which one word or passage is related to another word or passage, and the meaning of a given word (represented as a dimensional vector) is the sum of those relationships. In practice, LSA takes a word and computes its distance from other words appearing in very large corpora. The results are cosine values that range from −1 to 1; values below zero are infrequent, and higher values indicate higher degrees of similarity in meaning. For the examples above, the cosine for mice is .15. In contrast, the cosine for book is .01. The LSA website offers options that allow the researcher to tailor the analysis to the research goals and maximize validity, options such as the specific corpus and the number of factors used to define a word. For replication purposes, note that we used the Word Embedding Analysis Website (http://wordvec.colorado.edu) to conduct the analysis, using these options: (a) one-to-many comparison, (b) LSA embedding method, (c) general reading up to sixth grade embedding space (a fourth-grade embedding space was not an option), (d) document-to-document comparison type, and (e) maximum factors. An explanation of these options is available on the website.

Statistical Analysis

The adults' responses were analyzed descriptively. For the children's responses, the primary analyses involved regression models including, where relevant, a fixed within-subject factor of modality (listening or reading), word class (noun, verb, adjective), or context (global, local) and a between-subjects effect of diagnosis (DLD or TLD). All analyses were linear models fit using the generalized least squares, including a compound symmetry correlation matrix to account for within-subject correlation using the nlme package (Version 3.1.163; Pinheiro et al., 2023), and plots were created using the tidyr package (Version 1.3.0; Wickham et al., 2023) in R (Version 4.3.1; R Core Team, 2020) using RStudio (Version 2023.9.0 + 463; RStudio Team, 2020). Detailed specifications for all analyses appear in the Results section of the supplemental material (see https://osf.io/9gubc).

Results

Adults

As anticipated, the adults did not always respond with the target word in the passage; on average, only 26% of the adults responded with a target for any given item (SD = 26): 36% (SD = 34) of adults responded with the target noun, 35% (SD = 18) with the target verb, and 9% (SD = 12) with the target adjective. The range of target responding was large even within word class. For example, in the "Animal Eyes" passage, 90% of adults responded with the target noun *prey* when given the frame, "animals use sight to hunt __ and to avoid predators." In contrast, in the "Astronauts" passage, none responded with the target noun *bodies* when given the frame, "Once in their suits, astronauts breathe pure oxygen for a few hours so that they will not get gas bubbles in their ______ s." Instead, the most common response was *lungs*. This variation from the target does not undermine our decision to consider all adult responses correct. The LSA cosines for the original targets were .09 (SD = 0.08), relative to the passage as a whole, and .06 (SD = 0.08), relative to the sentence frame. The LSA cosines were just as high for the other responses in the adult data set: passage = .10 (SD = 0.01) and sentence = .08 (SD = 0.01). Instead, this variability suggests that the texts did not include cues that obligated a single specific inference and that some texts provided more cues to constrain inferences than others. LSA cosines for adult responses to each story and word class target appear in the Stories section of the supplemental material (see https://osf.io/9gubc).

Children

Accuracy of Inferences by Diagnostic Group

Mean accuracy in the DLD group was 0.41 (SD = 0.12) compared to 0.65 (SD = 0.11) in the TLD group, b = 0.244, $t(64)$ = 8.282, $p \leq .001$. There was no within-group relationship between language scores on the TNL-2 as measured in Year 4 (the same time as the inference task) and inference accuracy: DLD: b = 0, $t(22)$ = 0.164, p = .871; TLD: b = 0.001, $t(39)$ = 0.609, p = .546. In other words, the presence, but not severity, of the language impairment was predictive. That said, the ability to infer new words was related to vocabulary size as measured by the PPVT-5 in the DLD group, b = 0.007, $t(23)$ = 3.31, p = .003, but not in the TLD group, b = 0.001, $t(39)$ = 0.936, p = .355, likely because all children in the TLD group had age-appropriate vocabulary levels (standard scores > 90) and they were working with age-appropriate texts.

Error Types by Diagnostic Group

We characterized inaccurate responses by type (see Table 3). The three most frequent error types in the DLD group were problems with semantic fit, repetitions, and "I don't know" responses. Answers that reflected problems with semantic fit made for well-formed sentences but were outside the adults' response set. They reflected problems with inferring word meanings, although some came closer to target meanings than others. (We return to this issue in

Table 3. Error types by diagnostic group.

