

Research Article

Encoding Deficits Impede Word Learning and Memory in Adults With Developmental Language Disorders

Karla K. McGregor,^a Katherine Gordon,^a Nichole Eden,^a
Tim Arbisi-Kelm,^a and Jacob Oleson^a

Purpose: The aim of this study was to determine whether the word-learning challenges associated with developmental language disorder (DLD) result from encoding or retention deficits.

Method: In Study 1, 59 postsecondary students with DLD and 60 with normal development (ND) took the California Verbal Learning Test–Second Edition, Adult Version (Delis, Kramer, Kaplan, & Ober, 2000). In Study 2, 23 postsecondary students with DLD and 24 with ND attempted to learn 9 novel words in each of 3 training conditions: uncued test, cued test, and no test (passive study). Retention was measured 1 day and 1 week later.

Results: By the end of training, students with DLD had encoded fewer familiar words (Study 1) and fewer novel words (Study 2) than their ND peers as evinced by word recall. They also demonstrated poorer encoding as evinced

by slower growth in recall from Trials 1 to 2 (Studies 1 and 2), less semantic clustering of recalled words, and poorer recognition (Study 1). The DLD and ND groups were similar in the relative amount of information they could recall after retention periods of 5 and 20 min (Study 1). After a 1-day retention period, the DLD group recalled less information that had been encoded via passive study, but they performed as well as their ND peers when recalling information that had been encoded via tests (Study 2). Compared to passive study, encoding via tests also resulted in more robust lexical engagement after a 1-week retention for DLD and ND groups.

Conclusions: Encoding, not retention, is the problematic stage of word learning for adults with DLD. Self-testing with feedback lessens the deficit.

Supplemental Materials: <https://doi.org/10.23641/asha.5435200>

Efficient learning of verbal information is characteristic of people with developmental language disorders (DLDs; also known as specific language impairment). The deficit is apparent on tasks that involve new configurations of familiar material, such as learning a list of familiar words (Sheng, Byrd, McGregor, Zimmerman, & Bludau, 2015) or recall of narrative passages (Plante, Ramage, & Magloire, 2006); tasks that involve unfamiliar material, such as learning new words (McGregor, Arbisi-Kelm, & Eden, 2017; McGregor et al., 2013); and tasks that involve both familiar and unfamiliar material, such as learning from a classroom lecture (Becker & McGregor, 2016). In this article, we ask whether the word-learning challenges faced by young adults with DLD are the result of deficits that impede the encoding

of information or deficits in long-term memory that impede the retention of that newly encoded information.

Encoding and Retention

Encoding is a set of processes involved in creating an initial memory trace. These processes include sensory perception, attention, and rehearsal. In the case of learning a new spoken word, encoding begins with the acoustic encoding of an echoic memory (Buchsbaum, Olsen, Koch, & Berman, 2005). Attention to the new word increases the likelihood that this ultrashort echoic memory will be held as a short-term memory in the phonological loop (Kane, Bleckley, Conway, & Engle, 2001). Subvocal rehearsal increases the likelihood that this short-term memory will become a long-term memory (Baddeley, 2003). The short-term storage and the processes acting upon information in the short-term store are captured by the construct of working memory.

Baddeley (2003) models working memory as a set of interacting modular components. The central executive

^aThe University of Iowa, Iowa City

Correspondence to Karla K. McGregor: karla-mcgregor@uiowa.edu

Editor-in-Chief: Sean Redmond

Editor: Lisa Archibald

Received January 23, 2017

Revision received May 26, 2017

Accepted June 5, 2017

https://doi.org/10.1044/2017_JSLHR-L-17-0031

Disclosure: The authors have declared that no competing interests existed at the time of publication.

controls attention to relevant incoming information. The phonological loop is a temporary, limited capacity store of information that is verbal. The visual-spatial sketchpad is a temporary, limited capacity store of information that is visual. The episodic buffer binds visual and verbal information as well as information in working memory with relevant information from long-term memory (Baddeley, 2003). Cowan (2010) models working memory as a temporary activation of items from long-term memory. The activated items are the focus of a limited capacity, domain-general attention system. New information that is the focus of attention can be encoded into long-term memory (although Cowan, 2010, does not address learning specifically). A central executive can place and displace items into the focus of memory as needed. What is common to the two models is the recognition that working memory is limited in capacity and that executive functions allocate attention to relevant information in working memory while shifting attention away from irrelevant information.

Retention of newly encoded information in long-term memory depends, in part, upon processes of consolidation. Through consolidation, the memory becomes more stable and more integrated with existing memories. Consolidation occurs in two overlapping waves. In the first wave, the memory trace becomes less prone to interference as a result of the modification and reorganization of synaptic proteins (Walker, 2005). In the second wave, the memory trace continues to strengthen, retrieval links form, and the new memory integrates into the existing memory network at the level of the neocortex (Gupta & Tisdale, 2009; McClelland, McNaughton, & O'Reilly, 1995; Walker, 2005). This wave is slower than the first, requiring hours, days, or even weeks (Dudai, 2004; Dudai & Eisenberg, 2004).

These various memory systems can be differentially impaired. For example, individuals with schizophrenia present with deficits in encoding but not retention, whereas individuals with depression tend to encode and retain, but they present with retrieval difficulties (Delis, Kramer, Kaplan, & Ober, 2000). Based on work in our laboratory and other laboratories, we predict that encoding, not retention, is the bottleneck to learning in young adults with DLD. In the next section, we summarize the previous work.

Previous Evidence of Encoding Deficits

In two studies from our laboratory, we compared the ability of young adults with DLD and their unaffected age-mates to encode and retain novel words and their meanings. In McGregor et al. (2013), the DLD group performed lower on recall of new forms and meanings immediately after training, suggesting an encoding problem. There was a significant and moderate correlation between encoding performance and scores on a measure of extant receptive vocabulary, the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007), as would be expected given that encoding is the initial stage of learning vocabulary words. Over the course of the week after training, the performance gap between the DLD and unaffected

groups remained steady for the recall of word meaning, but it increased for the recall of word forms. This result could indicate a retention problem, but the experiment was not designed to test that hypothesis because additional encoding opportunities occurred during the retention interval.

In McGregor et al. (2017), a new sample of young adults completed training on novel words. Again, those with DLD recognized and recalled words less accurately than their peers immediately after training, suggesting an encoding problem. When the word recall task was repeated 1 week later with no intervening opportunities for additional encoding, the size of the performance gap between the DLD and unaffected groups remained the same, suggesting intact retention. Word recognition performance after the 1-week interval also suggested good retention. Specifically, in a visual world paradigm where lexical cohorts compete for activation, newly trained words (e.g., *backθs*) slowed identification of familiar English cohorts (e.g., *bucket*) in both DLD and unaffected groups. We concluded that both groups demonstrated retention and engagement of the words they had encoded.

Other evidence that DLD is characterized by verbal encoding deficits but intact retention comes from studies of children. Bishop and Hsu (2015) compared the learning performance of 7- to 11-year-olds with DLD, their unaffected age-mates, and unaffected children who were younger and had similar grammatical skills as the DLD group. In one condition, the children were trained to associate abstract visual patterns with complex nonverbal sounds. In the other condition, they were trained to associate unfamiliar English words with pictures of the animals they named. Training in both conditions took place in three blocks administered in each of 4 training days. The children with DLD had no difficulty learning the nonverbal material. They did evince problems with word learning, and these problems manifested as lower accuracy of identification compared to age-mates during the first block. They did not lose ground from session to session, suggesting intact retention. Moreover, they did not differ from their peers in the rate of learning across sessions. The authors conclude that long-term declarative memory is intact, but initial encoding of novel phonological strings is impaired.

Nichols et al. (2004) provided additional evidence from a task that involves encoding new configurations of familiar phonological strings, rather than novel strings. Children with DLD, who were 6–14 years old, completed the California Verbal Learning Test–Children's Version (CVLT-C; Delis, Kramer, Kaplan, & Ober, 1994). The CVLT-C begins with the examiner reading List A, a set of 15 common words from the categories fruit, clothing, and toys in a fixed but random order. The child immediately attempts to name the words on the list. This procedure is repeated for five total administrations. These are considered the encoding trials. Then the examiner presents List B, and the child attempts to name these words. After recalling List B, the child must again attempt to recall List A, first uncued and then cued by semantic category. These are the short-delay retention trials. After a 20-min delay, free recall

and cued recall of List A are again elicited as long-delay retention trials. Finally, recognition memory is tested. The child hears 45 words, some from List A, some from List B, some nonlisted items that are similar in sound or meaning to the listed words, and some unrelated words. Poor recall during encoding and retention trials with better performance on the recognition task is considered a sign of retrieval problems. The children with DLD presented with a profile commensurate with an encoding deficit. Specifically, they were poorer than unaffected peers on all encoding trials except the first. There were strengths as well. The two groups did not differ in their use of semantic clustering or in primacy and recency effects on recall. The children with DLD were able to retain and retrieve the information that had been encoded.

