Research Article

Autism & Developmental Language Impairments

Autism & Developmental Language Impairments
Volume 5: 1–19

DOI: 10.1177/2396941519899311
journals.sagepub.com/home/dli
SAGE

The word learning profile of adults with developmental language disorder

Karla K McGregor
Boys Town National Research Hospital, Omaha, NE, USA; University of Iowa, Iowa City, IA, USA

Timothy Arbisi-Kelm and Nichole Eden
Boys Town National Research Hospital, Omaha, NE, USA

Jacob Oleson
University of Iowa, Iowa City, IA, USA

Abstract

Background and aims: Previous investigations of word learning problems among people with developmental language disorder suggest that encoding, not retention, is the primary deficit. We aimed to replicate this finding; test the prediction that word form, not the linking of form to referent, is particularly problematic; and determine whether women with developmental language disorder are better word learners than men with developmental language disorder.
Methods: Twenty adults with developmental language disorder and 19 age-, sex-, and education-matched peers with typical language development attempted to learn 15 words by retrieval practice. Their retention was measured one day-, one week-, and one month after training.
Results: The participants with developmental language disorder required more exposures to the word-referent pairs than the participants with typical language development to reach mastery. While training to mastery, they made more errors in word form production, alone or in combination with errors in linking forms to the correct referents, but the number of no attempts and pure link errors did not differ by group. Women demonstrated stronger retention of the words than men at all intervals. The developmental language disorder and typical language development groups did not differ at the one-day- and one-month retention intervals but the developmental language disorder group was weaker at the one-week interval. Review via retrieval practice at the one-day and one-week interval enhanced retention at the one-month interval; the review at one week was more beneficial than the review at one day. Women benefitted more from the review opportunities than men but the developmental language disorder and typical language development groups did not differ.
Conclusions: Adults with developmental language disorder present with weaknesses in the encoding of new words but retention is a relative strength. Encoding word forms is especially challenging but encoding word-to-referent links is not. We interpret this profile, and the evidence of a female advantage, as consistent with the Procedural Circuit Deficit Hypothesis.
Implications: When treating a client with developmental language disorder whose goal is to increase vocabulary knowledge, the interventionist should anticipate the need for multiple exposures to new words within activities that highlight the forms of the words and support their memory and production. Periodic review should serve to support long-term retention.

Keywords

Developmental language impairment, lexical development, memory, specific language impairment, vocabulary

Corresponding author:
Karla K McGregor, Boys Town National Research Hospital, 555 N. 30th St., Omaha, NE 68131, USA.
Email: karla.mcgregor@boystown.org

Introduction

Developmental language disorder (DLD), sometimes referred to as specific language impairment, is a highly prevalent (Norbury et al., 2016) neurodevelopmental condition (APA, 2013) that is characterized by below age-level attainments in spoken language comprehension and expression (Leonard, 2014), heightened risk for reading and other learning disabilities (Young et al., 2002), and notable rates of comorbidity with Developmental Coordination Disorder and Attention-Deficit/Hyperactivity Disorder (ADHD) (Beitchman et al., 1996; Hill, 2001). Although the condition is evident in early childhood, it persists into adulthood (Johnson et al., 1999). People with DLD may present with word learning problems, which, if longstanding, can lead to vocabularies that lack breadth and depth (McGregor, Oleson, Bahnsen, & Duff, 2013) and, ultimately, to functional limitations. Specifically, people who have limited vocabularies are disadvantaged in academic settings and this is true for children (Bleses, Makransky, Dale, Hojen, & Ari, 2016; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015) and adults (Dollinger, Matyja, & Huber, 2008).

Word learning problems among adults with DLD

Whereas it is well documented that word learning problems affect children with DLD (see Kan & Windsor, 2010 for a meta-analysis), our understanding of the extent and nature of word learning problems among adults with DLD is limited. Jackson, Leitão, Claessen, and Boyes (2019b) identified 70 papers on word learning among people with DLD, only five of which included adults. In one, Alt and Gutmann (2009) compared college students with spoken and written language disorders, with or without Attention-Deficit-Hyperactivity Disorder (ADHD) to their peers with typical language and attention on a task tapping the fast mapping of semantic features and word forms. Probes immediately after presentation of the novel objects and names revealed poorer recognition of both semantic features and word forms on the part of the language disorder + ADHD group as compared to the typical group. The students with language disorders but no attention problems scored between the two other groups and differed from neither. Long-term retention of the word forms and the semantic information was not measured.

In a second study, Bishop, Barry, and Hardiman (2012) examined the learning and retention of new word forms, in the absence of semantic referents, in children with DLD and their adult relatives compared to children with typical language development and their relatives. The task involved repeating novel words five times each in succession, a 1-hour break, and then repeating the novel words again. Whereas a single administration of a nonword repetition task is generally viewed as a measure of phonological short-term memory, the investigators took improvement over the five repeated administrations as indicative of learning and performance after the 1-hour break as evidence of retention. The study was not meant to be about the manifestation of DLD in adulthood, and thus DLD among the adult relatives was not verified. The adults from families with DLD did, however, present with poorer performance on a single administration of the standard NEPSY nonword repetition test than the adults from typical language families. Nevertheless, they did not demonstrate problems learning or retaining the word forms after repeated administrations.

The other three papers reviewed in Jackson et al. (2019b) were from our laboratory. In contrast to the other two, we have consistently found that word learning is challenging for adults with DLD, but our studies differed from Alt and Gutmann (2009) by measuring word learning with production rather than recognition probes and from Bishop et al. (2012) by verifying the DLD diagnosis in the adult participants. We also routinely measure longer term retention.

The primary goal of the three aforementioned studies (McGregor, Arbisi-Kelm, & Eden, 2017; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017; McGregor, et al., 2013) was to determine whether the word learning problem was a matter of encoding or retaining new words. Each study involved an independent sample of college students with DLD and their peers with typical language development (TD) who were similar in age and education. Although methodological details varied, in all studies, we trained the students on a set of novel words and their referents. The DLD and TD groups received the same numbers of exposures to the to-be-learned information. We then compared the groups on recall immediately after training and one week later. In McGregor et al., (2013), the students with DLD recalled fewer word forms and meanings than their TD peers immediately after training, suggesting that encoding is problematic. Over the week, the DLD-TD gap in the recall of meaning remained stable, suggesting no problems with retention. However, the DLD-TD gap in the recall of word forms widened, suggesting, perhaps, that retention of word forms is a problem. That said, there were also 12-hour and 24-hour retention tests that served to provide more exposure to the word forms (and thus more opportunities for encoding) before the one-week test. In the next two studies, we took greater care to reduce this confound. In McGregor, Arbisi-Kelm et al., (2017) and again in McGregor, Gordon, et al., (2017), we found additional evidence of an encoding deficit but no evidence of a retention deficit. Moreover, we found that having the participants engage in retrieval-based practice during training, a method known to facilitate encoding (Karpicke & Grimaldi, 2012), served to close the DLD-TD word learning gap (McGregor, Gordon, et al., 2017).

Word learning problems among children with DLD

Although contrary evidence exists (Rice, Oetting, Marquis, Bode, & Pae, 1994; Riches, Tomasello, & Conti-Ramsden, 2005), the word learning problems of children with DLD often present as encoding deficits but intact retention as well (Bishop & Hsu, 2015; Haebig et al., 2019; Leonard et al., 2019; Storkel, Komesidou, Fleming, & Romine, 2017). In Haebig et al. (2019) and Leonard et al. (2019), two different cohorts of 5-year-olds with DLD and their TD age-matched peers learned novel words via retrieval practice and their recall of the words was tested after a 5-minute and one-week delay. In both cases, the children with DLD named the novel referents less accurately than their peers on the first retrieval trial but their rate of learning thereafter was similar. One week later, the groups did not differ in their recall of the words. In Bishop and Hsu (2015), 8-year-olds with DLD and their TD age-matched and grammar-matched peers were taught the names for unfamiliar animals over two weeks. The children also learned nonverbal sounds linked to visual patterns via the same training protocol. The children with DLD learned fewer of the animal names than their age-mates, but they did just as well as their peers when learning the sound-visual associations. As in Haebig et al. (2019) and Leonard et al. (2019), the word learning problem was evident early in training, and the children with DLD did not fall further behind their TD peers from session to session. Thus, initial encoding, but not subsequent encoding or retention, was problematic. In a clinical treatment study involving teaching kindergartners with DLD new words via interactive book reading, Storkel et al. (2017) found variable presentations such that individual children had problems with encoding, retention, or both, but the overall pattern was that most of the children who responded poorly to the intervention were struggling with encoding.

If the problem is typically one of encoding new information into long-term memory, then more exposure to that information should attenuate the problem. In fact, a frequent finding is that children with DLD need more exposures to words and their referents than their typical age-mates (Gray, 2003, 2004, 2005; Rice et al., 1994) and, in some cases, younger typical language mates (Riches et al., 2005; Windfuhr, Faragher, & Conti-Ramsden, 2002) to attain the same word learning outcomes.