Error type	Example	DLD errors (*N* = 476)		TLD errors (*N* = 428)	
		n	Proportion of errors	*n*	Proportion of errors
I don't know		61	.13	5	.01
Word form	He helped choose the place to /əkstrʌkt/ our capital city. (aiming for "construct")	3	.006	1	.002
Phrases	When it's dark out, holes called pupils do something to cover the front of the owls' eyes.	13	.03	6	.01
Repetitions	They leave the spacecraft through a airlock door called an airlock.	179	.38	129	.30
Argument structure	A businessman started buying supplies . . . to the cowboys.	23 (3 reps)	.05 (total) .04 (w/out reps)	22 (4 reps)	.05 (total) .04 (w/out reps)
Tense or agreement	They make sure their trucks is clean.	2 (1 rep)	.004 (total) .002 (w/out reps)	3 (1 rep)	.007 (total) .002 (w/out reps)
Word class errors	Luckily for the zebra, the disease fly doesn't land on it.	52 (37 reps)	.11 (total) .03 (w/out reps)	24 (5 reps)	.06 (total) .04 (w/out reps)
Semantic fit	Everyone hoped that he would be our first president because he was very good.	198	.42	258	.60

Note. DLD = developmental language disorder group; TLD = typical language development group; w/out = without.

the LSA results below.) These were also the most frequent error types among the children with TLD; in fact, the DLD group was 43% less likely to make semantic fit errors than their peers. In contrast, they were 130% more likely to respond with "I don't know" and 27% more likely to respond with a repetition than the TLD group. The DLD group also made more word class errors than their peers, for example, by answering with a noun instead of an adjective or a verb instead of a noun. However, once repetitions were accounted for, word class errors were rare in both groups. All other error types constituted less than 5% of the total errors in both groups.

Accuracy of Inferences by Word Class and Diagnostic Group

The effect of word class was statistically significant, $\chi^2(2, N = 198) = 21.28$, $p < .001$; the interaction between word class and diagnostic group was not, $\chi^2(2, N = 198) = 2.10$, $p = .35$ (see Figure 1). Thus, any differences in word class accuracy for children with DLD are similar for children with TLD. Children with DLD were significantly less accurate in inferring verbs ($M = 0.31$, $SD = 0.15$) than nouns ($M = 0.50$, $SD = 0.20$), $b = -0.18$, $t(192) = -4.6$, $p < .001$; adjectives ($M = 0.42$, $SD = 0.18$) than nouns ($M = 0.50$, $SD = 0.20$), $b = -0.08$, $t(192) = -2$, $p = .047$; and verbs ($M = 0.31$, $SD = 0.15$) than adjectives ($M = 0.42$, $SD = 0.18$), $b = -0.10$, $t(192) = -2.6$, $p = .01$.

Accuracy of Inferences by Modality and Diagnostic Group

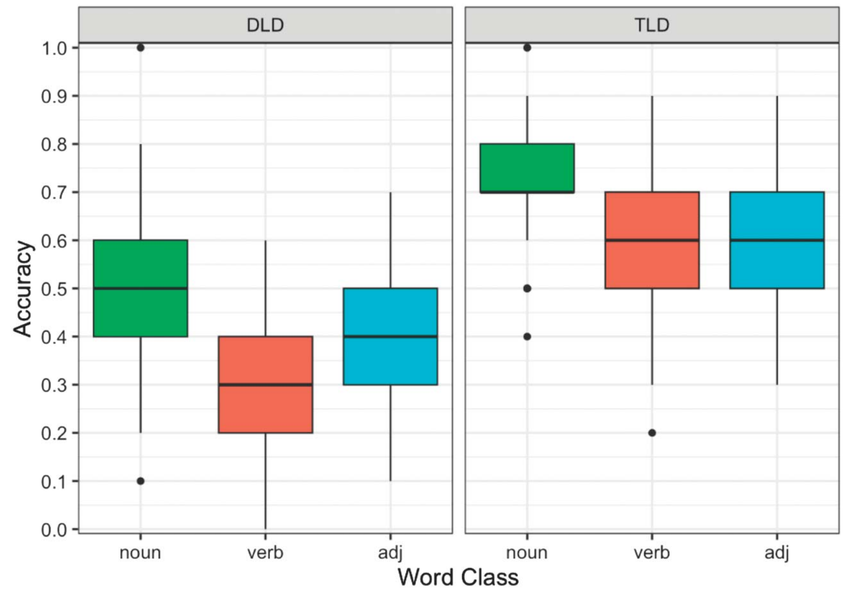
The effect of modality was not significant, $b = 0.02$, $t(128) = 0.58$, $p = .561$. Children with DLD were similarly accurate at inferring words when reading ($M = 0.40$, $SD = 0.15$) and listening ($M = 0.42$, $SD = 0.15$) to the texts. The interaction between group and modality was not statistically significant, indicating that the effect was similar for both groups, $b = -0.02$, $t(128) = -0.32$, $p = .75$ (see Figure 2).