Current Approach

DLD in Young Adulthood

DLD can present lifelong challenges for those affected (Clegg, Hollis, Mawhood, & Rutter, 2005). Given a conservative prevalence estimate of 5% (Nippold & Schwarz, 2002), over 12 million adults in the United States are affected by DLD. The current study is one of a series on word learning and memory among young adults with DLD (Becker & McGregor, 2016; Hall, McGregor, & Oleson, 2017; McGregor et al., 2013, 2017; Sheng et al., 2015).

We are particularly interested in adults enrolled in postsecondary education. According to self-report, nearly 6% of university students in the United States have a specific learning disability—an umbrella term that encompasses DLD as it is applied in U.S. postsecondary contexts (McGregor et al., 2016) and prevalence is higher on community college campuses than university or 4-year-college campuses (Newman et al., 2011; Sanford et al., 2011). New information on postsecondary students with DLD holds clinical importance for these citizens who wish to pursue higher education but who, without adequate support, are also at a higher risk of failure than the general college population (Newman et al., 2011). Moreover, there is scientific advantage to limiting our study to comparisons of postsecondary students with DLD to those without: We minimize potential confounds between language status and educational opportunities/socioeconomic status. When studying vocabulary, these are serious confounds (Hart & Risley, 1995).

There are no agreed-upon methods for identifying DLD in adulthood. Reliance upon a positive history of DLD or related language learning disabilities alone is not satisfactory because of underidentification (Tomblin et al., 1997). Poll, Betz, and Miller (2010) examined the sensitivity and specificity of three behavioral tasks for identifying DLD in young adults enrolled in postsecondary vocational schools—nonword repetition, sentence recall, and grammaticality judgment. Of the three, sentence recall was best. Fidler, Plante, and Vance (2011) examined the sensitivity and specificity of nine behavioral tasks for identifying DLD in three groups of adults known to be at high risk: university students receiving accommodations for learning disabilities, university students who reported a history of

speech-language services during childhood, and parents of children with diagnosed DLD. They determined whether the measures identified DLD at expected rates of impairment in these groups and in matched control groups. The measures that maximized sensitivity and specificity were the Modified Token Test—a measure of grammatical understanding and verbal working memory (all at-risk groups), a spelling test (postsecondary groups), and a word defining test (learning disabled and parent groups; Morice & McNicol, 1985). In the earliest work on identification of DLD in adults, Tomblin, Freese, and Records (1992) also found the Modified Token Test and a spelling test to be useful. Following this evidence base, we identified our participants on the basis of the weighted scoring of the Modified Token Test and a spelling test as recommended in Fidler et al. (2011). We also administered a sentence recall test, but this was used descriptively, not for the purposes of identification.

Hypotheses and Predictions

In this project, we took a two-pronged approach to testing the hypothesis that encoding, not retention, impedes word learning in young adults with DLD. In Study 1, we present descriptive data collected via a standardized test of verbal learning and memory. We predicted that adults with DLD would perform more poorly than unaffected age-mates on aspects of the test that reflect initial encoding of information but equally well on aspects that reflect retention relative to the amount of information that they did manage to encode. In Study 2, we present experimental data collected in a word-learning task that included varied supports for the encoding of unfamiliar words and then assessed both shorter- and longer-term retention. We predicted that the young adults with DLD would “close the gap” between their performance and that of their unaffected age-mates when provided support for encoding.

Study 1: Is Encoding the Problem?

Method

Participants

There were 59 participants with DLD and 60 with normal development (ND). Forty-one participants with DLD and 39 with ND also participated in McGregor et al. (2016). Their ages ranged from 18 to 24 years, and all were students in postsecondary institutions in the Midwest of the United States. Participants attended a variety of postsecondary institutions, which in our sample were university (DLD 40, ND 46), 4-year college (DLD 3, ND 0), and community college (DLD 16, ND 14). Demographic information appears in Supplemental Material S1.

To confirm self-reported diagnoses of DLD, we used two measures documented to maximize sensitivity and specificity of DLD identification among young adults with a history of speech-language services: a 15-word spelling test (Fidler et al., 2011) for assessing knowledge of irregularly spelled words and the Modified Token Test for testing syntactic comprehension. The measures were weighted

in accord with the procedure used by Fidler et al. (2011). We had originally enrolled 64 people with previous DLD diagnoses, but we excluded five who did not meet the score threshold to be categorized as having DLD on the Fidler et al. (2011) measures.

In addition to the diagnostic language measures, we administered the nonverbal matrices of the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2004), requiring a standard score of 85 or better by all participants for enrollment in the study. Participants were also required to pass a pure-tone audiometric screening at 0.5, 1, 2, and 4 kHz at 25 dB bilaterally. We also administered a number of probes and standardized tests to assess vocabulary, memory, and additional language skills (Supplemental Material S1). Although their performance was consistently significantly lower than that of the ND group, the DLD group scored within 1 *SD* of the normative mean on all of these standardized tests, as might be expected of students undertaking postsecondary studies.

Stimuli

We administered the California Verbal Learning Test–Second Edition, Adult Version (CVLT-II; Delis et al., 2000), a standardized, norm-referenced test designed to measure the amount of verbal information one can remember and the processes involved in remembering (or forgetting). According to the manual (Delis et al., 2000), the participants in the normative sample were 1,087 16- to 89-year-olds who were representative of the demographics of the U.S. population. There are separate norms for men and women. The test has high internal consistency (split-half correlation $r = .94$). Test–retest reliability varies with the score under consideration from a low of $r = .27$ for total learning slope to a high of $r = .88$ for the accuracy of long-delay free recall. The high correlations ($r = .63$ to $r = .86$) between the CVLT-II scores and the California Verbal Learning Test–First Edition scores speak to its validity (Delis, Kramer, Kaplan, & Ober, 1987). The CVLT-II is significantly correlated with verbal IQ as measured by the vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999).

Procedure

The procedures here and in Study 2 met the ethical standards of the University of Iowa Internal Review Board and the Helsinki Declaration of 1975, as revised in 1983. The third author, a certified speech-language pathologist with extensive assessment experience, administered the CVLT-II according to the standard procedures specified in the test manual. The procedure is like that of the CVLT-C administered in Nichols et al. (2004), except that there are 16, rather than 15, words per list, and these belong to four different semantic categories: furniture, vegetables, ways of traveling, and animals. The procedure involves reading List A, a list of 16 randomly ordered words, five times to the participant. After each reading, the participant is asked to recall the words. The examiner then reads List B, an interference list of 16 unrelated words, and the participant is asked to recall them. Next, the examiner administers

short-delay free recall and short-delay cued recall tasks of List A. The cuing involves presentation of the semantic category labels. There is then a 20-min delay during which the participant completes a nonverbal task. Afterward, the examiner administers long-delay free recall and long-delay cued recall tasks. Finally, the examiner administers a forced-choice recognition task. The number of hits and false positives on the recognition task is used to compute the Discriminability Index. Given this mix of recognition and recall tasks at various delays with and without cuing, conclusions can be drawn about the relative integrity of various memory processes. The CVLT-II Comprehensive Scoring System (Delis & Fridlund, 2000) software automatically computes the raw and standardized scores.

Results

Comparison to CVLT-II Normative Data

As a preliminary step, we asked how the participants' performance compared to that of the normative sample of the CVLT-II. The CVLT-II norms are expressed as z scores with a mean of zero. In the DLD group, the median performance was -0.5 on immediate recall (Trial 5) and also -0.5 at the short-delay and long-delay recall tests. Thus, most participants in the DLD group were just under the expected mean, and their performance relative to normative expectations did not decline from encoding short-term to long-term retention period. Only a handful of scores were clinically significant: 12 participants earned a score that was 2 *SDs* or more below the mean of the normative sample on immediate recall, as did five on short-delay recall and six on long-delay recall. In the ND group, the median performance was zero at all recall intervals. Thus, most of these participants performed at the expected mean. Two participants scored 2 *SDs* or more below the mean on the immediate recall test, as did two participants at the short-delay recall test.