In a series of studies, Gray (2003, 2004, 2005) taught four novel word-referent pairs to four- and five-year-old until they reached an a priori criterion of comprehension and production mastery. In each study, the children needed more exposure to produce the word correctly than to match a given word to its referent in a 4-alternative forced choice (4AFC) task, like that used in the Peabody Picture Vocabulary Test (PPVT-4, Dunn & Dunn, 2007). Across studies, the children with DLD required between 1.18 times the number of exposures (the 4AFC task in Gray, 2004) and 2.07 times the number of exposures (the production task in Gray, 2003) required by their peers to reach mastery. In Gray's more recent work, she and her colleagues did NOT find differences between DLD and age-matched TD groups in word learning—whether measured by production or comprehension tasks during encoding or retention (Gray & Brinkley, 2011; Gray, Brinkley, & Svetina, 2012). In these studies, the authors used different word sets so that they could manipulate the neighborhood density and phonotactic probability of the words to be learned and they trained the words in a supportive, interactive setting. In Gray, Pittman, and Weinhold (2014), using a third set of words and computerized training, children with DLD learned to identify trained referents as quickly as their age-mates but they were slower to learn to produce the names of those referents. Thus, it appears that children with DLD often, but not always, require more exposures than their peers to learn new words.

The present study

The encoding deficit hypothesis. In the current study, we manipulated the number of exposures to the words in the training set. Our goal was not to test a clinical intervention per se, but rather to test the hypothesis that the word learning problems characteristic of adults with DLD involve challenges with encoding but not retention. Unlike our previous studies, we did not provide the same number of exposures to each participant. Instead, after one exposure to each referent and its name, the participant proceeded toward mastery via retrieval practice, that is, by naming the referent and receiving feedback on accuracy until he or she named at least 13 of 15 targets correctly. We then examined retention of the mastered word set by asking each participant to name 1/3 of the referents one day later, another 1/3 one week later, and all referents (the final 1/3 followed by the other two sets) one month later.

Given the encoding deficit hypothesis, we predicted that the adults with DLD would require more exposures to the word-referent pairs than the TD adults to reach equivalent levels of mastery. Moreover, we predicted that the adults with DLD would retain the words that they had encoded to mastery as well as their TD peers. Finally, because we asked the participants to name all referents at the one-month interval, we were able to determine how the opportunity to name the first set at the one-day interval and the second set at the one-week interval facilitated longer term retention. We predicted that such review would be beneficial for the TD group. We did not make a firm prediction about the DLD group. In the sense that naming the referents was an opportunity for further encoding and we had hypothesized encoding weaknesses for the DLD group, it could be that they would benefit less from these intervening exposures than their TD peers. On the other hand, at the one-day and one-week intervals, the participants were naming what was, by then, familiar information. Because the encoding problem that characterizes DLD tends to occur earlier, not later, in learning (Bishop & Hsu, 2015; Haebig et al., 2019; Leonard et al., 2019; McGregor, Gordon et al., 2017), the review could be more accurately considered as an opportunity to enhance the already encoded representation. In that sense, the DLD group might reap similar benefits as the TD group.

The word configuration hypothesis. The training to mastery paradigm allowed answers to an additional question about the word learning profiles of adults with DLD; namely, whether learning the word form configuration or linking the configuration to its referent is the greater challenge. Specifically, because the training involved learning via retrieval, the participants not only attempted to name the targets at the retention intervals but also during training. Therefore, we could explore the route to mastery by examining the errors made during each block of training. Imagine learning two new words: a/sɪmbək/is a prickly plant and a/peɪntɪʧ/ is a contraption that presses metal. Two possible naming errors are informative. Upon seeing the prickly plant, one learner might call it a/sɪmbəp/instead of a/sɪmbək/, but another learner might call it a/peɪntɪʧ/. The first error reveals a problem with the word form configuration, whereas the second reveals a problem linking a correct word form to its referent. Gray (2004) found that children with DLD who were poor word learners comprehended most of a set of newly trained words and could draw pictures of their referents well enough to demonstrate semantic knowledge, but they struggled to produce the word forms. Given this, together with our previous finding that, among adults with DLD, newly learned lexical-semantic knowledge was more stable than newly learned form configurations (McGregor et al., 2013), we predicted that the learning of word form configurations would present the greater challenge for adults with DLD. The associative linking of word forms to their referents would be a relative strength.

Framing this prediction is the Procedural Circuit Deficit Hypothesis which posits that people with DLD have weak procedural memory systems but intact declarative systems (Ullman, Earle, Walenski, & Janacsek, 2019; Ullman & Pierpont, 2005). This hypothesis could account for the general profile associated with DLD, that is, extraordinary problems with grammar but less severe problems with vocabulary (Leonard, 2014). Weaknesses in another procedural task, the serial reaction time task (Lum, Conti-Ramsden, Morgan, & Ullman, 2014), lend support to the hypothesis. Moreover, the procedural system is supported by frontal/basal ganglia circuits and these circuits also play a role in attention and working memory (McNab & Klingberg, 2008), additional areas of deficit among many with DLD (Archibald & Gathercole, 2006; Finneran, Francis, & Leonard, 2009; Lum, Conti-Ramsden, Page, & Ullman, 2012; Spaulding, Plante, & Vance, 2008).

Both declarative and procedural systems are critical to language learning and use but in different ways. The procedural system supports the implicit learning of new skills, especially complex motoric or cognitive sequences such as riding a bicycle, tying a shoe, or formulating a grammatical sentence. The declarative system supports the implicit and explicit learning and use of idiosyncratic knowledge such as facts and landmarks. There is some redundancy between the two systems and there can be tradeoffs between them according to task, context, and the individual learner's abilities (Ullman et al., 2019). Because lexical knowledge is explicitly available, we traditionally think of the lexicon as a declarative system. In truth, it is supported by both memory systems. For example, the declarative system supports long-term memory for the association of words to their referents, whereas the procedural system supports the segmentation of new words from the speech stream (Ullman et al., 2019). We argue that the procedural system must also play a role in learning unfamiliar word forms which are, after all, complex motoric sequences. In fact, learning longer words is harder for children with DLD than learning shorter words, evidence that suggests more complex motoric sequences are especially challenging (Jackson, Leitão, Claessen, Boyes, 2019a). A meta-analysis of the statistical learning of verbal sequences revealed reliable, age-invariant weaknesses among people with DLD (Lammertink, Boersma, Wijnen, & Rispens, 2017).

The female advantage hypothesis. A third and final question was whether the word learning profiles we obtain differ by sex. We had not explored sex differences in our earlier papers on word learning but in an analysis of children's verbal definitions, we found that girls exhibited greater breadth and depth of extant vocabulary than boys, and this was true of both typical and DLD groups (McGregor, Oleson, Bahnsen, & Duff, 2013). Given a female advantage for the learning, memory and use of verbal information in the general population of children (Bauer, Goldfield, & Reznick, 2002; Bornstein & Cote, 2005; Burman, Bitan, & Booth, 2008; Kramer, Delis, Kaplan, O'Donnell, & Prifitera, 1997) and adults (Halpern & LeMay, 2000; Kaushanskaya, Marian, & Yoo, 2011), we predicted that the women would encode and retain the new words better than the men. Ullman (2016) argues that females have stronger declarative memory systems than males and, thus, when procedural memory is weak, as is the case in DLD, affected females can draw upon their declarative memory to compensate. Therefore, we predicted the female advantage for both typical and DLD groups. A viable alternative prediction is that we would find no sex differences at all. We could be under-powered to detect small effects, and our participants are adults. There is evidence to suggest that sex differences in language abilities are small and that these small differences dwindle by adulthood (Wallentin, 2009).

In summary, the goal of the current study was to establish a profile of word learning among adults with DLD. Specifically, we aimed to replicate the finding that encoding is a relative weakness, but retention is a relative strength for individuals with DLD, and to test the hypothesis that word form configuration is more challenging than the linking of forms to referents. Finally, we asked whether women with DLD demonstrate milder word learning problems relative to their typically developing peers than men with DLD.

Method

Participants

Participants were recruited and tested according to a protocol approved by the Internal Review Board at the University of Iowa in compliance with the Declaration of Helsinki. Forty-six college students (aged 18–25) from the Midwest of the USA answered various campus-wide advertisements for people with and without language learning difficulties to participate in research. To qualify for the study, the student had to earn a standard score of 85 or better on the nonverbal matrices of the Kaufman Brief Intelligence Test (KBIT-2; Kaufman & Kaufman, 2004). One person scored below this threshold and so did not enroll in the study. The students were required to pass a pure-tone audiometric screening at 0.5, 1, 2, and 4 kHz at 25 dB bilaterally. One person was excluded for this reason.