Use of Global and Local Cues by the Two Diagnostic Groups

When comparing LSA values for *correct* answers, there was no significant effect of diagnostic group, $b = -0.01$, $t(128) = -0.80$, $p = .423$. Children in the DLD group ($M = 0.10$, $SD = 0.03$) provided correct answers that were similar in semantic fit to children in the TLD group ($M = 0.10$, $SD = 0.02$). There was a significant effect of cue context, $b = -0.02$, $t(128) = -2.93$, $p = .004$. In both diagnostic groups, cosines were higher for correct responses relative to the passage as a whole than the immediate sentence frame, suggesting that both groups mined the more extensive information in the global cues to support inferences. The interaction between group and cue context was not statistically significant, indicating that the effect was similar for both groups, $b < .0004$, $t(128) = 0.05$, $p = .962$. For the DLD group, the LSAs for the correct responses in relation to sentence frames (local cues) averaged .09 ($SD = 0.04$), and in relation to passages (global cues) averaged .11 ($SD = 0.03$).

We were especially interested in the semantic fit of *incorrect* responses because it was apparent that some errors were better than others. For this analysis, we deleted "I don't know," phrasal responses, and repetitions of words in the passage and sentence frames, the latter

Figure 1. Accuracy of inferences by word class and diagnostic group. DLD = developmental language disorder group; TLD = typical language development group; adj = adjective.

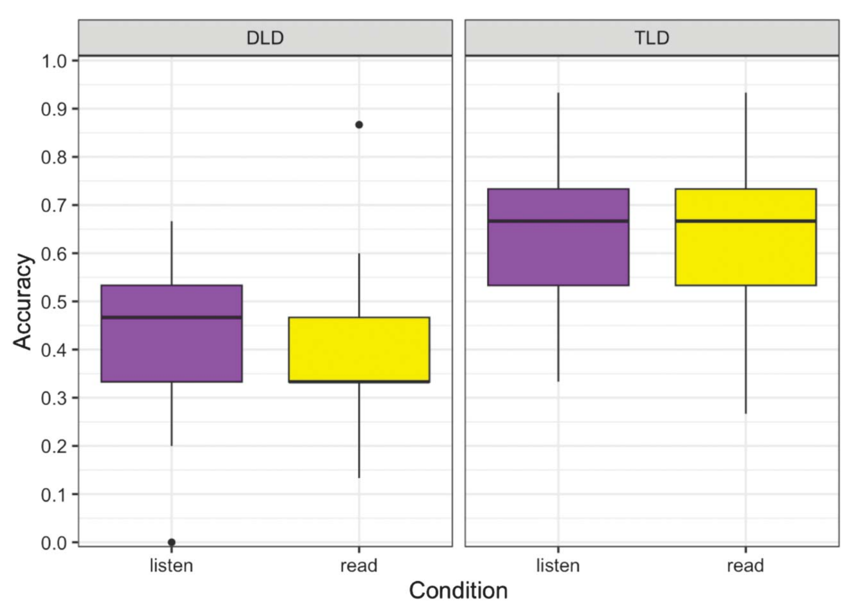


two because these would artificially inflate the LSA values. After removing correct responses and these three error types, there was no significant effect of group, $b = -0.004$, $t(128) = -0.80$, $p = .428$. Children in the DLD group ($M = 0.04$, $SD = 0.02$) provided answers that were similar in semantic fit to children in the TLD group ($M = 0.04$, $SD = 0.02$). There was no significant effect of cue context, $b = -0.01$, $t(128) = -1.18$, $p = .242$, or interaction between group and cue context, $b = -0.01$, $t(128) = -1.25$, $p = .213$.