Patterns of Responses on the CVLT-II

We focused on scores that the test developers classify as revealing the integrity of encoding information into memory and retaining that information over time. Table 1 summarizes the scores we analyzed, the interpretation of those scores as recommended by the authors of the test, and the differences we found between the DLD and ND groups. We present the statistical analyses of these scores in the next sections.

In SAS PROC MIXED, we used a linear regression model with an unstructured correlation matrix to evaluate the relationship between the outcome variable (number of words recalled) and the predictor variables (diagnostic group, delay, and recall condition [cued or uncued]). The unstructured correlation matrix accounts for the within-subject correlation due to multiple responses per subject and allows each of the response conditions to have unique correlations with each other. Diagnostic group was a between-subjects variable; delay and recall conditions were within-subject variables. The recall conditions were measured

Table 1. CVLT-II measures, their interpretation, and differences between diagnostic groups in Study 1.

Measures	Interpretation	Outcome
Measures of encoding		
Correct recall on List A Trial 1	Lower recall suggests poorer phonological short-term memory.	DLD < ND
Correct recall on List A Trials 2–5	Lower recall suggests poorer encoding.	DLD < ND
Rate of learning, Trials 1 and 2	Lower rate of change suggests poorer encoding.	DLD < ND
Semantic clustering of List A recall	Less semantic clustering suggests less effective encoding strategy.	DLD < ND
Serial clustering of List A recall	More serial clustering suggests less effective (stimulus-bound) encoding strategy.	DLD = ND
Primacy of List A recall	Lower recall from primacy region of the list suggests poorer encoding.	DLD = ND
Recency of List A recall	Words recalled primarily from the recency region of the list suggest poor encoding.	DLD = ND
Measures of retention and retrieval		
Short-delay free recall	Steeper drop in recall from Trial 5 to the short delay suggests poorer retention and/or poorer ability to resist retroactive interference.	DLD = ND
Short-delay cued recall	Greater boost in recall after cuing suggests that more words were encoded than could be retrieved.	DLD = ND
Long-delay free recall	Steeper drop in recall from the short delay to the long delay suggests poorer retention.	DLD = ND
Long-delay cued recall	Greater boost in recall after cuing suggests that more words were encoded than could be retrieved.	DLD = ND
Recognition accuracy	Accurate recognition but poor recall indicates retrieval deficits. Poor recognition indicates encoding deficits.	DLD < ND

Note. CVLT-II = California Verbal Learning Test–Second Edition, Adult Version; DLD = developmental language disorder; ND = normal development.

at the short- and long-delay intervals only, as these were the only points that included both cued and uncued retrieval tasks. Therefore, each student participated in nine conditions (Trial 1, Trial 2, Trial 3, Trial 4, Trial 5, Short Uncued, Short Cued, Long Uncued, and Long Cued). Given an unbalanced design, we focused directly on all of the pairwise tests of interest. To be conservative, we reported *p* values adjusted for false discovery rate. A residual analysis yielded no evidence of lack of normality.

The ND group recalled more words than the DLD group at all times and in both recall conditions (see Table 2). Both groups improved with each subsequent time point except the Trial 5 to short-term delay comparison where

performance dropped, likely because of the intervening interference list (see Table 3). For both groups, cued recall was stronger than uncued recall. One disadvantage of the pairwise tests is that interaction effects are not tested directly; however, these can be discerned in Table 2 and Figure 1. The difference between the DLD and ND groups roughly doubles from Trial 1 to Trial 2 (–0.9664 to –1.9987), revealing a slower rate of encoding on the part of the DLD group. The between-groups differences are roughly steady from Trials 2 to 4. The initial pattern reverses; from Trials 4 to 5, the DLD group encodes at a faster rate than their ND peers so that the difference between them nearly halves (–2.0383 to –1.2700). The difference between the

Table 2. The effect of diagnostic group (DLD vs. ND) on recall performance at each time point and in each recall condition.

Time/condition	Estimate	SE	<i>t</i>	<i>p</i>	FDR	95% CI
Trial 1	–0.9664	0.3390	–2.85	.0052	0.0068	[–1.6377, –0.2950]
Trial 2	–1.9989	0.4333	–4.61	< .0001	< 0.0001	[–2.8570, –1.1407]
Trial 3	–1.7997	0.4004	–4.49	< .0001	0.0001	[–2.5927, –1.0068]
Trial 4	–2.0393	0.3999	–5.10	< .0001	< 0.0001	[–2.8312, –1.2474]
Trial 5	–1.2701	0.3987	–3.19	.0019	0.0025	[–2.0597, –0.4804]
S/uncued	–1.8006	0.4487	–4.01	.0001	0.0003	[–2.6891, –0.9120]
L/uncued	–1.7257	0.4696	–3.67	.0004	0.0007	[–2.6558, –0.7956]
S/cued	–1.7393	0.4314	–4.03	< .0001	0.0003	[–2.5936, –0.8849]
L/cued	–1.6667	0.4286	–3.89	.0002	0.0003	[–2.5155, –0.8179]

Note. The estimate indicates the difference in number of words recalled by the two groups at each time or condition. Degrees of freedom for each comparison = 117. DLD = developmental language disorder; ND = normal development; FDR = false discovery rate; S = short delay; L = long delay.

Table 3. The effect of time and recall condition on recall performance within each diagnostic group.

Dx	Time/condition	Time/condition	Estimate	SE	t	p	FDR	95% CI
DLD	1	2	-2.0508	0.252	-8.13	< .0001	< 0.0001	[-2.5501, -1.5512]
	2	3	-1.9492	0.205	-9.51	< .0001	< 0.0001	[-2.3551, -1.5432]
	3	4	-0.6271	0.222	-2.83	.0055	0.0067	[-1.0658, -0.1884]
	4	5	-1.1525	0.193	-5.96	< .0001	< 0.0001	[-1.5353, -0.7698]
	5	S/uncued	1.8305	0.219	8.34	< .0001	< 0.0001	[-2.2654, -1.3956]
	S/uncued	L/uncued	-0.4915	0.201	2.44	.0160	0.0182	[0.0931, 0.8899]
	S/cued	L/cued	-0.3559	0.133	2.68	.0084	0.0099	[0.0930, 0.6189]
	S/cued	S/uncued	0.6780	0.191	3.55	.0005	0.0009	[0.3003, 1.0557]
	L/cued	L/uncued	0.5424	0.147	3.68	.0003	0.0006	[0.2508, 0.8339]
	ND	1	2	-3.0833	0.250	-12.32	< .0001	< 0.0001
2		3	-1.7500	0.203	-8.61	< .0001	< 0.0001	[-2.1526, -1.3474]
3		4	-0.8667	0.220	-3.95	.0001	0.0003	[-1.3017, -0.4316]
4		5	-0.3833	0.195	-2.00	.0478	0.0478	[-0.7629, -0.0038]
5		S/uncued	1.3000	0.218	5.97	< .0001	< 0.0001	[-1.7313, -0.8687]
S/uncued		L/uncued	-0.4167	0.200	2.09	.0389	0.0403	[0.0216, 0.8117]
S/cued		L/cued	-0.2833	0.132	2.15	.0335	0.0363	[0.0226, 0.5441]
S/cued		S/uncued	0.6167	0.189	3.26	.0015	0.0021	[0.2421, 0.9912]
L/cued		L/uncued	0.4833	0.146	3.31	.0012	0.0019	[0.1942, 0.7725]

Note. The estimate indicates the difference in the number of words recalled at the two time points or cuing conditions being compared. Degrees of freedom for each comparison = 117. Dx = diagnostic group; FDR = false discovery rate; DLD = developmental language disorder; S = short delay; L = long delay; ND = normal development.

groups remains steady between the short-term and long-term delays (-1.7 or -1.8, depending on whether the recall was cued or uncued). In summary, the DLD group performed poorer than the ND group on all encoding trials, although the size of the performance gap varied across trials. The performance gap was similar in size at the short and long delays, that is, the participants with DLD did not demonstrate declining performance over the retention interval.