To qualify for the DLD group, students had to report concerns about their ability to learn and use language. We asked whether they now, or in the past, had difficulty in understanding or using spoken or written language. We also sought to verify their DLD with standardized tests; however, few language tests for young adults exist and those few do not report their sensitivity and specificity for DLD identification (Fidler, Plante, & Vance, 2011). Fidler et al. (2011) administered a variety of measures that tapped phonology, morphology, syntax, semantics, and narrative to typical adults and adults who had learning disabilities, a history of speech-language services, or a family history of language impairments. They found three of these—a 15-item spelling test for assessing knowledge of irregularly spelled words; the word definition subtest of the Clinical Evaluation of Language Fundamentals-4 (CELF-4, Semel, Wiig, & Secord, 2003); and the Modified Token Test (Morice & McNicol, 1985) for assessing syntactic comprehension—to most accurately identify those who had learning disabilities. Two of those three, the spelling and Token tests most accurately identified those with a history of speech-language therapy. Because we wished to establish a word learning profile for adults with DLD in general, not just those who have clinically significant word learning problems, we did not want to bias our sample by depending on a word definition task for identification. Therefore, we used the recommended cutoff on the two-part screener, the spelling and Token tests. Combined, these have a sensitivity of 80% and a specificity of 87% for identification of a history of speech-language needs that required service. The measures were weighted per the procedure used by Fidler et al. (2011) as modified in Fidler, Plante, and Vance (2013). Five people who reported concerns did not meet the score threshold set by these measures and thus did not enroll in the study. To qualify for the TD group, students had to report no concerns about their spoken or written language ability, and score above the threshold recommended in Fidler et al. (2011).

Given these criteria, 20 students (10 women) qualified for the DLD group and 19 (10 women) for the TD group and enrolled in the study. Among the DLD group, 12 received classroom accommodations because of problems related to speaking, reading or writing, four had not sought accommodations, one was in the process of being evaluated for accommodations, and three did not report the status of any accommodation. One person reported a diagnosis of ADHD. In the TD group, two people reported a diagnosis of ADHD; one of the two had classroom accommodations, and the other had not sought accommodations. One person reported a diagnosis of clinical anxiety, and that person was also receiving accommodations given this condition.

Both groups were primarily Caucasian. Two participants with DLD were African American. Two participants in the TD group reported their ethnic background as Hispanic or Latino, and one reported race as Asian. Participants were matched for type of post-secondary institution, which in our sample were major universities (10 DLD, 10 TD) and community colleges (nine DLD, nine TD). One participant with DLD attended a four-year college.

To further describe the two participant groups, we administered standardized tests to assess vocabulary, memory, and additional language skills. The DLD group scored significantly lower than the TD group on phonological memory, sentence memory, receptive vocabulary, and defining vocabulary and these were large effects (Table 1); although, not surprising given a sample recruited from post-secondary settings, the mean scores of the DLD group were still within an average range on most. In the DLD group, there were below-average scores from four people on the CELF-3 repeating sentences, nine people on the CVLT-2, one person on the PPVT-4, and one person on the EVT-2. In the TD group, eight people scored

Table 1. Demographic characteristics and cognitive/linguistic scores of the DLD and TD groups.

			DLD	TD			
Domain	Measure	Unit of measure	Mean (SD)	Mean (SD)	Normed mean	p	d
Age	Self-report	Years	20.7 (1.2) 18.75–23.00	20.9 (1.9) 18.42–24.00		.67	.14
Education	Self-report	Years	13.7 (1.3) 12–16	14.1 (1.8) 12–17		.48	.25
Nonverbal IQ	KBIT-2	Standard score	98 (10.8) 85–120	110 (12.8) 90–130	100 (15)	.0025*	1.01
Spelling	Probe[a]	Raw score out of 15	4.6 (2.9) 0–11	11.8 (2.3) 7–15		<.0001*	2.75
Sentence comprehension	Token test	Raw score out of 44	35 (5.5) 23–44	40 (3.1) 33–43		.0016*	1.12
Phonological memory	Probe[b]	Raw score out of 96	86 (6.4) 74–96	91 (3.3) 81–95		.0045*	.98
Sentence memory	CELF-3 Sentence recall	Standard score	9.6 (3.0) 3–15	11.68 (2.1) 7–15	10 (3)	.0179*	.80
List memory	CVLT-2	t-score	42 (11.7) 20–70	45 (11.3) 29–68	50 (10)	.40	.26
Receptive vocabulary	PPVT-4	Standard score	100 (13.4) 71–130	110 (9.9) 96–131	100 (15)	.0136*	.85
Expressive vocabulary	EVT-2	Standard score	106 (14.6) 81–127	113 (8.2) 97–130	100 (15)	.10	.59
Defining vocabulary	CELF-4 Word definition	Standard score	11.4 (2.3) 8–15	13.1 (1.3) 11–15	10 (3)	.0091*	.91

KBIT: Kaufman Brief Intelligence Test; CELF: clinical evaluation of language fundamentals; CVLT: California Verbal Learning Test; PPVT: peabody picture vocabulary test; EVT: expressive vocabulary test.
[a]Fidler et al. (2011).
[b]Dollaghan & Campbell (1998).
*$p < .02$.

below average on the CVLT-2 but no one scored below average on the language tests.

Stimuli

The stimuli were novel words that named unusual objects. We used novel words to ensure that no one had prior knowledge of the word forms. The decision to use these novel words as nouns that name concrete objects reflected the goal of testing the encoding deficit hypothesis. Specifically, to test that hypothesis, we needed a learning task that depended largely on the ability to encode the co-occurrence of a word and a visual referent, which is the case with concrete object names. Other sorts of words, like adjectives or verbs, would depend on additional processes such as sentence comprehension and the use of that sentence to infer meaning and word class.

There were three sets of five novel word stimuli (Table 2). To better simulate real word learning, we derived the novel words from real English words by manipulating only the final consonants. All English words were disyllabic nouns with primary stress on the first syllable. The English words had familiarity ratings (Nusbaum, Pisoni, & Davis, 1984) between 5 and 7 (scale 1–7; mean 6.83, SD 0.33); and each had an age of acquisition rating (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) below 10 years (mean 6.36, SD 1.85). The mean of the positional segmental frequency of all segments in English words was .063 (SD = .014). After modifying the final consonant, the mean of the positional segmental frequency of all segments in the novel words was .057 (SD = .013) (Vitevitch & Luce, 2004).

To equate the learnability of the three sets, structure variations (i.e. CV.CVC, CVC.CVC, CCV.CVC) and phonological feature distribution (i.e. place, manner, and voicing) of word-initial and word-final segments were balanced across the three sets. No word-initial or word-final segment occurred more than twice within each set.

A female native speaker of Standard American English recorded the audio stimuli and task instructions in a sound-treated booth using a Larson-Davis 2560 0.5-inch random incidence microphone. The recorded signals were routed to the computer through a Larson-Davis PRM902 preamplifier, a LISTEN SoundConnect microphone power supplier, and a MOTU Ultralite-mk3 Hybrid sound interface. The stimuli were then digitized at a 44.1 kHz sampling rate and 16-bit resolution using the sound-editing software Adobe Audition 1.0.

Each novel word was randomly assigned a visual referent, specifically an unusual plant, animal, or inanimate object that none of the laboratory personnel could name. These were depicted in digital images found via Google. Each of the three sets included three animate two inanimate referents. The ordering of sets during training was counterbalanced across participants.

Protocol

Data collection

Each participant visited the laboratory to give informed consent and complete the battery of standardized tests used for inclusion, exclusion, and description of language and cognitive abilities. The experimental protocol began when they returned to learn 15 novel words and their referents until reaching mastery, an a priori criterion of 13 of 15 referents named with an entirely correct word form. The participants returned at set intervals so that we could measure retention.

For training and retention, each participant sat facing a laptop wearing closed circumaural headphones with a Blue Snowball USB microphone placed approximately 18 inches away from the mouth. Audio and visual stimuli were presented using E-Prime 2.0, on a Dell Vostro laptop. The data collection procedure is summarized in Table 3 and explained below.

Training

The first step was passive exposure to one of the three sets of five pictures and their names. The participant imitated each word immediately after hearing it without feedback. The next step involved active retrieval trials. The participant saw the five pictured referents

Table 2. Training stimuli presented in phonetic notation, matched with orthographic English neighbors.

Set A		Set B		Set C	
Trained	Neighbor	Trained	Neighbor	Trained	Neighbor
blɑsəd	blossom	gɑrlɪd	garlic	ʧɪpmʌz	chipmunk
dræɡəs	dragon	ʤækɪz	jacket	dɛzɚg	desert
kæktəb	cactus	mæɡnəf	magnet	kɑtəf	cottage
mɛləg	melon	baɪsət	bison	peɪntəʧ	painter
sɪmbək	cymbal	skutəv	scooter	spaɪdəp	spider

Table 3. Protocol.