Relationship Between Inferencing Skill and Independent Measures of Memory and Attention

As hypothesized, inference performance varied with memory and attention skills. When collapsing across groups, there was a significant effect of verbal short-term

Figure 2. Accuracy of inferences by modality and diagnostic group. DLD = developmental language disorder group; TLD = typical language development group.



memory skill, NWR: b = 0.011, $t(63)$ = 5.493, $p \leq .001$; verbal working memory skill, backward digit total correct (BDT): b = 0.015, $t(64)$ = 4.68, $p \leq .001$; and sustained attention, Track-It: b = 0.229, $t(60)$ = 2.406, p = .019. When examined within group, these effects became marginal or nonsignificant (ps = .063–.352; see the Exploratory 3 section in the supplemental material [see https://osf.io/9gubc]), with the one exception being that, for children with TLD, there was a significant effect of verbal short-term memory as measured by the NWR, b = 0.007, $t(39)$ = 2.423, p = .02. Therefore, these significant relationships between inferencing accuracy and memory and attention could reflect, in part, broader differences between the DLD and TLD groups. To explore this possibility, we reran the model testing the effect of diagnostic group on inference accuracy while adding in scores on the Track-It and BDT as covariates. We did not include NWR scores in the model because the collinearity between BDT and NWR was high, r = .52. The result was a significant effect of BDT, b = 0.007, $t(58)$ = 2.34, p = .023. For each 1-point increase in BDT, children's accuracy in inferring the missing words increased by 0.7%. There was also a marginal effect of Track-It, b = 0.12, $t(58)$ = 1.782, p = .08. For each .1-point increase in Track-It accuracy (i.e., 10%), children's accuracy in inferring the missing words increased by 1.2%. Crucially, after accounting for group differences in BDT and Track-It and their effects on inferring missing words, there was *still* a significant effect of group, b = 0.202, $t(58)$ = 6.27, $p \leq .001$. Children in the TLD group were more accurate than children in the DLD group, although the size of this effect was smaller compared to the model without these covariates. Together, these results confirm that variation in inferencing skill relates to independent measures of working memory (and potentially attention). However, group differences in these executive function abilities do not entirely account for the group differences in inference ability.

Discussion

We asked fourth graders with DLD and their age-mates with TLD to infer missing nouns, verbs, and adjectives while listening to or reading grade-appropriate texts. The original words in the texts and all of the adults' responses comprised the set of correct responses. As predicted, the children with DLD gave fewer correct responses than their peers, scoring, on average, 24% lower.

Error Types

As in Cain et al.'s (2003) study, the most frequent error type in both diagnostic groups involved answers that made for well-formed sentences and that related, more or less, to the target word, sentence frame, or passage as a whole. In other words, the children made use of the linguistic context, but they could not always infer a word that was as good a semantic fit as the adults. These response types were less common in the DLD group than the TLD group; however, the LSA revealed no differences between groups in the "semantic fit" of these erred responses. That is, when they did produce this error type, they came as close to the targets as their peers.

Word class errors were not especially problematic. Once we accounted for the errors that were also repetitions, word class errors were not frequent in either group. In the passages we presented, word class was largely cued by word order (e.g., "they float at the surface and ______ shells on their bellies"), although in some cases, inflections provided additional cues (e.g., "it can pass on deadly diseases as it ______s"). These fourth graders with DLD were able to mine these cues.

Two error types were proportionately more common in the DLD group than in the TLD group, and they both indicate inference failures. Some were repetitions of a word that had appeared in the sentence frame or passage characterized. These might be employed strategically when the child finds it difficult to infer the correct word; after all, repetitions are almost always semantically relevant to the passage, albeit pragmatically odd. Many others were "I don't know." Given the overall pattern of errors, we conclude that, relative to their peers, the DLD group made not only more errors but also more of the worst errors—those that were nonresponses.

Responses by Word Class

As predicted, the children with DLD (and their age-mates with TLD) were more successful in inferring the meanings of nouns than verbs and adjectives. Although the nouns were easiest, the children with DLD found adjectives easier to infer than verbs; this was also the case in Oetting et al.'s (1995) study. There was no indication from the adult data that the passages provided more support for inferring adjectives; in fact, a higher proportion of adults responded with target verbs than target adjectives. Instead, the children may have been more familiar with the adjectives than the verbs. Although AoA did not differ significantly by word class, the noun and verb sets each had a mean AoA of 6.6 years, whereas the adjectives had a mean of 5.6 years. This tentative conclusion awaits future study.