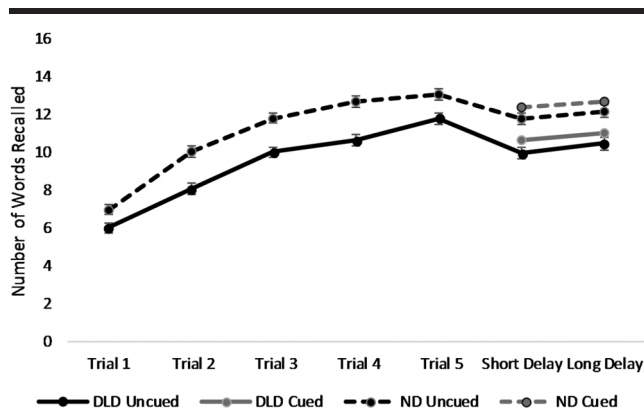
The order in which participants recall items reflects the organization of the items during encoding. Participants can organize items within semantic categories (semantic clustering) or by the order that they are presented (serial clustering). We compared the encoding strategies of the two diagnostic groups. The dependent variables were the

semantic cluster score and the serial cluster score. Note that these scores are chance-adjusted for the number of items produced at each time point. For example, the semantic cluster score is the number of clusters minus the number of clusters expected based on the total words recalled in the trial. The semantic cluster scores were not normally distributed, so we employed a nonparametric test. The DLD group ($M = 0.61$, $SE = 0.20$, 95% CI [0.22, 1.01]) used less semantic clustering than the ND group ($M = 1.37$, $SE = 0.20$, 95% CI [0.98, 1.76]), Mann-Whitney $U = 1,240.5$, $z = -2.81$, $p = .005$. This was a small effect, $abs(r) = .26$. There were no group differences in serial clustering, $t(117) = 0.46$, $p = .65$ (DLD $M = 0.93$, $SE = 0.13$, 95% CI [0.66, 1.17]; ND $M = 1.0$, $SE = 0.12$, 95% CI [0.76, 1.25]).

We also compared the two groups on the average percentage of items recalled from the beginning of the list, the first four items (primacy), and the end of the list, the last four items (recency) across Trials 1-5, by conducting a mixed analysis of variance with group (DLD, ND) as the between-subjects variable and list position (beginning, end) as the within-subject variable. There was a main effect for list position, $F(1, 117) = 15.94$, $p < .0001$, $\eta_p^2 = .12$ (medium effect), but no significant main effect for group, $F(1, 117) = 1.65$, $p = .20$, and no significant interaction, $F(1, 117) < 1$. Participants recalled a higher percentage of words from the beginning of the list ($M = 30.20$, $SE = 0.48$, 95% CI [29.25, 31.16]) than the end of the list ($M = 26.96$, $SE = 0.54$, 95% CI [25.89, 28.02]).

Finally, we compared the recognition accuracy across the two groups by examining the number of hits, the number of false positives, and recognition discriminability—which takes into account the individual's hit rate relative to his or her false positive rate. None of these scores were normally distributed, so we employed nonparametric tests. All three

Figure 1. The word recall of the developmental language disorder (DLD) and normal development (ND) groups after each encoding trial and at short and long delays. Cued recall at short and long delays is also depicted. Standard error bars are included.



scores differed between groups: recognition hits, Mann–Whitney $U = 1,041$, $z = -3.87$, $p = .0001$, $\text{abs}(r) = .35$ (medium effect); false positives, Mann–Whitney $U = 1,164$, $z = 3.22$, $p = .001$, $\text{abs}(r) = .30$ (medium effect); and recognition discriminability, Mann–Whitney $U = 942.5$, $z = -4.40$, $p = .00001$, $\text{abs}(r) = .40$ (medium effect). Individuals with ND scored higher on number of hits ($M = 15.17$, $SE = 0.17$, 95% CI [14.84, 15.50]) and recognition discriminability ($M = 3.52$, $SE = 0.09$, 95% CI [3.35, 3.69]) but lower on false positives ($M = 0.92$, $SE = 0.33$, 95% CI [0.27, 1.56]) than individuals with DLD (hits $M = 14.20$, $SE = 0.17$, 95% CI [13.87, 14.54]; recognition discriminability $M = 2.93$, $SE = 0.09$, 95% CI [2.76, 3.10]; false positive $M = 2.27$, $SE = 0.33$, 95% CI [1.62, 2.92]). Together, these results reveal poorer recognition on the part of the DLD group, consistent with an encoding deficit.

Relationship Between the CVLT-II and PPVT-4 Scores

Because we are interested in encoding as the set of initial processes involved in word learning, we ran a correlation between the PPVT-4 vocabulary scores, a proxy for the end state of that learning, and the encoding indices that differentiated DLD and ND groups. For the DLD group, PPVT-4 raw scores were positively predicted by the amount of correct recall during the encoding trials, $r = .29$, $p = .03$, and the extent of semantic clustering, $r = .34$, $p = .009$; both were medium effect sizes. These relationships were not significant for the ND group (PPVT-4 by correct recall, $r = .22$, $p = .09$; PPVT-4 by semantic clustering, $r = .14$, $p = .30$).

Discussion

Interpretation of Study 1 Data

The evidence from Study 1 supported the prediction that encoding poses a bottleneck to learning for young adults with DLD. When compared to their unaffected peers, the participants with DLD evinced an encoding deficit by recalling fewer words over learning trials and using less semantic clustering—an effective encoding strategy. That said, they did not differ from their peers in all aspects of encoding. They did not recall words primarily from the end of the list, a pattern that would have suggested the use of short-term memory rather than true encoding into memory, nor did they depend heavily on serial clustering—an ineffective encoding strategy.

Still, the weight of the evidence points to encoding rather than retention as the crux of the problem. The performance gap between the DLD and ND groups remained steady from the shorter to the longer retention interval. The evidence also suggests that retrieval mechanisms are not at fault. Performance after cuing should improve if retrieval is a problem, but response to cuing was similar in the DLD and ND groups. Finally, if individuals with DLD struggled primarily with retrieval, we would expect good recognition accuracy, because recognition tasks minimize demands on retrieval. Instead, we found that the recognition task differentiated the DLD and ND groups, with poorer recognition on the part of those with DLD.

Comparison to Children’s Performance on the CVLT

The profile of these adults with DLD—poor encoding with good retention and retrieval—is remarkably similar to the profile of the children with DLD studied by Nichols et al. (2004). There were two differences between their study and ours. They found no difference between participants with and without DLD on Trial 1. We did find a difference there, albeit with a smaller effect size than on subsequent trials. We found less use of semantic clustering on the part of the DLD group, but they found similar semantic clustering in the two diagnostic groups. Recall that the adult and child versions of the CVLT use different semantic categories; perhaps this is relevant to the inconsistency.

Motivation for Study 2

Given that young adults with DLD struggled with encoding and that encoding performance accounted for significant variance in their receptive vocabulary scores, we pursued a more direct test of the effect of encoding on new vocabulary learning in Study 2. Specifically, we manipulated the encoding experience to see if we could close the learning gap between the diagnostic groups. There is consistent evidence that encoding new information in tasks that demand retrieval practice is more effective than encoding in response to equivalent but passive exposure to the information (for a review, see Karpicke & Grimaldi, 2012). For example, college students who engaged in repeated testing of Swahili–English word pairs that required them to retrieve the words repeatedly recalled more words a week later than students who engaged in repeated study trials (Karpicke & Roediger, 2008). Thus, in Study 2 we investigated whether retrieval practice via repeated tests boosts encoding for college students with DLD.

Study 2 Approach

When determining how to design the tests to be used for training, we grappled with two concerns. First, we knew from Study 1 that participants with DLD would likely answer the test questions (i.e., recall the words) less accurately than those with ND. Evidence from adults with ND confirms the detrimental effect of incorrect retrieval on learning (Butler, Karpicke, & Roediger, 2008; Fazio, Huelser, Johnson, & Marsh, 2010; Metcalfe & Kornell, 2007; Pashler, Cepeda, Wixted, & Roherer, 2005; Pashler, Zarow, & Triplett, 2003). However, the effect is ameliorated by providing feedback during the test (Butler & Roediger, 2008; Metcalfe, Kornell, & Finn, 2009; Pashler et al., 2003, 2005). Therefore, we designed a test with feedback.