Day	Goal	Stimuli	# of items	# of blocks	Participant's task
Training day	To begin encoding	Set 1[a] pictures and names	5	1	Look, listen, imitate
		Set 1 pictures	5	5[b]	Name w/feedback
		Set 2 pictures and names	5	1	Look, listen, imitate
		Set 2 pictures	5	5	Name w/feedback
		Set 3 pictures and names	5	1	Look, listen, imitate
		Set 3 pictures	5	5	Name w/feedback
		Set 1 pictures	5	2	Name w/feedback
		Set 2 pictures	5	2	Name w/feedback
		Set 3 pictures	5	2	Name w/feedback
	To reach mastery	Sets 1, 2, & 3 pictures	15	1 to 10[c]	Name w/feedback
	To reduce short-term memory for newly encoded words	Arithmetic problems			Solve for 20 seconds
	To measure encoding outcomes	Sets 1, 2, & 3 pictures	15	1	Name w/out feedback
1 day post	To measure retention over one day	Set 1 pictures	5	1	Name w/out feedback
	To enable re-encoding	Set 1 pictures	5	1	Name w/feedback
1 week post	To measure retention over one week	Set 2 pictures	5	1	Name w/out feedback
	To enable re-encoding	Set 2 pictures	5	1	Name w/feedback
1 month post	To measure retention over one month	Set 3 pictures	5	1	Name w/out feedback
	To determine whether re-encoding opportunities boosted long-term retention one month after initial training	Sets 1 & 2 pictures	10	1	Name w/out feedback

[a]The order of sets was counterbalanced across participants.
[b]When multiple blocks were presented, the order of items within blocks was randomized anew each time.
[c]Blocks continued until the participant named at least 13 of 15 items correctly or until the 10th block was completed.

one at a time and attempted to name each one. A chime played after a correct response. After an incorrect response, the participant heard the correct production and imitated it. Although the chimes and the productions of the correct words were computer generated, the examiner sitting next to the participant controlled those presentations. The block of five items in the first training set was presented for active retrieval practice five times. In each recurrence, the items within the block were presented in a new random order.

The one block of passive exposure followed by the five blocks of active retrieval with feedback was repeated for the second set of five items; then for the third set of five items. To give the participants a chance to practice after other items had been interleaved, we then cycled back through active retrieval practice: Naming with feedback began again for two blocks of the first set, followed by two blocks of the second set, and finally, two blocks of the third set. In total, all participants received one passive exposure to each item, and seven $(5 + 2)$ active exposures to each item, for a total of eight exposures before mastery training.

Mastery

The participant saw the 15 pictures from the three training sets combined in randomized order and named each with feedback as described above. The stimulus presentation program would automatically calculate a participant's total number of correct items from the decisions that the examiner had entered in the moment. If at least 13 of the 15 items were named correctly, the task would end. If fewer than 13 items were correct, participants undertook another try with the same 15 words, in a re-randomized order. The maximum number of presentations was 10.

Post-mastery naming task

Immediately following the mastery task, participants solved math problems for 20 s to reduce the likelihood that subsequent recall of the newly learned words reflected only short-term memory. Afterwards, participants were presented with all 15 of the trained picture items once more and asked to name them without feedback.

Retention

Approximately one day $(24 \pm 2$ hours) after training, participants returned to name their first training set of five items, once without feedback to determine retention, using the same procedure as post-mastery naming. Then they named them again with feedback to allow active review. Approximately one week (7 ± 1 day) after training, participants returned to name their second training set once without feedback, then again with feedback. Approximately one month (30 ± 3 days) after training, they returned to name their third training set without feedback. Finally, they named the first two sets (combined and in random order) without feedback.

Data analysis

To test the prediction that encoding, not retention is problematic, we examined the number of blocks to mastery as well as accuracy one day-, one week-, and one month after mastery in the DLD and TD groups.

To test the prediction that the word form, not the linking of forms to their referents, is problematic, we first transcribed the naming responses. Two independent transcriptionists working on the same 20% of the sample agreed segment-to-segment 97.6% of the time. The transcriptions were entered segment by segment into an excel spreadsheet. We then applied a matching algorithm, programmed in R, which automatically compared each production to each of the 15 targets. The result was a set of similarity metrics for each production that we used to determine the response type, as explained in Table 4. We binned picture naming attempts during training by response type, time, and diagnostic group.

To test both the encoding and word form predictions, we built relevant statistical models in SAS v9.4. The details of these appear alongside the results below. The data themselves and the code used for statistical analysis are available in Open Science Forum (https://osf.io/4v7j6/).

Results

Encoding

Five students with DLD did not reach mastery after 10 cycles. Given that their standard scores on the CELF-3 Recalling Sentences Subtest ranged from 3 to 8 (DLD peers ranged from 6 to 15) and on the PPVT-4 ranged from 71 to 95 (DLD peers ranged from 88 to 130), they were, arguably, among those with the most severe DLD presentations. None of these students reported a history of ADHD. One of the students with TD did not reach mastery. His test scores did not distinguish him from his TD peers. On the CELF-3 Recalling Sentences Subtest, he scored 14 (TD peers ranged from 7 to 15) and on the PPVT-4 he scored 113 (TD peers ranged from 96 to 131); he did not report ADHD.

We used a linear regression model to evaluate the effect of diagnostic group and sex, on the number of cycles required for mastery. The participants who did not reach mastery were included in this analysis and the number of cycles for them was entered as 10. This was a conservative decision in that it reduced the chances of finding the predicted difference between the DLD and TD groups. Diagnostic group, sex, and the diagnostic group × sex interaction explained 25% of the variance in cycles required for mastery with diagnostic group being statistically significant, $\beta = 2.89$, $t = 2.92$, $p = .0019$, 95% CI (0.88, 4.90), while sex was not, $\beta = -0.39$, $t = -0.28$, $p = .7774$, 95% CI (−3.16, 2.38), and the interaction was not, $\beta = 0.59$, $t = 0.31$, $p = .7591$, 95% CI (−3.28, 4.56). On average, students with DLD required 5.9 mastery cycles (SD = 3.4), and students with TD required 2.7 (SD = 2.5). Given that everyone received eight exposures prior to mastery training, the DLD group required 1.39 times the amount of exposure of the TD group to meet the mastery criterion (DLD: $8 + 5.9 = 13.9$; TD: $8 + 2.7 = 10.7$; $13.9/10.7 = 1.39$). The number of cycles required for mastery was significantly correlated with vocabulary size as measured by the PPVT-4 standard scores, for the DLD group, $r = -48$, $p = .032$, but not for the TD group, $r = -.08$, $p = .74$. Among the participants with DLD, the more encoding cycles required, the smaller the size of the extant vocabulary.

Table 4. Decision guide for interpreting similarity metrics.

Response Type	Definition	Example
Correct	100% match between the form produced and the form of the target name	/dræ.gəs/for the target/dræ.gəs/
Form error	<100% match between the form produced and the form of the target name; best match of the 15 possibilities	/dæ.gəs/for the target/dræ.gəs/
Link error	100% match between the form produced and the form of a non-target name	/dræ.gəs/for the target/kɑ.təf/
Link + form error	<100% match between the form produced and the form of a non-target name; best match of the 15 possibilities	/dæ.gəs/for the target/kɑ.təf/

Retention

Having demonstrated compromised encoding related to vocabulary outcomes among people with DLD, the next step was to test the prediction that retention is intact. For those who reached mastery, we asked whether naming performance differed by time, diagnostic group or sex. The primary research question had a dichotomous outcome and involved between-subject variables, so we used a longitudinal logistic regression analysis with generalized estimating equations (GEE) (Liang and Zeger, 1986) and an exchangeable correlation structure. The GEE approach for estimating regression parameters adjusts the model estimated standard errors due to correlated values that result from repeated measures by using the so-called sandwich estimator. The sandwich estimator provides valid standard errors resulting in valid statistical tests for the logistic regression beta coefficients. The fixed effects were time, group, sex, group × time and group × sex. The interaction term for group by sex was not included in the final model, X^2 (2 df) = 3.71, $p = 0.1566$.

There was a significant effect for sex with females (X = .59, SD = .04) scoring higher than males (X = .42, SD = .04), $\beta = .71$, z = 2.94, $p = .0033$, 95% CI (0.23, 1.18). There was a significant effect of time, with better performance at the one-day interval (X = .74, SD = .03) than the one-month interval (X = .21, SD = .04), $\beta = 2.34$, z = 8.03, $p < .0001$, 95% CI (1.77, 2.91), better performance at the one-day interval than the one-week interval (X = .60, SD = .05), $\beta = .61$, z = 2.60, $p = .0093$, 95% CI (.15, 1.07), and better performance at the one-week interval than the one-month interval, $\beta = 1.73$, z = 6.51, $p < .0001$, 95% CI (1.20, 2.26). There was a significant effect of diagnostic group with the TD group (X = .57, SD = .04) scoring higher than the DLD group (X = .44, SD = .05), $\beta = -1.06$, z = −2.65, $p = .0081$, 95% CI (−1.85, −.28). When we examined diagnostic group × time, we found no difference at the one-day interval, $\beta = -0.19$, z = −0.57, $p = .5662$, 95% CI (−.86, .47), or at the one-month interval, $\beta = -.36$, z = −.82, $p = .4134$, 95% CI (−1.23, .41). However, there was a difference at the one-week interval, $\beta = -1.06$, z = −2.65, $p = .0081$, 95% CI (−1.85, −.28). This pattern is evident in Table 5.

Review

One month after training, we compared recall of the first and second word sets, those reviewed at the one-day and one-week retention intervals, to recall of the third set, the set that had not been named since the training day. We use a GEE analysis with an exchangeable correlation structure. The fixed effects were diagnostic group, sex, time, diagnostic group × time, and diagnostic group × sex. There was no diagnostic group × time interaction effect, X^2 (2 df)=1.42, $p = .4924$, or diagnostic group × sex interaction, z = 0.99, $p = .3232$, so they were removed from the final model.