Responses by Modality

The children with DLD were poor readers. On the TOSREC, a test that measures reading comprehension and fluency, all but four children with DLD scored more than 1 *SD* below the mean of the normative sample.

However, their inferences in the reading modality were no less accurate than in the listening modality. Thus, it was not reading difficulties that limited their ability to draw inferences. We state this conclusion with the caveat that three children with DLD were such weak readers that they could not complete the inferencing task in the reading modality. Our findings put into perspective those of Steele and Watkins (2010), who examined word inferences during reading only. Although it reinforces their conclusion that inferring word meanings is difficult for children with DLD, it also demonstrates that the problem exists regardless of modality. That said, the reading problem may influence inferencing in indirect ways. A child who has difficulty comprehending what they read (or hear) will build a weaker language system than a child who is a good comprehender and thus have a weaker system from which to make inferences.

The Nature of the Problem

Given the pattern of responses, we can reject several explanations of the inferencing problem. The difficulty is not attributable to poor reading abilities per se, as the children with DLD had equivalent problems while listening and reading. Their ability to use word order and word inflection as cues was not problematic, as they rarely mistook one word class for another. Semantic activation and integration of semantic cues do not seem to be a barrier, as the semantic fit of their correct and incorrect responses was comparable to that of their peers.

There is partial support in the data for the executive function and linguistic knowledge explanations of the inferencing problem. Given that problems with attention and verbal short-term and working memory often accompany DLD, we had predicted that participants with DLD would find global cues spread throughout the passage more problematic than local cues, those in the immediate sentence frame. However, the children with DLD made equivalent use of local and global cues, just like their peers with TLD. Still, there was some indication of the relevance of executive function. Specifically, verbal working memory scores accounted for variance in inference performance, a finding broadly consistent with Blom and Boerma (2019), Cain et al. (2003), O'Hara and Johnston (1997), and Shulman and Guberman (2007).

After accounting for the effect of executive function on inferencing, there remained a significant difference between the diagnostic groups. Here, the linguistic knowledge hypothesis is relevant. Taking the "I don't know" responses literally, one might conclude that the children with DLD simply did not know or could not fully process all of the linguistic structures, vocabulary words, or semantic relationships in the text and, therefore, were unable to use them as a basis for inference. Within the DLD group, vocabulary knowledge as measured by the PPVT-5 accounted for variance in inferring missing words, whereas broader language abilities measured by the TNL-2 administered concurrently with the inferencing task did not.

The relatively high rate of repetition errors in the DLD group suggests a third possibility: a problem with lexical inhibition. In a series of studies, McMurray and colleagues have demonstrated that adolescents with DLD are similar to peers in activating lexical competitors and recognizing a target in a visual world paradigm. However, they continue to fixate on competitors late in the trial (McMurray et al., 2014, 2019, 2022). In other words, they demonstrate deficits in the inhibition of related lexical items. In the current study, the relatively high rate of repetitions of words from the text suggests difficulty inhibiting semantically related but incorrect responses.

Limitations

One of the advantages of this study was its ecological validity. These fourth graders were making inferences from authentic fourth-grade texts. However, that strength is also a weakness, given the inherent lack of control over the predictability of the target words. Tighter matching of the AoA of the target words and a posttest to measure the children's knowledge of the targets would be valuable additions to future studies. The sample size was also small and constrained to children in Grade 4, thus limiting external validity.

Finally, our executive function measures were limited to three tasks, and two key components of executive function—inhibition and shifting attention—were not measured. A more comprehensive battery would be helpful.

Clinical Implications

Although we have known for some time that children with DLD have difficulty learning new words from text (Steele & Watkins, 2010), this study isolates inference as a culprit. Had we administered this task in the reading modality only, we would likely have concluded that the children with DLD could not infer as well as their peers because they could not read as well. Nearly all of the participants with DLD had or were at risk for reading disability. However, reading is not the direct barrier that interferes with their inferencing. These children have a broader problem that stands to limit their word learning in both spoken and written modalities. Service needs must be determined with the breadth of the problem in mind. Specifically, a child who is having difficulty with reading comprehension and inferencing should be tested to determine whether that problem is specific to reading or instead reflects a broader linguistic deficit that extends to spoken language.