The second concern was how to elicit the optimal level of effort for promoting encoding. Among college students with ND, free recall enhances encoding and retention more than cued recall (Carpenter, 2009; Carpenter & DeLosh, 2006). This effect is not a matter of practicing the type of question that will appear on the test, as free recall during training enhances retention even when the later test involves cued recall (Carpenter & DeLosh, 2006). Instead, it seems that the higher level of effort required by the free recall questions is optimal for encoding (Bjork, 1994). However,

free recall is not the most effective study strategy for all learners. Grade-school children benefited more from answering specific questions about science texts than from free recall of science facts (Karpicke, Blunt, Smith, & Karpicke, 2014). Not knowing the optimal effort level for the DLD group, we designed two test conditions. Both included feedback, as described above, but one did no cued naming of new referents whereas the other did cued naming by presenting the first syllable of each two-syllable name.

In Study 2, participants completed the CVLT-II and completed a novel word-learning task. In the word-learning task, we taught adults novel word-referent pairs in three conditions: free recall, cued recall, and passive study. To minimize the effect of incorrect answers on encoding, we provided feedback in the free and cued recall conditions. To control for individual differences, we used a within-subject design. Each participant learned three sets of word-referent pairs under each of three conditions. We measured participants' memory for the words by their naming performance on the final encoding trial (for the free and cued recall conditions only) and their naming performance 24 hr later. Also, we used a visual world paradigm to measure their memory and lexical engagement of the words 1 week after the encoding trials. The advantage of the visual world paradigm is that it is a sensitive measure of lexical memory and lexical competition that does not require overt free recall, a feat of memory likely to be fragile after 1 week in the absence of additional exposures to the words (McGregor et al., 2017).

Study 2: Does Learning via Testing Improve Encoding?

Method

Participants

Forty-seven additional college students (ages 18–25 years) from the Midwest of the United States participated. Twenty-three participants qualified as having DLD, and 24 qualified as having ND. Participants were similar in type of postsecondary institution, which in our sample was university (DLD 17, ND 18), 4-year college (DLD 4, ND 5), and community college (DLD 2, ND 1).

The participants met the same inclusionary and exclusionary criteria, measured with the same tools, described in Study 1. Demographic information as well as the results of standardized tests administered to describe their language and memory skills appear in Supplemental Material S1. The DLD group scored within 1 *SD* of the normative mean on all of these nondiagnostic standardized tests, although their performance was significantly lower than that of the ND group on all tests other than the Kaufman Brief Intelligence Test.

Stimuli

Training. Three sets (A/B/C) of nine novel word stimuli were constructed (Supplemental Material S2). We derived each novel word from a real word by modifying either one or two final consonants. To ensure familiarity,

we selected real English words with age-of-acquisition ratings below 10 years (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). The English words were disyllabic nouns with primary stress on the first syllable. The prosodic structure of the real words was CV.CVC (e.g., “faucet”), CVC.CVC (e.g., “garlic”), or CCV.CVC (e.g., “stomach”). The real words also satisfied the following lexical criteria: log lexical frequency (Brysbaert & New, 2009) between 1 and 4 (mean = 2.45, base = 10), concreteness rating (Coltheart, 1981) between 400 and 700 (scale 100–700, mean = 574.8, *SD* = 51.8), imageability rating (Coltheart, 1981) between 400 and 700 (scale 100–700; mean = 574.9, *SD* = 38.1), and familiarity rating (Nusbaum, Pisoni, & Davis, 1984) between 4 and 7 (scale 1–7; mean = 6.9, *SD* = 0.25).

To equate learnability, the novel words were proportionately distributed across the three sets so that (a) no word onset segment was represented more than twice per set, (b) phonological feature distribution (i.e., place, manner, and voicing) of onset segments was proportionally balanced across the three sets, and (c) phonological feature distribution of onset segments was similarly balanced across each of the three prosodic shapes (CV/CVC/CCVC).

We paired each of the 27 novel words with a visual referent, selected from digital images of unusual objects that have no commonly known English name. Objects were divided into three sets, each containing three plants, three animals, and three inanimate objects. The stimulus sets were counterbalanced to training conditions across participants.

Visual world paradigm. For the visual world paradigm, we needed stimuli to populate four trial types: real English word with cohort (e.g., letter vs. lettuce), real English word with noncohort (e.g., letter vs. bacon), trained novel word versus real English cohort (e.g., /fəʊsɪb/ vs. faucet), and trained novel word versus real English noncohort (e.g., [fəʊsɪb] vs. desert). Therefore, in addition to the 27 novel words and their English base words, we selected 27 more English word pairs with the same phonological and lexical characteristics described above. A summary of stimulus pairings for the visual world paradigm task appears in Supplemental Material S2.

Recording. A female native speaker of Standard American English recorded audio stimuli and task instructions in a sound-treated booth using a Larson Davis 2560 0.5-in. random incidence microphone (Larson Davis, Inc.). The recorded signals were routed to the computer through a Larson Davis PRM902 preamplifier (Larson Davis, Inc.), a LISTEN SoundConnect microphone power supplier (Listen, Inc.), and a MOTU Ultralite-mk3 Hybrid sound interface (MOTU). The stimuli were then digitized at a 44.1-kHz sampling rate and 16-bit resolution using the sound-editing software Adobe Audition (Version 1.0; Adobe Audition, 2003).

Procedure

Participants attended four sessions of 1–2 hr in length. During the first session, we obtained informed consent. Participants completed a demographic questionnaire and a series of standardized tests, including the CVLT-II. During the second session, the participants completed the

training tasks. Session 3, 24 hr after training, involved a picture-naming task, and Session 4, 1 week after training, involved a word recognition task in the form of a visual world paradigm.

Training. There were three training conditions: learning via uncued testing, learning via cued testing, and learning via passive study (no test). Participants attempted to learn nine words in each of these conditions; the order of conditions was counterbalanced across participants. Participants faced a Dell Vostro laptop (Dell, Inc.) and wore closed circumaural headphones. A Snowball USB microphone (Blue Microphones) was placed approximately 18 in. away from the speaker's mouth to record naming attempts in the cued and uncued test conditions.

Each condition began with the nine word–picture stimuli presented via E-Prime 2.0 in random order. Participants were instructed to watch and listen carefully, as the experimenters would be observing how many of the words they learn. After seeing and hearing each stimulus pair once, participants engaged in a 15-s distractor task solving simple math problems. Afterward, the training continued as follows.

In the uncued test condition, participants were asked to name each picture upon presentation. They heard a chime if they answered correctly. When participants answered incorrectly (or if they had no answer at all), the experimenter would play the audio sample of the entire target word once. Participants cycled through the 9-item set 3 times so that they had four exposures per word–picture pair (once via passive exposure and thrice via correct uncued naming or the audio sample feedback if incorrect). The cued test condition was identical to the uncued test condition, except that the first syllable of the target word was aurally presented along with its corresponding picture. Participants were asked to name each picture using the entire target word after hearing this cue. They again received feedback in the form of a chime or the correct production, and they again cycled through the set three times for a total of four exposures (one passive and three via correct cued naming or the audio sample feedback if incorrect). In the third task—the no-test condition—participants were instructed to silently watch and listen as the stimuli from the relevant set were presented both visually and aurally. They did this three times after the first exposure for a total of four passive exposures.

Picture naming. Approximately 24 hr following the exposure and training conditions, participants were asked to name all 27 picture stimuli as they were presented in random order. There was no cue or feedback.

Visual world paradigm. Participants returned 1 week after training for a word recognition task completed in a visual world paradigm. Each trial began with the simultaneous presentation of two pictures, one each in the upper corners of a computer screen. The participants were instructed to move a cursor up a “runway” (vertical bar) located in the center of the screen. Once the cursor reached the top of the runway, they heard the name of one of the two pictures, and they moved the cursor to click on the named picture. Participants were told that the task was timed and they should, therefore, answer as quickly as possible.

Transcription. The fourth author, a linguist, transcribed all naming responses. A research assistant independently transcribed a random sample of 10 participants' responses. Segment-by-segment agreement between the two transcribers was 93% for the naming responses collected on the training day and 94% for the naming responses collected at the 1-day retention interval.

Results

CVLT-II

As a preliminary step, we asked whether this group of participants presented similar standard scores on the CVLT-II as those in Study 1. In the DLD group, the median performance on immediate recall was $z = -1$; the median performance on short-delay recall was $z = -0.5$, and the median performance on long-delay recall was also $z = -0.5$; therefore, at the group level, there was improvement rather than decline of performance over retention intervals relative to normative expectations. Only a handful of scores were clinically significant: Three participants earned a score that was 2 *SDs* or more below the mean of the normative sample on immediate recall, as did one on short-delay recall and two on long-delay recall. In the ND group, the median performance at all recall intervals was zero. Two of the participants earned a score that fell 2 *SDs* below the mean for immediate recall. These profiles are similar to those in Study 1 in that most participants in the DLD group performed just below the expected mean whereas those in the ND group performed at the mean.