There was no diagnostic group effect, $\beta = -0.21$, z = −0.72, $p = .4694$, 95% CI (−.79, .36). There was a significant effect of sex, $\beta = .59$, z = 1.99, $p = .0462$, 95% CI (.01, 1.18), with females (X = .48, SD = .04) having a higher performance than males (X = .34, SD = .06). There was a time effect, X^2 (2 df)=19.84, $p < .0001$, such that words that were reviewed one week after training resulted in the highest mean performance (X = .63, SE = .05) at the one-month test; review one day after training was also helpful as it resulted in a mean performance (X = .42, SE = .05) that was higher than no review (X = .22, SE = .04). At the one-month test, the word sets reviewed one day after training, one week after training, or not at all were significantly different from each other (Table 6).

Naming response profiles during training

We asked whether naming response profiles differed by diagnostic group or sex as measured at each of the seven active training blocks, the first mastery block,

Table 6. Effect of review on naming performance one month after training.

Review interval	Estimate	Z	p	95% CI
1 day vs. 1 week	−.86	−4.12	<.0001	−1.27, −.45
1 day vs. no review	.96	3.49	.0005	.42, 1.49
1 week vs. no review	1.82	−6.38	<.0001	1.26, 2.38

Table 5. The effect of retention interval by group.

Group	Time	Estimate	Z	p	95% CI
DLD	Day vs. month	1.34	6.84	<.0001	.96, 1.72
DLD	Day vs. week	1.04	3.51	.0004	.46, 1.62
DLD	Month vs. week	−0.30	−1.45	0.1472	−.70, .11
TD	Day vs. month	1.34	3.83	.0001	.66, 2.03
TD	day vs. week	−0.17	−0.63	.6270	−53, .88
TD	Month vs. week	−1.17	−3.81	.0001	−1.77, −.57

and the post-mastery block for each of three sets of five items on the training day. For analysis purposes, the sets are collapsed such that each block includes 15 naming attempts. We ran a generalized linear mixed model for each of the possible response types: correct, form error, link error, link + form error, and no attempt. The outcome variable was 1 or 0 (that particular response was present or absent) with a logit link function. The fixed effects in the model were diagnostic group, sex, time, group $\times$ sex, and group $\times$ time. The time variable was categorical with nine levels. A random intercept for subject was included to account for within-subject correlation over time, and a random intercept for trial was included because the correlation between trials within a block was different from the correlation between trials from different blocks. A Bonferroni correction yields a significance level .05/9 = .0055. The analysis was carried out in SAS v9.4 Proc Glimmix using Laplace Approximation and residual degrees of freedom.

Correct responses. There was neither an effect for sex, $\beta = .06$, $t_{5180} = 0.17$, $p = .7975$, 95% CI (−0.67, 0.80), nor a diagnostic group x sex interaction, $\beta = -0.60$, $t_{5180} = -0.80$, $p = .4250$, 95% CI (−2.06, 0.87). There were effects for time, $F_{8,5180} = 86.23$, $p < .0001$, diagnostic group, $\beta = -1.45$, $t_{5180} = -3.88$, $p = .0001$, 95% CI (−2.19, 0.72), and their interaction, $F_{8,5180} = 2.74$, $p = .0052$. Correct responses, and all incorrect response types, are plotted by time and diagnostic group in Figure 1.

The diagnostic group by time interaction is presented in Table 7. There were no significant differences between the groups at block 1 when both groups were near floor; block 6 when both groups dipped as they shifted from word set 3 back to word set 1; or post-mastery when both groups approached ceiling. The groups differed at all other blocks and, in each case, the DLD group gave fewer correct names than the TD group. Except for dips at block 6 and beginning mastery corresponding to changes in the sets being named, correct naming responses tended to increase over time.

Form error responses. There was no effect of sex, $\beta = 0.11$, $t_{5180} = -0.51$, $p = .6083$, 95% CI (−0.32, 0.54), or diagnostic group $\times$ sex interaction, $\beta = 0.25$, $t_{5180} = 0.58$, $p = .5640$, 95% CI (−0.61, 1.12). There were effects of time, $F_{8,5180} = 16.43$, $p < .0001$, diagnostic group, $\beta = 0.85$, $t_{5180} = 3.87$, $p = .0001$, 95% CI (0.42, 1.29) and their interaction, $F_{8,5180} = 3.30$, $p = .0009$. The interaction between diagnostic group and time is presented in Table 7. Although the DLD group made more form errors than the TD group at each block, these differences were significant at block 4, block 5, block 7, beginning mastery, and post mastery only.

Link error responses. Link errors were infrequent. There were no link errors on the beginning mastery block for the TD group and none on the post mastery block for the DLD group. There was no effect of diagnostic group, $\beta = 1.29$, $t_{5180} = 0.04$, $p = .9679$, 95% CI

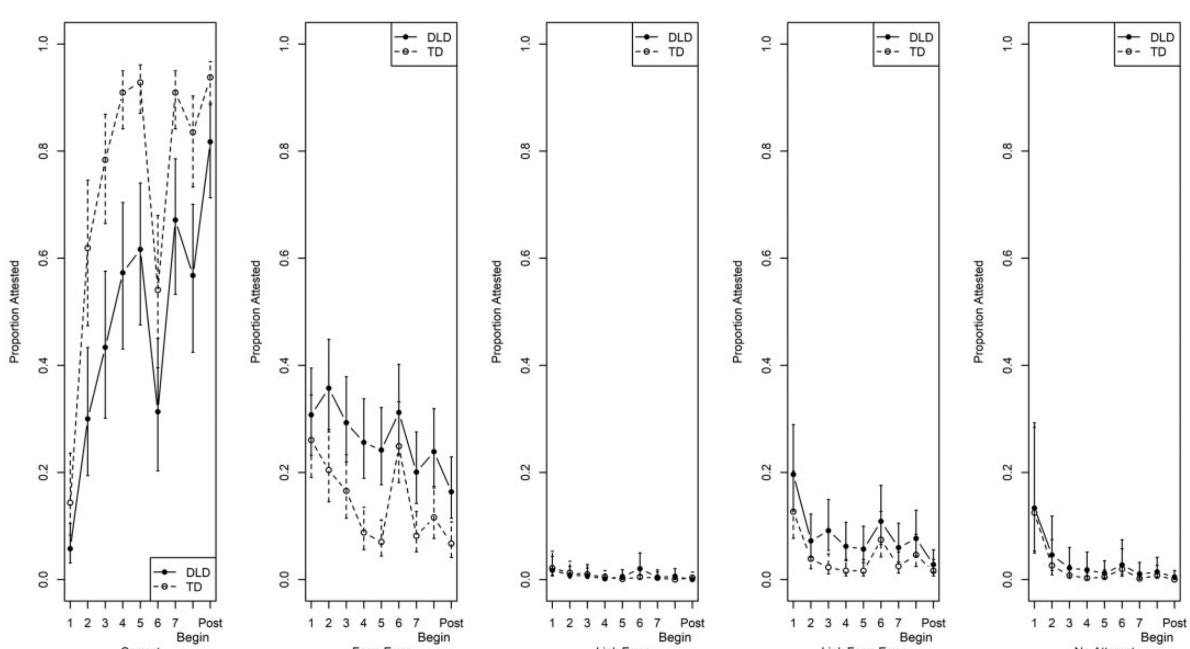

Figure 1. Naming responses during training by diagnostic group, time, and type expressed as proportion of total responses.

Table 7. Diagnostic group × time interactions in the correct responses and form errors made during training.

Response	Time	β	t_{5180}	p	95% CI
Correct	Block 1	−1.00	−2.22	.0265	−1.89, −0.12
	Block 2	−1.33	−3.16	.0016*	−2.16, −0.51
	Block 3	−1.56	−3.67	.0002*	−2.39, −0.72
	Block 4	−2.01	−4.61	<.0001*	−2.87, −1.16
	Block 5	−2.08	−4.71	<.0001*	−2.94, −1.21
	Block 6	−0.95	−2.24	.0252	−1.78, −0.12
	Block 7	−1.59	−3.62	.0003*	−2.45, −0.73
	Beginning mastery	−1.35	−3.15	.0017*	−2.19, −0.51
	Post mastery	−1.21	−2.70	.0069	−2.09, −0.33
Form error	Block 1	0.23	−0.82	.4105	−0.32, 0.79
	Block 2	0.77	2.70	.0070	0.21, 1.33
	Block 3	0.74	2.51	.0120	0.16, 1.31
	Block 4	1.28	4.04	<.0001*	0.66, 1.90
	Block 5	1.44	4.41	<.0001*	0.80, 2.09
	Block 6	0.31	−1.09	.2755	−0.25, 0.88
	Block 7	1.04	3.17	.0015*	0.40, 1.68
	Beginning mastery	0.88	2.84	.0046*	0.27, 1.48
	Post mastery	1.00	2.97	.0029*	0.34, 1.66

*$p < .0055$.