The current study indicates the need for interventions that support word inferencing and, more specifically, suggests what the focus of that intervention should be. The children with DLD did not have difficulty inferring word class; like their peers, they tended to infer nouns, verbs, or adjectives as the contexts indicated. Moreover, their inferences tended to be age appropriate in terms of semantic fit. The problem was mainly one of response failure, the inability to generate an inference in the moment. We also offer preliminary insights into the mechanisms that might contribute to the inferencing problem: deficits in verbal working memory, vocabulary knowledge, and lexical inhibition may play a role.

We did not investigate the malleability of inferencing skills or the effectiveness of different approaches to teaching those skills; however, there is evidence that directly intervening to teach inference may be effective (Fukkink & de Glopper, 1998). For example, 5- to 6-year-olds with DLD who received a 16-session treatment designed to enhance inferential comprehension of narratives improved, generalized, and maintained inferential comprehension at levels higher than a comparison group randomly assigned to phonological awareness training over the same period (Dawes et al., 2019). The inferences targeted were causal, informative, or evaluative (e.g., *Why did Bear jump out of the mud? How do you think he felt?*); they were meant to promote understanding of the narratives rather than word learning. When teaching word inferences, consider having children read the sentences around an unfamiliar word, identify its antonyms and synonyms in the text, and break it into its root and affixes. These strategies were more effective than direct instruction on word meanings for helping typical fifth graders infer new words (Baumann et al., 2003).

Conclusions

Fourth graders with DLD were less able to infer word meaning from linguistic context than their grademates with TLD. Inferring while listening was just as challenging as inferring while reading. The clinical implications are twofold. First, children who have difficulty inferring word meanings while reading should be evaluated for their ability to make inferences while listening. Second, inferencing is a likely intervention target for children with DLD, one that could benefit both language comprehension and word learning. From a theoretical perspective, the weakness in using language to learn language that we documented here may help to account for the persistent nature of DLD. With this avenue toward language development partially blocked, it will be difficult for individuals with DLD to catch up to unaffected peers.

Data Availability Statement

The raw data, analysis code, and supplemental materials are available in the Open Science Framework (https://osf.io/9gubc).

Acknowledgments

This work was funded by National Institute on Deafness and Other Communication Disorders Grant 5 R01 DC011742-07 awarded to Karla K. McGregor (Principal Investigator). We thank the caregivers and children who participated. Justin Kueser suggested the latent semantic analysis, and Jacob Oleson confirmed our analytic approach.

References

Adams, C., Clarke, E., & Haynes, R. (2009). Inference and sentence comprehension in children with specific or pragmatic language impairments. *International Journal of Language and Communication Disorders, 44*(3), 301–318. https://doi.org/10.1080/13682820802051788

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

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

American Speech-Language-Hearing Association. (n.d.). *Childhood hearing screening* [Practice Portal]. https://www.asha.org/practice-portal/professional-issues/childhood-hearing-screening/#collapse_1

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development, 62*(5), 875–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Baumann, J. F., Edwards, E. C., Boland, E. M., Olejnik, S., & Kame'enui, E. J. (2003). Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. *American Educational Research Journal, 40*(2), 447–494. https://doi.org/10.3102/00028312040002447

Blom, E., & Boerma, T. (2019). Reciprocal relationships between lexical and syntactic skills of children with developmental language disorder and the role of executive functions. *Autism & Developmental Language Impairments, 4*. https://doi.org/10.1177/2396941519863984

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature Human Behaviour, 5*(8), 1046–1054. https://doi.org/10.1038/s41562-021-01145-1

Booth, A. E., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development, 80*(1), 15–22. https://doi.org/10.1111/j.1467-8624.2008.01242.x

Botting, N., & Adams, C. (2005). Semantic and inferencing abilities in children with communication disorders. *International Journal of Language & Communication Disorders, 40*(1), 49–66. https://doi.org/10.1080/13682820410001723390

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal and Social Psychology, 55*(1), 1–5. https://doi.org/10.1037/h0041199

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850–859. https://doi.org/10.3758/BF03196414

Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language, 30*(3), 681–694. https://doi.org/10.1017/S0305000903005713