Naming

Twenty-four hours after novel word training, the participants with ND accurately named an average of 5.4 of the 27 words correctly, whereas those in the DLD group named only 3.51. At first, this might seem surprisingly low given that, at the longest delay, the ND participants in Study 1 named an average of 12.18 words and the DLD participants named 10.48 (see Figure 1). However, these differences are less surprising upon considering that Study 1 involved real words, not novel words; 16 words, not 27; and a retention interval of 20 min, not 24 hr. Poor naming performance after learning multiple new words is a common finding among children (Booth, McGregor, & Rohlfing, 2008; Gray, 2003; Horst & Samuelson, 2008; Munro, Baker, McGregor, Docking, & Arciuli, 2012) and adults (Storkel, Armbrüster, & Hogan, 2006).

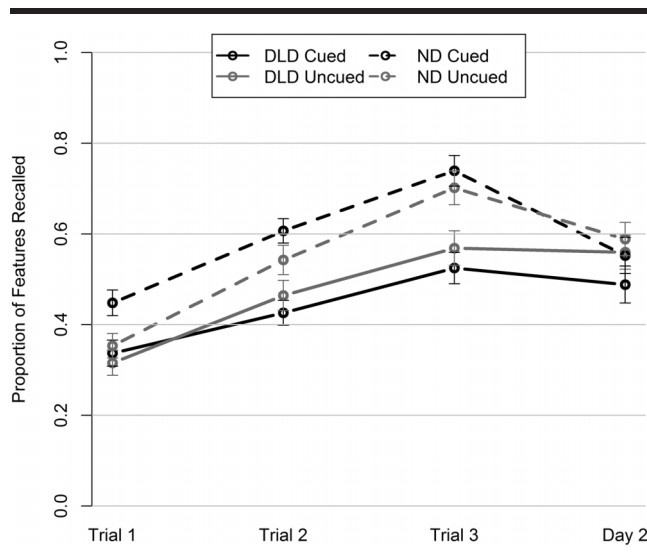
Given the near-floor level naming performance of both the ND and DLD groups in Study 2, we turned away from a binary score to a more nuanced analysis of naming performance. We compared productions to targets feature by feature to capture how well the participants recalled the place, manner, and voicing of each consonant segment and the height, backness, and tenseness of each vowel. The dependent variable was the proportion of the features that were produced correctly. That way, if the participant said “mɛləd” for the target /mɛləg/, for example, he or she would receive some credit for remembering most of the

word form. Note that for cued naming during the training, this dependent variable was derived from the second syllable productions only because the participants heard the first syllable as a cue. In all other cases, the dependent variable was derived from analysis of the whole word production.

In SAS PROC MIXED, we used a linear regression model with an unstructured correlation matrix to evaluate the relationship between the outcome variable (number of features recalled from the second syllable of each target word) and the predictor variables (diagnostic group, delay, and encoding condition). Diagnostic group was a between-subjects variable; time and training condition were within-subject variables. Each student participated in nine conditions (Cued Trial 1, Cued Trial 2, Cued Trial 3, Cued Day 2, Uncued Trial 1, Uncued Trial 2, Uncued Trial 3, Uncued Day 2, No Test Day 2). Due to this unbalanced design, we did not conduct a formal main effect test or test of interaction. Instead, we focused directly on all of the pairwise tests of interest, which we used to discuss the interaction effect. We used the false discover rate to adjust the 37 pairwise tests of interest to control for Type I error. A residual analysis yielded no evidence of lack of normality.

Depending on time and condition, ND participants averaged 0.35–0.74 proportion of features recalled in the trained words, whereas the DLD participants averaged 0.32–0.57 (see Figure 2). By the end of the encoding trials in the cued and uncued test conditions on Day 1, the DLD group could recall fewer features than the ND groups. On the second day, the two groups did not differ from each other in the recall of words that they had encoded in the test conditions, but they did differ in the recall of words

Figure 2. Mean proportion of features correctly recalled by the developmental language disorder (DLD) and normal development (ND) groups plotted by time and encoding condition. Error bars depict standard error.



that they had encoded in the no-test condition, $p = .04$, false discover rate = .06, CI [-0.2, -0.01] (Table 4; Figure 2). Thus, the prediction that encoding support would close the long-term gap between the groups held.

During the period of encoding itself, there were different rates of change in the two diagnostic groups. We found an approximate doubling of the between-groups effect size from Trial 1 to Trial 2 (from $-.038$ to $-.078$ in the uncued test condition and from $-.111$ to $-.181$ in the cued test condition). In other words, the DLD group began to lag farther behind the ND group. This difference in rate fit the same pattern we found in Study 1, suggesting that building on the first memory trace was particularly difficult for participants with DLD.

The within-subject comparisons appear in Table 5. First, consider change over trials in the two training conditions involving testing. The DLD group recalled more features from one encoding trial to the next on Day 1, and the amount of recall did not differ in the uncued and cued test conditions. The ND group also recalled more features from one trial to the next, and their recall on Trial 1 was higher in the cued than in the uncued condition.

Next, consider the change in performance from Day 1 to Day 2. From the final encoding trial on Day 1 to the naming test on Day 2, the DLD group's performance remained stable in both uncued and cued conditions (Table 5; Figure 2). The ND group's performance dropped in both conditions. On Day 2, both groups performed better in the test conditions (whether cued or uncued) than in the no-test condition; performance in the two test conditions did not differ (Table 5; Figure 3).

Visual World Paradigm

To examine lexical engagement after a week-long retention interval, we analyzed the results of the visual world paradigm. The dependent variable was the median reaction time in milliseconds for correct identifications of the referents being named by the English words. Accuracy was high for the DLD group ($M = .989$, $SE = .002$, 95% CI [.984, .993]) and the ND group ($M = .990$, $SE = .002$, 95% CI [.986, .994]).

To explore the variables that affected reaction time, we applied an analysis of variance with diagnostic group (DLD, ND) as a between-subjects variable and training condition (uncued test, cued test, no test) and cohort status of foil (cohort, noncohort) as within-subject variables. There were no violations of homogeneity of variances or normality.

There was no main effect of diagnostic group, $F(1, 45) < 1$, or training condition, $F(2, 90) = 1.18$, $p = .31$. There was a main effect for cohort, $F(1, 45) = 82.76$, $p < .0001$, $\eta_p^2 = .65$, that was qualified by a Training Condition \times Cohort interaction, $F(2, 90) = 8.93$, $p = .0003$, $\eta_p^2 = .17$. According to a Bonferroni post hoc test, word-referent recognition took longer in the presence of a lexical cohort ($M = 1,113$, $SE = 13.76$, 95% CI [1,085, 1,141]) than a noncohort ($M = 1,042$, $SE = 16.75$, 95% CI [1,008, 1,076]), and this was true in the uncued, cued, and no-test conditions, $ps < .0001$. The interaction occurred because the participants

Table 4. The effect of diagnostic group (developmental language disorder vs. normal development) on recall performance at each time point and in each encoding condition.

Encoding condition	Time (Day.Trial)	Estimate	SE	t	p	FDR	95% CI
Uncued	1.1	-0.0377	0.0387	-0.98	.3347	0.3885	[-0.1155, 0.0402]
Uncued	1.2	-0.0782	0.0463	-1.69	.0985	0.1408	[-0.1715, -0.0152]
Uncued	1.3	-0.1330	0.0533	-2.49	.0164	0.0283	[-0.2404, -0.0256]
Uncued	2.1	-0.0292	0.0525	-0.56	.5808	0.5990	[-0.1349, 0.0765]
Cued	1.1	-0.1110	0.0404	-2.75	.0086	0.0158	[-0.1922, -0.0297]
Cued	1.2	-0.1806	0.0383	-4.72	< .0001	0.0001	[-0.2577, -0.1035]
Cued	1.3	-0.2144	0.0483	-4.44	< .0001	0.0002	[-0.3116, -0.1172]
Cued	2.1	-0.0642	0.0571	-1.13	.2665	0.3261	[-0.1792, 0.0508]
No test	2.1	-0.1043	0.0485	-2.15	.0368	0.0609	[-0.2019, -0.0067]

Note. The estimate indicates the difference in the proportion of features recalled by the two groups. Degrees of freedom for each comparison = 45. FDR = false discovery rate.

took longer to recognize English words referents in the presence of a lexical cohort when that cohort was a novel word trained in the uncued test condition ($M = 1,131$, $SE = 15.08$, 95% CI [1,100, 1,161]) or the cued test condition ($M = 1,113$, $SE = 13.73$, 95% CI [1,085, 1,140]) than the no-test condition ($M = 1,096$, $SE = 15.15$, 95% CI [1,065, 1,126]), $ps < .003$. The difference between the uncued and cued test conditions was not significant, $p = .67$. In other words, training via tests resulted in more robust lexical competition 1 week later than passive training. This was true for both groups.