(−61.42, 64.00), no effect for sex, β =−0.75, $t_{5180} = -1.51$, $p = .1314$, 95% CI (−1.73, 0.23), no diagnostic group × sex interaction, β =−0.24, $t_{5180} = -0.24$, $p = .8123$, 95% CI (−2.19, 1.71), and no diagnostic group × time interaction, $F_{8,5180} = 1.52$, $p = .1569$. There was a time effect, $F_{8,5180} = 62.20$, $p < .0001$, involving a significant increase in link errors from block 5 to 6, β=1.38, $t_{5180} = 2.16$, $p = .0310$, 95% CI (0.12, 2.63), but not block 1 to 2, β=−0.34, t_{5180}−1.93, $p = .0542$, 95% CI (−1.32, 0.01), block 2 to 3, β =−0.22, $t_{5180} = -0.56$, $p = .5788$, 95% CI (−0.01, 0.56), block 3 to 4, β =−1.20, $t_{5180} = -1.88$, $p = .0595$, 95% CI (−2.45, 0.05), block 4 to 5, β =−0.01, $t_{5180} = -0.01$, $p = .9946$, 95% CI (−1.57, 1.56), block 6 to 7, β =−0.90, $t_{5180} = -1.75$, $p = .0802$, 95% CI (−1.90, 0.11), or beginning mastery to post mastery, β =4.69, $t_{5180} = 0.03$, $p = .9741$, 95% CI (−277.46, 286.83). In other words, when accuracy faltered as the training shifted from word set 3 on block 5 back to word set 1 on block 6, it was because of an increase in link errors among both groups.

Link + form error responses. There was neither a significant effect for sex, $\beta = 0.40$, $t_{5180} = 1.20$, $p = .2305$, 95% CI (−0.25, 1.04), nor an interaction between diagnostic group and sex, $\beta = 0.84$, $t_{5180} = 1.28$, $p = .2008$, 95% CI (−0.45, 2.13). There was an effect of time, $F_{8,5180} = 18.43$, $p < .0001$, and diagnostic group, $\beta = 0.86$, $t_{5180} = 2.56$, $p = .0104$, 95% CI (0.20, 1.52), but no diagnostic group × time interaction, $F_{8,5180} = 1.64$, $p = .2008$. Link + Form errors were infrequent, and, with a few exceptions, they tended to decline over time in both groups; however, the DLD group made more link + form errors overall than the TD group.

No attempts. No attempts did not vary with sex, β = −0.78, $t_{5180} = -1.13$, $p = .2587$, 95% CI (−2.14, 0.57), diagnostic group, β = 2.38, $t_{5180} = 0.04$, $p = .9687$, 95% CI (−116.63, 121.39), diagnostic group × sex, β = 1.25, $t_{5180} = 0.90$, $p = .3685$, 95% CI (−1.48, 3.98), or diagnostic group × time, $F_{8,5180} = 0.81$, $p = .3685$. There was a significant effect of time, $F_{8,5180} = 31.58$, $p < .0001$: Failures to attempt naming were infrequent and they tended to decline over time.

In summary, correct responses tended to increase with training while errors decreased. There were a few exceptions to the decreasing trend in errors that had to do with changes in the order of the sets being named. By the second training block, the DLD group made fewer correct responses and more form and link + form errors than the TD group. Link errors and no attempts did not differentiate the DLD and TD groups.

Naming response profiles during retention

For those who reached mastery, we asked whether naming response profiles differed by diagnostic group or sex as measured at each retention interval. During each interval, each participant named five items. We ran a generalized linear mixed model for each of the possible error responses: form error, link error, link + form error, and no attempt. The outcome variable was 1 or 0 (that response was present or absent) with a logit link function. The fixed effects in the model were diagnostic group, sex, group × sex and group × time. The time variable was categorical with three levels. A random intercept for subject was included to account for within-subject correlation over time, and a random intercept for trial was included because the correlation between trials within a given retention interval was different from the correlation between intervals. The required Bonferroni correction given three comparisons, yields a significance level .05/3 = . 016. The analysis was carried out in SAS v9.4 Proc Glimmix using Laplace Approximation and residual degrees of freedom. The naming responses by retention interval and diagnostic group appear in Figure 2.

Form error responses. There was no effect of time, $F_{2,487} = 0.01$, $p = .9891$, sex, $\beta = -0.37$, $t_{487} = -1.27$, $p = .2055$, 95% CI (−0.93, 0.20), diagnostic group × time, $F_{2,487} = 2.86$, $p = .0585$, or diagnostic group × sex, $F_{1,487} = 0.00$, $p = .9493$. There was a significant diagnostic group effect, $\beta = 1.12$, $t_{487} = 3.77$, $p = .0002$, 95% CI (0.54, 1.71). The DLD group made more form errors than the TD group.

Link error responses. There was no effect of diagnostic group, $\beta = -.51$, $t_{487} = -0.47$, $p = .6394$, 95% CI (−2.67, 1.64), sex, $\beta = -0.13$, $t_{487} = -0.21$, $p = .8333$, 95% CI (−1.27, 1.02), diagnostic group × sex, $F_{1,487} = 0.46$, $p = .4985$, or diagnostic group × time, $F_{2,487} = 0.19$, $p = .8277$. There was a time effect $F_{2,487} = 6.23$, $p = .0021$. The DLD group had no link errors at the one-day interval. There was an increase in link errors from the one-week- to the one-month interval, $\beta = 1.76$, $t_{487} = 3.53$, $p = .0005$, 95% CI (0.78, 2.75).

Link+form error responses. There was no effect of diagnostic group, $\beta = -.31$, $t_{487} = -0.46$, $p = .6439$, 95% CI (−1.62,1.00), sex, $\beta = -0.53$, $t_{487} = -0.85$, $p = .3935$, 95% CI (−1.75, 0.69), diagnostic group × sex, $F_{1,487} = 1.01$, $p = .3142$, or diagnostic group × time, $F_{2,487} = 2.04$, $p = .1305$. There was a time effect $F_{2,487} = 9.55$, $p < .0001$, characterized by a significant increase from the one-day- to the one-month interval, $\beta = 1.86$, $t_{487} = 3.67$, $p = .0003$, 95% CI (0.87, 2.86) and the one-week- to the one-month-interval, $\beta = 1.43$, $t_{487} = 3.35$, $p = .0009$, 95% CI (0.59., 2.27), but not the one-day- to the one-week-interval, $\beta = 0.43$, $t_{487} = 0.78$, $p = .4377$, 95% CI (−0.66., 0.52).

No attempts. There was no effect of diagnostic group, $\beta = 0.28$, $t_{487} = 0.27$, $p = .7846$, 95% CI (−1.70, 2.26), sex, $\beta = -0.73$, $t_{487} = -1.14$, $p = .2544$, 95% CI (−1.98, 0.52), diagnostic group × sex, $F_{2,487} = 0.36$, $p = .5491$, or diagnostic group × time, $F_{2,487} = 0.75$, $p = .4711$. There was a significant time effect $F_{2,487} = 7.71$, $p = .0005$. The TD group had 0 no attempts at the one-day interval so we performed no comparisons involving that interval. There was a significant increase from the one-week- to the one-month-interval, $\beta = 1.53$, $t_{487} = 3.93$, $p < .0001$, 95% CI (0.76, 2.29).

In summary, all error types increased as the length of the retention interval increased. Only form errors

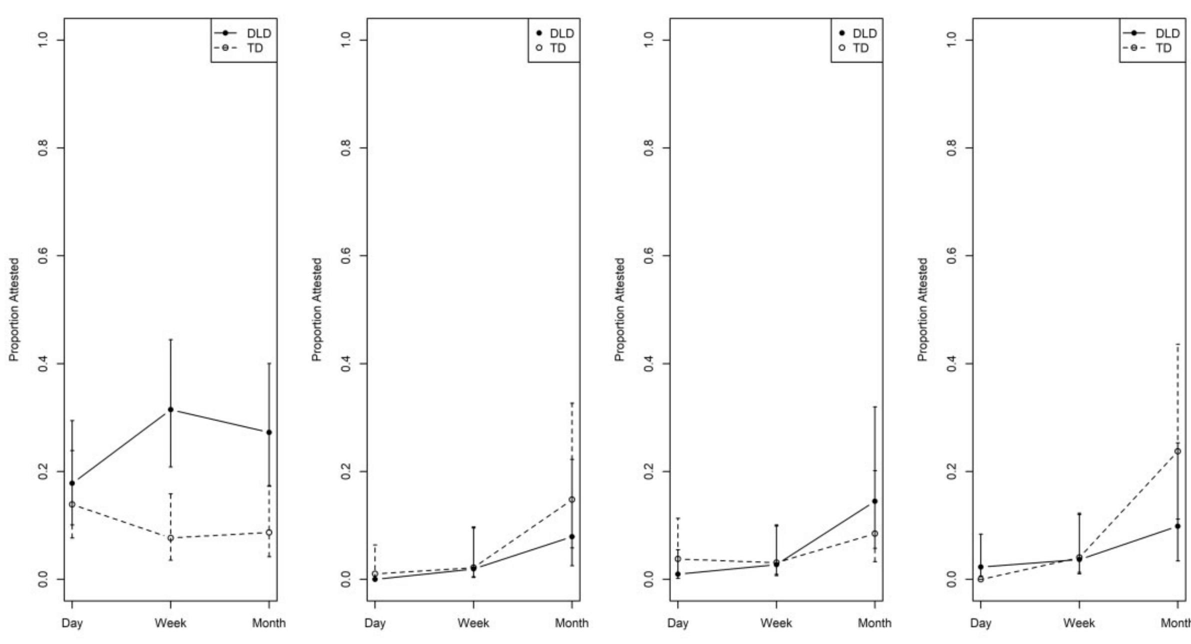

Figure 2. Naming responses during retention by diagnostic group, time, and type expressed as proportion of total responses.

distinguished the retention performance of the DLD and TD groups.