Cain, K., Oakhill, J. V., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671–681. https://doi.org/10.1037/0022-0663.96.4.671

Common Core State Standards Initiative. (2020). *Key shifts in language arts*. https://www.thecorestandards.org/other-resources/key-shifts-in-english-language-arts/

Davies, C., Ebbels, S., Nicoll, H., Syrett, K., White, S., & Zuniga-Montanez, C. (2023). Supporting adjective learning by children with developmental language disorder: Enhancing metalinguistic approaches. *International Journal of Language & Communication Disorders, 58*(2), 629–650. https://doi.org/10.1111/1460-6984.12792

Dawes, E., Leitão, S., Claessen, M., & Kane, R. (2019). A randomized controlled trial of an oral inferential comprehension intervention for young children with developmental language disorder. *Child Language Teaching and Therapy, 35*(1), 39–54. https://doi.org/10.1177/0265659018815736

de Bree, E. H., Boerma, T., Hakvoort, B., Blom, E., & van den Boer, M. (2022). Word reading in monolingual and bilingual children with developmental language disorder. *Learning and Individual Differences, 98,* Article 102185. https://doi.org/10.1016/j.lindif.2022.102185

Dixon, P., LeFevre, J. A., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80*(4), 465–472. https://doi.org/10.1037/0022-0663.80.4.465

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Language and Communication Disorders 43*(2), 201–218. https://doi.org/10.1080/13682820701366147

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

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test–Fifth Edition* [Measurement instrument]. Pearson.

Estes, K. G., 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)

Evans, J. L. (2002). Variability in comprehension strategy use in children with SLI: A dynamical systems account. *International Journal of Language & Communication Disorders, 37*(2), 95–116. https://doi.org/10.1080/13682820110116767

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 321–335. https://doi.org/10.1044/1092-4388(2009/07-0189)

Eyer, J. A., Leonard, L. B., Bedore, L. M., McGregor, K. K., Anderson, B., & Viescas, R. (2002). Fast mapping of verbs by children with specific language impairment. *Clinical Linguistics & Phonetics, 16*(1), 59–77. https://doi.org/10.1080/02699200110102269

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

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86,* 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research, 68*(4), 450–469. https://doi.org/10.3102/00346543068004450

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

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

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

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 193–202. https://doi.org/10.3758/BF03195564

K5 Learning. (2023). *Grade 4 children's stories & reading worksheets.* https://www.k5learning.com/reading-comprehension-worksheets/fourth-grade-4/childrens-stories

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)

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavioral Research Methods, 44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*. Psychology Press. https://doi.org/10.4324/9780203936399

Lavelli, M., Barachetti, C., Majorano, M., Florit, E., Brotto, C., & Miottello, P. (2019). Impacts of a shared book-reading intervention for Italian-speaking children with developmental language disorder. *International Journal of Language & Communication Disorders, 54*(4), 565–579. https://doi.org/10.1111/1460-6984.12460

Leonard, L. B., Schwartz, R. G., Chapman, K., Rowan, L. E., Prelock, P. A., Terrell, B., Weiss A. L., & Messick, C. (1982). Early lexical acquisition in children with specific language impairment. *Journal of Speech and Hearing Research, 25*(4), 554–564. https://doi.org/10.1044/jshr.2504.554

Lynch, J. S., Van Den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology, 29*(4), 327–365. https://doi.org/10.1080/02702710802165416

McGregor, K. K. (2023). *Dynamics of word learning*. OSF. https://osf.io/r56s7/?view_only=ca5d0d5bc05a4862b44f356e7d0c277d

McGregor, K. K., Ohlmann, N., Eden, N., Arbisi-Kelm, T., & Young, A. (2023). Abilities and disabilities among children with developmental language disorder. *Language, Speech, and Hearing Services in Schools, 54*(3), 927–951. https://doi.org/10.1044/2023_LSHSS-22-00070

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., Pomper, R., Eden, N., Arbisi-Kelm, T., Ohlmann, N., Gajre, S., & Smolak, E. (2021). Children's language abilities predict success in remote communication contexts. *Language Development Research, 1*(1), 245–282. https://doi.org/10.34758/rzsq-zd16

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), Article e13094. https://doi.org/10.1111/cogs.13094