Discussion

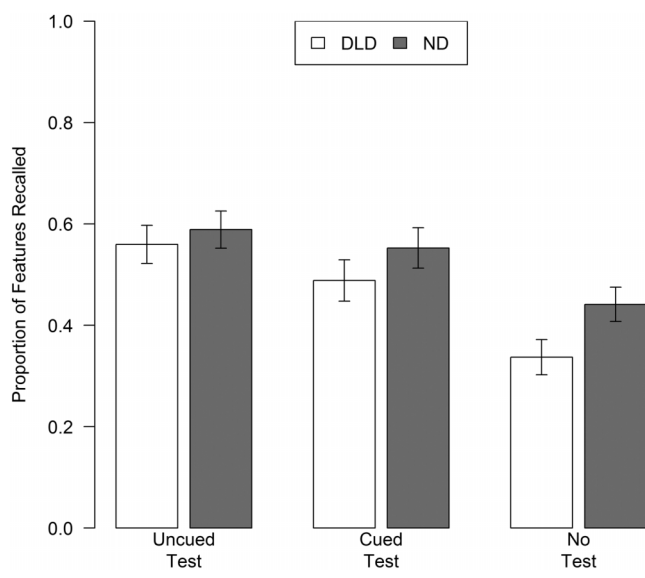
The early stages of word learning, those processes involved in encoding new information, are challenging to young adults with DLD. The later stages of learning, those processes involved in retaining encoded information, are not problematic. We found this to be true of their performance on a standardized test that measured their ability to encode and retain a list of familiar words (Study 1) and on probes of encoding and retention administered after training unfamiliar words (Study 2). In Study 2, the

Table 5. The effect of encoding condition and time on recall performance within each diagnostic group.

Dx	Encoding condition	Time (Day.Trial)	Time (Day.Trial)	Estimate	SE	t	p	FDR	95% CI
DLD	Uncued	1.1	1.2	-0.1487	0.020	-7.29	< .0001	< 0.0001	[-0.1898, 0.1076]
		1.2	1.3	-0.1043	0.021	-5.02	< .0001	< 0.0001	[-0.1462, -0.0625]
		1.3	2.1	-0.0091	0.021	-0.44	.6630	0.6630	[-0.0511, 0.0328]
	Cued	1.1	1.2	-0.0891	0.025	-3.57	.0009	0.0019	[-0.1394, -0.0389]
		1.2	1.3	-0.0987	0.023	-4.27	.0001	0.0002	[-0.1453, -0.0521]
		1.3	2.1	0.0365	0.031	1.17	.2478	0.3261	[-0.0263, 0.0993]
ND	Uncued	1.1	1.2	-0.1139	0.020	-5.54	< .0001	< 0.0001	[-0.1540, -0.0719]
		1.2	1.3	-0.1592	0.020	-7.83	< .0001	< 0.0001	[-0.2001, -0.1182]
		1.3	2.1	0.1129	0.020	5.54	< .0001	< 0.0001	[0.0719, 0.1540]
	Cued	1.1	1.2	-0.1587	0.024	-6.50	< .0001	< 0.0001	[-0.2079, -0.1096]
		1.2	1.3	-0.1325	0.023	-5.85	< .0001	< 0.0001	[-0.1781, -0.0869]
		1.3	2.1	0.1867	0.031	6.11	< .0001	< 0.0001	[0.1252, 0.2482]
DLD	Time (Day.Trial)	Encoding condition	Encoding condition						
	1.1	Cued	Uncued	0.0213	0.032	0.67	.5044	0.5419	[-0.0425, 0.0851]
	1.2	Cued	Uncued	-0.0383	0.034	-1.11	.2713	0.3261	[-0.1075, 0.0309]
	1.3	Cued	Uncued	-0.0439	0.039	-1.11	.2720	0.3261	[-0.1234, 0.0356]
	2.1	Cued	Uncued	-0.0713	0.040	-1.78	.0816	0.1221	[-0.1519, 0.0093]
	2.1	Cued	No test	0.1513	0.034	4.48	< .0001	0.0002	[0.0833, 0.2193]
ND	2.1	Uncued	No test	0.2226	0.037	6.00	< .0001	< 0.0001	[0.1479, 0.2973]
	1.1	Cued	Uncued	0.0946	0.031	3.05	.0038	0.0074	[0.0322, 0.1570]
	1.2	Cued	Uncued	0.0642	0.034	1.91	.0628	0.1029	[-0.0036, 0.1319]
	1.3	Cued	Uncued	0.0375	0.039	0.97	.3372	0.3885	[-0.0404, 0.1154]
	2.1	Cued	Uncued	-0.0363	0.039	-0.93	.3598	0.3926	[-0.1152, 0.0427]
	2.1	Cued	No test	0.1112	0.033	3.37	.0016	0.0032	[0.0447, 0.1778]
2.1	Uncued	No test	0.1475	0.036	4.06	.0002	0.0004	[0.0744, 0.2206]	

Note. The estimate indicates the difference in the proportion of features recalled at the two time points or encoding conditions being compared. Degrees of freedom for each comparison = 45. Dx = diagnostic group; FDR = false discovery rate; DLD = developmental language disorder; ND = normal development.

Figure 3. Naming accuracy by diagnostic group and training condition. Accuracy is expressed as the mean proportion of features correctly recalled after a 24-hr retention interval (Day 2). Error bars depict standard error. DLD = developmental language disorder; ND = normal development.



participants with DLD, like their peers with ND, demonstrated engagement between new and familiar words 1 week after encoding, constituting still more evidence of intact retention.

An important next step will be to determine the specific aspects of encoding that are at fault. In the next section, we review literature and patterns within our data that provide direction.

Processes of Encoding

Working memory is the gateway to encoding words that can later be consolidated and retained in long-term memory. Lum, Ullman, and Conti-Ramsden (2015) conducted a study of the relationship between working memory and long-term declarative memory that took advantage of a naturally occurring difference in the presentation of DLD. They compared two groups of children with DLD, one group with concomitant low working memory and the other with average working memory, to age-mates who had neither DLD nor working memory deficits. They administered a list-learning task much like the one we used in Study 1, and they found that only the children with DLD plus working memory impairments had difficulty encoding the listed words as evinced by poor recall of the list during the encoding phase and poor recognition of the words after a delay. These findings parallel those from our DLD group (undifferentiated by working memory ability). Lum et al. (2015) also found poorer retrieval after a delay, even after controlling for the number of words initially encoded, but they attribute this not to a retention problem but to the demands that retrieval places on working memory.

They concluded that long-term declarative memory per se is intact in people with DLD but that working memory deficits impede encoding of information.

Bishop and Hsu (2015) agree. Because they found children with DLD to be deficient in verbal learning (but not nonverbal learning) and because the deficit relative to unaffected children was manifest only in the earliest stages of learning, they concluded that long-term declarative memory is intact. They attribute the early bottleneck to procedural memory limitations. Specifically, the procedural system supports the learning of the motoric sequences that comprise the phonological form of the word to be learned. The procedural system also supports working memory and word retrieval; therefore, the problem should extend not only to the learning of new words but also to the learning of new configurations of familiar words—the sort of problem we verified in Study 1 and replicated at the beginning of Study 2. There is one difference between the learning patterns demonstrated by the children with DLD in Bishop and Hsu (2015) and those demonstrated by the young adults with DLD in the current studies. Bishop and Hsu found differences between children with DLD and unaffected age-mates on the first learning block, and these persisted in magnitude over learning trials; that is, the rate of encoding in the two groups did not differ. We found differences in encoding rate in both Studies 1 and 2, such that adults with DLD evinced a slower rate of encoding than those with ND from the first to second trial. If one assumes a deficit in procedural memory, this inconsistency between studies is to be expected. Bishop and Hsu measured encoding with a recognition task, whereas we measured with a production task. The recognition task minimizes verbal working memory load; thus, it should be sensitive only to the earliest encoding attempts, attempts that involve fully novel strings. Another way to think about this is that production of newly learned words is more likely to reveal the deficit than recognition of newly learned words, a documented finding (Gray, 2003, 2004).