Discussion

In this study, people with DLD needed 39% more exposure than their peers with TD to commit a set of new words to long-term memory. In this specific case, the learners were young adults who, despite finding language to be challenging, were successful enough to enter postsecondary education. The training set comprised 15 disyllabic nonwords that named object referents. Each word had at least one English neighbor and, as a set, they had specific phonotactic properties. The training protocol involved retrieval practice plus feedback, and the outcome measure was the accuracy of referent naming. We list these details to emphasize that there is nothing magical about the 39% value. The amount of extra exposure required for people with DLD will vary with characteristics of the learner, stimuli, training, and outcome measure. In Gray's earlier work, 4- and 5-year-old with DLD needed 18% more exposure than their age-mates to learn four words when the outcome measure was to select the referent when given the word (Gray, 2004), and 107% more exposure when the outcome measure was naming the referent (Gray, 2003). The general point is that people with DLD, as a group, tend to require more exposures than other learners to encode new words into long-term memory.

Examination of the naming response profiles during training revealed that the encoding problem is specific to word forms. Although the participants with DLD did make linking errors and sometimes gave no response, they did so no more often than their peers with TD. In the next two sections, we summarize the evidence in support of the conclusion that the encoding of word forms is the primary bottleneck to word learning among people with DLD.

Encoding is problematic

That the participants with DLD required more exposures than the adults with TD to reach mastery during the training session stands as the primary evidence that encoding is problematic for people with DLD. The finding is congruent with our previous studies of adults with DLD (CITATIONS-REMOVED-FOR-BLIND-REVIEW) and with studies of children with DLD (Bishop & Hsu, 2015; Gray, 2003, 2004, 2005; Haebig et al., 2019; Leonard et al., 2019; Rice et al., 1994 ; Riches et al., 2005; Storkel et al., 2017; Windfuhr et al., 2002). The negative correlation between the number of encoding exposures required and the PPVT-4 scores of the participants with DLD suggests, but does not prove, that the encoding problem limits the eventual size of the lexicon.

The adults with DLD were not different from their peers with TD in the extent to which their long-term recall benefitted from re-encoding opportunities afforded by retrieval practice after training. For both groups, naming one month after training was improved by retrieval practice one day or one week after training, with greater benefit from the one-week practice. This finding reflects the nonmonotonic lag effect; namely, for any given retention interval, there is a practice interval that optimizes accuracy (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Simply put, to stave off forgetting, one needs to find a 'sweet spot' for review, not too close to teaching because of the risk of forgetting again by the time of the retention session, yet not too close to the retention session because, by then, what was initially learned may have been completely forgotten. To relate these patterns to our data, given a session of learning to mastery and a retention session one month later, the sweet spot for review was one week after learning/three weeks before retention, although perhaps a sweeter spot would have been identified had we tried other options. The participants with DLD, like their TD peers, demonstrated the monotonic lag effect.

The mechanism at play during review is debated. Cepeda et al. (2006) present two possibilities that have potential for explaining the current results. It could be that, upon review, a second memory trace for a given word is formed that inherits the consolidated state of the initial trace for that word. Another possibility is that the review provided an opportunity to retrieve the long-term memory trace established during the learning session. The act of retrieval strengthened that memory trace and increased contextual associations that could aid subsequent retrieval at the retention session. Note that, in either case, the proposed mechanisms involve supporting additional encoding via long-term memory. Therefore, the ability of the DLD group to benefit from review similarly to their typical peers is not inconsistent with the encoding hypothesis: It is not the re-encoding of familiar word forms but, instead, the encoding of NEW word forms that is problematic for adults with DLD.

By hypothesizing that encoding is the bottleneck that limits word learning, we imply the corollary, that retention is intact. We found this to be the case for the one-day and one-month interval: the naming accuracy of the DLD and TD groups did not differ at these points. However, we did find the DLD group to name less accurately than the TD group at the one-week interval. Both groups declined significantly between one day and one month, but for the DLD group, that decline occurred mainly over the first week; whereas, for the TD group, the decline occurred somewhere between the one-week- and one-month interval. In McGregor, Gordon, et al., (2017), we did not find differences between the DLD and TD groups in retention after a one-week interval, but the outcome measure was a task that taps competition between the new words and their English lexical neighbors. In McGregor, Arbisi-Kelm, et al., (2017), we did not find differences between the DLD and TD groups in retention one week after training and, in fact, we did not find declines in retention in either group, but the words and referents were trained via passive study not retrieval practice, and the production measure was stem completion rather than picture naming. That said, when examining the individual differences within the DLD and TD groups (Figure 2, McGregor, Arbisi-Kelm, et al., 2017), it is apparent that some people did experience notable forgetting (also see Storkel et al., 2017).

The result we obtained here, intact retention at one day and one month with a faster rate of forgetting between those two points on the part of the DLD group welcomes further exploration to determine whether different task demands or different presentations of DLD among affected individuals are at play. We conclude that there is robust evidence for a deficit in the encoding of new information but also a possibility that retention is problematic for some individuals at some retention intervals as well. We turn next to evidence that confirm that the encoding of word forms is especially fragile.

Word form configurations are problematic

The trajectories of growth in naming accuracy over the training interval were similar for the DLD and TD groups. The two groups did not differ at block one when performance was near floor nor at the post-mastery block when, by definition, all had trained to a high level of accuracy. During the sixth block, training shifted from the third word set back to the first and, in response, performance dipped for both groups. With those exceptions, the students with DLD demonstrated lower performance than the TD group during training due to more form errors, whether alone or in combination with link errors. As Bishop and Hsu (2015), Haebig et al. (2019), and Leonard et al. (2019) reported for children with DLD, the problem was apparent early in training, in this case, by the second training block. During retention, accuracy decreased over time for both DLD and TD groups, although at a steeper rate for the DLD group, as discussed above. Again, where performance faltered, the form errors, not the other error types, distinguished the DLD and TD groups. In fact, the rates of no responses and link errors were virtually identical in the DLD and TD groups.

We view this pattern as consistent with the Procedural Circuit Deficit Hypothesis (Ullman et al., 2019). Encoding the sequence of phonemes that comprise a word form early in the course of learning when that complex motoric sequence is highly unfamiliar may depend upon the procedural system. Retrieving and producing that word form after learning may depend, in part, upon the frontal/basal ganglia circuits that underlie procedural memory (McNab & Klingberg, 2008). Linking the word form to its referent, an instance of associative learning that is supported by the declarative system, is relatively intact (Ullman, 2016). A study comparing problems with word forms and grammar within the same sample of participants with DLD would be useful for determining whether sequential learning and memory deficits in the procedural system provide a parsimonious explanation for these seemingly different problems with language.

Females demonstrate stronger performance than males

Although men and women did not perform differently during training, the women demonstrated stronger retention and greater gains from review during the retention interval than the men. The female advantage held for both the TD and DLD groups, as it did in McGregor, Oleson, et al., (2013), where we mined a longitudinal corpus of verbal definitions collected from boys and girls with and without DLD.

Kaushanskaya et al. (2011) attribute the difference to women's ability to recruit long-term knowledge about their native language phonology as a support for word learning. Kramer and colleagues (Kramer, Delis, & Daniel, 1988; Kramer et al., 1997) attribute the female advantage to more efficient application of semantic clustering strategies that can facilitate verbal memory. Either of these possibilities could apply here as the novel words were derived from real English words, thus making native phonology relevant, and they named exemplars from plant, animal, and artifact categories, thus making semantic clustering relevant.

Ullman (2016) makes the case that females have stronger declarative memory systems than males and that, in tasks that can be supported by both declarative and procedural systems, females depend more on declarative processes such as chunking the material to be learned or processed. If so, then females who have DLD may be better able to compensate for weaknesses in the procedural system than males with DLD. Ullman (2016) links the female advantage to estrogen and its effect on neural function. Specifically, estrogens enhance synaptic activity and plasticity, and thus learning (Garcia-Segura, 2008). The gene *Cyp19a1* regulates aromatase, the enzyme that synthesizes estrogens, and, of particular interest here, variations in *Cyp19a1* alleles among families of probands with speech sound disorder and/or dyslexia predicted the performance of family members on measures of extant vocabulary as well as their repetition of multisyllabic (nonsense) words and rate of repetition of disyllables (Anthoni et al., 2012).