McMurray, B., Apfelbaum, K. S., & Tomblin, J. B. (2022). The slow development of real-time processing: Spoken-word recognition as a crucible for new thinking about language acquisition and language disorders. *Current Directions in Psychological Science, 31*(4), 305–315. https://doi.org/10.1177/09637214221078325

McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition, 191,* Article 104000. https://doi.org/10.1016/j.cognition.2019.06.012

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research, 57*(4), 1344–1362. https://doi.org/10.1044/2014_JSLHR-L-13-0196

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237–270. https://doi.org/10.3102/00028312024002237

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

Nelson, N. W., Plante, E., Helm-Estabrooks, N., & Hotz, G. (2016). *Test of Integrated Language and Literacy Skills (TILLS)*. Brookes.

Norbury, C. F., & Bishop, D. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language and Communication Disorders, 37*(3), 227–251. https://doi.org/10.1080/13682820210136269

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick Incidental Learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research, 38*(2), 434–445. https://doi.org/10.1044/jshr.3802.434

O'Hara, M., & Johnston, J. (1997). Syntactic bootstrapping in children with specific language impairment. *International Journal of Language & Communication Disorders, 32*(2), 189–205. https://doi.org/10.3109/13682829709020403

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

Pile, E. J., Girolametto, L., Johnson, C. J., Chen, X., & Cleave, P. L. (2010). Shared book reading intervention for children with language impairment: Using parents-as-aides in language intervention. *Canadian Journal of Speech-Language Pathology & Audiology, 34*(2).

Pinheiro, J., Bates, D., & R Core Team. (2023). *nlme: Linear and nonlinear mixed effects models*. R package (Version 3.1-163). https://CRAN.R-project.org/package=nlme

Pomper, R., McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Ohlmann, N. (2022). Direct instruction improves word learning for children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 65*(11), 4228–4249. https://doi.org/10.1044/2022_JSLHR-22-00300

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. 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

Rice, M. L., Cleave, P. L., & Oetting, J. B. (2000). The use of syntactic cues in lexical acquisition by children with SLI. *Journal of Speech, Language, and Hearing Research 43*(3), 582–594. https://doi.org/10.1044/jslhr.4303.582

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

RStudio Team. (2020). *RStudio: Integrated Development for R*. RStudio. http://www.rstudio.com/

Schneider, J. M., Abel, A. D., & Maguire, M. J. (2023). Vocabulary knowledge and reading comprehension account for SES-differences in how school-aged children infer word meanings from sentences. *Language Learning and Development, 19*(4), 369–385. https://doi.org/10.1080/15475441.2022.2081573

Shulman, C., & Guberman, A. (2007). Acquisition of verb meaning through syntactic cues: A comparison of children with autism, children with specific language impairment (SLI) and children with typical language development (TLD). *Journal of Child Language, 34*(2), 411–423. https://doi.org/10.1017/S0305000906007963

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

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

Snedeker, J. (2000). *Cross-situational observation and the semantic bootstrapping hypothesis*. In *Proceedings of the 30th Annual Child Language Research Forum*. Center for the Study of Language and Information.

Steele, S. C., & Watkins, R. V. (2010). Learning word meanings during reading by children with language learning disability and typically-developing peers. *Clinical Linguistics & Phonetics, 24*(7), 520–539. https://doi.org/10.3109/02699200903532474

Tomblin, J. B., & Nippold, M. A. (2014). *Understanding individual differences in language development across the school years*. Psychology Press.

Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M. (2003). The stability of primary language disorder: Four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research, 46*(6), 1283–1296. https://doi.org/10.1044/1092-4388(2003/100)

Van der Lely, H. K. (1994). Canonical linking rules: Forward versus reverse linking in normally developing and specifically language-impaired children. *Cognition, 51*(1), 29–72. https://doi.org/10.1016/0010-0277(94)90008-6

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities, 34*(9), 2586–2597. https://doi.org/10.1016/j.ridd.2013.05.014

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *TOSREC: Test of Silent Reading Efficiency and Comprehension*. Pro-Ed.

Wechsler, D. (2011). *WASI-II: Wechsler Abbreviated Scale of Intelligence–Second Edition*. PsychCorp. https://doi.org/10.1037/t15170-000

Wickham, H., Vaughan, D., & Girlich, M. (2023). *tidyr: Tidy Messy Data*. R package (Version 1.3.0). https://CRAN.R-project.org/package=tidyr