Boys are more prone to neurodevelopmental disorders than girls and DLD is no exception. Community samples reveal that girls are affected about 22% less often than boys (Norbury et al., 2016). Lower prevalence among girls, together with findings of stronger performance among women in this study and girls in McGregor, Oleson, et al. (2013), suggest that being female is a protective factor that limits the likelihood and/or the severity of DLD. Given our small sample, we view the finding that females with DLD are better at retaining newly learned words than males with DLD as a hypothesis, not a conclusion. We have much to learn about how sex—and the sociocultural aspects of gender—influence language learning in general and manifestations of DLD in particular.

Conclusions

This paper is the fourth in a series of papers (McGregor, et al., 2013; McGregor, Arbisi-Kelm, et al., 2017; McGregor, Gordon, et al., 2017) in which we have documented weaknesses in encoding but relative strengths in retention in independent samples of college students with DLD. Given this consistency, we are confident in concluding that encoding is a bottleneck for word learning in this population. In fact, given that our sample included only mild cases, we might expect that encoding is an even greater challenge in the population as a whole. In this paper, we also found that the encoding of word forms, not the linking of forms to their referents, is the crux of the problem. This profile of strengths and weaknesses held for both men and women with DLD although the women tended to be stronger at retention than the men.

This profile may provide a useful starting point for clinicians working to improve the vocabulary knowledge of their clients with DLD. Two specific implications emerge. First, it will be crucial to pay attention to adequate dosage. It is now well documented that people with DLD need more exposure to language targets than their typical peers to demonstrate comparable learning. This is not to say that high dosage alone is enough. The training study here was meant to manipulate amount of exposure as a means of testing the extent of the encoding deficit. In an actual language intervention, the clinician would not only increase the quantity of exposures but also their quality. For example, exposures that allow for encoding across modalities visual, auditory, and kinesthetic modalities (Jewitt, Kress, Ogborn, & Charalampos, 2001; Valenzeno, Alibali, & Klatzky, 2003) are useful and spaced exposures are more helpful than massed exposures (e.g., Seabrook, Brown, & Solity, 2005). Second, activities that highlight word forms may be indicated. For example, in (McGregor, Arbisi-Kelm et al., (2017)), we found that asking adults to identify key features of word forms (e.g., did it start w a 'b' sound?) yielded better learning of their form configurations than asking them to identify key features of word referents (e.g., did it have eyes?).

In the future, we would do well to verify word learning profiles among children with DLD as well as adults with DLD who have more severe language deficits. It would be useful to relate the word learning profile to more extensive measures of the language abilities and disabilities of the participants, especially given that we know little about how DLD manifests in adults. Moreover, the profile established here could be extended as many aspects of word knowledge, for example, contextually appropriate usage and the nuances of semantic knowledge, remain relatively unexplored in this population. The data we have collected in this series of studies might well underestimate the severity of the problem but also, perhaps, the range of the strengths exhibited by people with DLD.

Acknowledgements

We thank Drs. Katie Gordon, Chris Conway, and Erin Smolak for comments on an earlier draft.

Data accessibility statement

The raw data and the code for statistical analysis are available in the Open Science Forum (https://osf.io/4v7j6/).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Institutes of Health-NIDCD #5R01DC011742 awarded to Karla McGregor.

ORCID iD

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

References

Alt, M., & Gutmann, M. L. (2009). Fast mapping semantic features: Performance of adults with normal language, history of disorders of spoken and written language, and attention deficit hyperactivity disorder on a word-learning task. *Journal of Communication Disorders*, *42*, 347–364.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington, DC: American Psychiatric Pub.

Anthoni, H., Sucheston, L. E., Lewis, B. A., Tapia-Páez, I., Fan, X., Zucchelli, M., ... Pennington, B. F. (2012). The aromatase gene CYP19A1: Several genetic and functional lines of evidence supporting a role in reading, speech and language. *Behavior Genetics*, *42*(4), 509–527.

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675–693.

Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics*, *23*(3), 313–335.

Beitchman, J. H., Brownlie, E. B., Inglis, A., Wild, J., Ferguson, B., Schachter, D., ... Mathews, R. (1996). Seven-year follow-up of speech/language impaired and control children: Psychiatric outcome. *Journal of Child Psychology and Psychiatry*, *37*(8), 961–970.

Bishop, D. V. M., Barry, J. G., & Hardiman, M. J. (2012). Delayed retention of new word-forms is better in children than adults regardless of language ability: A factorial two-way study. *PLoS One*, *7*, 1–6.

Bishop, D. V., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology*, *3*(1), 3.

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476.

Bornstein, M. H., & Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities. *Infancy*, *7*(3), 299–316.

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia*, *46*(5), 1349–1362.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3): 354.

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

Dollinger, S. J., Matyja, A. M., & Huber, J. L. (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality*, *42*(4), 872–885.

Dunn, L. M., and Dunn, D. M. (2007). *Peabody picture vocabulary test*. 4th edition. San Antonio, TX: Pearson.

Fidler, L. J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, *20*, 2–13.

Fidler, L. J., Plante, E., & Vance, R. (2013). Erratum: Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, *22*, 577.

Finneran, D. A., Francis, A. L., & Leonard, L. B. (2009). Sustained attention in children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *52*(4), 915–29.

Garcia-Segura, L. M. (2008). Aromatase in the brain: Not just for reproduction anymore. *Journal of Neuroendocrinology*, *20*(6), 705–712.

Gray, S. (2003). Word-learning by preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *46*, 56–67.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research*, *47*, 1117–1132.

Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1452–1467.

Gray, S., & Brinkley, S. (2011). Fast mapping and word learning by preschoolers with specific language impairment in a supported learning context: Effect of encoding cues, phonotactic probability, and object familiarity. *Journal of Speech, Language, and Hearing Research*, *54*, 870–884.

Gray, S., Brinkley, S., & Svetina, D. (2012). Word learning by preschoolers with SLI: Effect of phonotactic probability and object familiarity. *Journal of Speech, Language, and Hearing Research*, *55*, 1289–1300.

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

Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., ... 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.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, *12*(2), 229–246.

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149–171.

Jackson, E., Leitão, S., Claessen, M., & Boyes, M. (2019a). 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.

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

Jewitt, C., Kress, G., Ogborn, J., & Charalampos, T. (2001). Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review*, *53*(1), 4–18.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., . . . Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research*, *42*(3), 744–760.

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*, 739–756.

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*(3), 401–418.

Kaufman, A. S., & Kaufman, N.L. (2004). *Kaufman brief intelligence test*. 2nd edition. Circle Pines, MN: American Guidance Services.

Kaushanskaya, M., Marian, V., & Yoo, J. (2011). Gender differences in adult word learning. *Acta Psychologica*, *137*(1), 24–35.

Kramer, J. H., Delis, D. C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology*, *44* (6), 907–915.

Kramer, J. H., Delis, D. C., Kaplan, E., O'Donnell, L., & Prifitera, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology*, *11*(4), 577–584.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavioral Research Methods*, *44*, 978–990.

Lammertink, I., Boersma, P., Wijnen, F., & Rispen, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *60*, 12, 3474–3486.

Leonard, L. B. (2014). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., . . . 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.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, *51*, 1–10.

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex*, *48*(9), 1138–1154.

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

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.

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*, 1845–1856.

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.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, *86*(5), 1351–1370.

Morice, R., & McNicol, D. (1985). The comprehension and production of complex syntax in schizophrenia. *Cortex*, *21*(4), 567–580.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2016). The impact of non-verbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words*. Bloomington, IN: Indiana University, Psychology Department, Speech Research Laboratory (Research on Speech Perception, Progress Report No. 10).

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

Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI. *Journal of Speech, Language, and Hearing Research*, *48*, 1397–1411.

Seabrook, R., Brown, G. D., & Solity, J. E. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology*, *19*(1), 107–122.

Semel, E., Wiig, E. H., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals*, Fourth Edition. San Antonio, TX: The Psychological Corporation.

Semel, E. M., Wiig, E. H., & Secord W. (1995). *Clinical evaluation of language fundamentals-3*. San Antonio, TX: Psychological Corporation.

Spaulding, T. J., Plante, E., & Vance, R. (2008). Sustained selective attention skills of preschool children with specific language impairment: Evidence for separate attentional capacities. *Journal of Speech, Language, and Hearing Research*, *51*, 16–34.

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

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S.L. Small (Ed.), *Neurobiology of language* (pp. 953–968). Cambridge, MA: Academic Press.

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2019). The neurocognition of developmental disorders of language. *Annual Review of Psychology*, *71*. Advance online publication. doi: https://doi.org/10.1146/annurev-psych-122216-011555

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*, 399–433.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187–204.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavioral Research Methods, Instruments, and Computers*, *36*, 481–487.

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language*, *108*(3), 175–183.

Windfuhr, K. L., Faragher, B., & Conti-Ramsden, G. (2002). Lexical learning skills in young children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, *37*(4), 415–432.

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry*, *43*(5), 635–645.