Additional evidence of the working memory (procedural) deficits of young adults with DLD have been documented previously on three different sorts of tasks: competing language processing tasks, false memory tasks, and running span tasks. Competing language processing tasks (Gaulin & Campbell, 1994) require repetition of sentences or sentence-final words after judging the veracity of intervening sentences. Adults with DLD repeat less accurately than matched age-mates on this high-demand task (Isaki, Spaulding, & Plante, 2008; Poll, Miller, & van Hell, 2016), suggesting a more limited working memory capacity. The shifting of resources between storage (remembering the words or sentences) and processing (judging veracity) could also be at fault. This problem would be evident if not only the repetition but also the truth judgments were affected. That evidence is equivocal. Isaki et al. (2008) found both to be affected, whereas Poll et al. found only repetition accuracy to be affected.

The false memory task (Deese, 1959; Roediger & McDermott, 1995) requires the recall of verbally presented

lists of semantically or phonologically related words. False recall reflects the extraction of the “gist” of the list. For example, “sleep” might be recalled in error if related words like “pillow” and “dream” were on the list. Verbatim recall reflects the capacity of the phonological loop and the application of the subvocal rehearsal process that refreshes and maintains that phonologically encoded material. Compared to unaffected peers, adults with DLD do not differ on false memory errors, but they recall fewer words verbatim, suggesting a deficit in short-term capacity (Sheng et al., 2015). Deficient rehearsal mechanisms could also be at play, but this is unlikely given that the DLD group demonstrate robust primacy effects, a list-initial advantage attributable to rehearsal (Sheng et al., 2015).

The running span task (Cowan et al., 2005) requires memory for digits. Lists of 12–20 digits are presented. After each list, the participant is asked to recall the final five, six, or seven digits in order. Adults with DLD, again, perform more poorly than their peers, and this is thought to reflect their more limited short-term capacity (Poll et al., 2016).

Together, performance on these three tasks demonstrates that, as a group, young adults with DLD have limited working memory capacity. Given the data reported in Studies 1 and 2, we conclude that encoding is the problematic stage of word learning. Although word learning is typically considered to be a declarative task, the problem does not lie within the declarative system per se. Rather, given the evidence from Isaki et al. (2008), Poll et al. (2016), and Sheng et al. (2015), we hypothesize that working memory capacity is the locus of the problem. There is ample evidence that working memory capacity limits verbal learning in children with DLD as well (see summaries in Archibald & Gathercole, 2006; Lum, Conti-Ramsden, Page, & Ullman, 2012; Montgomery, 2003; Montgomery, Majimairaj, & Finney, 2010). Our finding that adults with DLD have particular problems building upon an initial memory trace (i.e., their slow growth from Trial 1 to Trial 2 in both studies) is compatible with this hypothesis. To build successfully would require holding the initial trace in phonological short-term memory while updating it upon hearing the next presentation of the word form.

Clinical Implications

As children with DLDs enter adolescence and adulthood, the academic challenges they face likely change, but they do not go away (Clegg et al., 2005; Elbro, Dalby, & Maarbjerg, 2011; Whitehouse, Watt, Line, & Bishop, 2009). Students with a variety of learning disabilities, including those with DLD, are attending university in record numbers, but once there, they report barriers to successful matriculation (McGregor et al., 2016). Academic success for postsecondary students with DLD will require supports for the initial stages of learning, especially when unfamiliar words must be encoded.

Here one strategy to support encoding was tested, retrieval-based testing. After a day-long retention interval, the DLD group performed more poorly than the ND group on items learned via passive study, but the two groups

performed equally well on items learned in the cued and uncued test conditions. With the right strategies, students with DLD can “close the gap.”

We were worried that the participants with DLD might need the extra support of cues to ensure adequate learning via tests, but the worry was unfounded. The participants with DLD performed equally well in the uncued and cued test conditions. The presence of feedback is likely key here. Chen and Liu (2014) presented children with DLD with a word-learning task that involved repeated testing but no feedback. The result was a markedly slower rate of encoding on the part of the children with DLD.

Retrieval-based testing provides both direct and indirect benefits (Roediger, Putnam, & Smith, 2011). A direct benefit is that the feedback involved in self-testing allows a student to relearn forgotten information. Self-testing might also present indirect benefits such as enhancing attention to the task and motivation for learning. Whether students with DLD experience the same benefits of retrieval-based testing as other students is a question for future research.

Conclusions

Young adults with DLD present with word-learning problems, and encoding—not retention or retrieval—is the culprit. We hypothesize that limited working memory capacity hinders the encoding of information. Integrating self-testing with feedback into study sessions is helpful for learners with DLD; it serves to close the long-term gap between their performance and that of their peers. This finding is relevant for the treatment of DLD.

Acknowledgment

This work was supported by the National Institutes of Health (5R01DC011742), awarded to Karla McGregor, principal investigator.

References

- Adobe Audition.** (2003). (Version 1.0) [Computer software]. San Jose, CA: Adobe Systems. Retrieved from www.adobe.com/products/audition.html
- Archibald, L. M., & Gathercole, S. E.** (2006). Short-term memory and working memory in specific language impairment. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders* (pp. 139–160). New York, NY: Psychology Press.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L.** (1995). *CELEX-2* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Baddeley, A.** (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839.
- Becker, T. C., & McGregor, K. K.** (2016). Learning by listening to lectures is a challenge for college students with developmental language impairment. *Journal of Communication Disorders*, 64, 32–44.
- 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), 1.

- Bjork, R. A.** (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Booth, A. E., McGregor, K. K., & Rohlfing, K. J.** (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning & Development, 4*, 179–202.
- Brysaert, M., & New, B.** (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods, 41*, 977–990.
- Buchbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F.** (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron, 48*(4), 687–697.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L., III.** (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 918.
- Butler, A. C., & Roediger, H. L.** (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616.
- Carpenter, S. K.** (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563.
- Carpenter, S. K., & DeLosh, E. L.** (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268–276.
- Chen, Y., & Liu, H. M.** (2014). Novel-word learning deficits in Mandarin-speaking preschool children with specific language impairments. *Research in Developmental Disabilities, 35*(1), 10–20.
- Clegg, J., Hollis, C., Mawhood, L., & Rutter, M.** (2005). Developmental language disorders—A follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry, 46*(2), 128–149.
- Coltheart, M.** (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.
- Cowan, N.** (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51–57.
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R.** (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42–100.
- Deese, J.** (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.
- Delis, D. C., & Fridlund, A. J.** (2000). *CVLT-II Comprehensive Scoring System and Computerized Report*. San Antonio, TX: The Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A.** (1987). *The California Verbal Learning Test*. San Antonio, TX: The Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A.** (1994). *The California Verbal Learning Test—Children's Version*. San Antonio, TX: The Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A.** (2000). *CVLT-II: California Verbal Learning Test: Adult Version*. San Antonio, TX: The Psychological Corporation.
- Dudai, Y.** (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55*, 51–86.
- Dudai, Y., & Eisenberg, M.** (2004). Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron, 44*, 93–100.
- Dunn, L., & Dunn, L.** (2007). *Manual: Peabody Picture Vocabulary Test—Fourth Edition*. Bloomington, MN: Pearson Assessments.
- Elbro, C., Dalby, M., & Maarbjerg, S.** (2011). Language-learning impairments: A 30-year follow-up of language-impaired children with and without psychiatric, neurological and cognitive difficulties. *International Journal of Language & Communication Disorders, 46*(4), 437–448.
- Fazio, L. K., Huelser, B. J., Johnson, A., & Marsh, E. J.** (2010). Receiving right/wrong feedback: Consequences for learning. *Memory, 18*(3), 335–350.
- Fidler, L. J., Plante, E., & Vance, R.** (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology, 20*(1), 2–13.
- Gaulin, C. A., & Campbell, T. F.** (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills, 79*(1), 55–64.
- Gray, S.** (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research, 46*(1), 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*(5), 1117–1132.
- Gupta, P., & Tisdale, J.** (2009). Word learning, phonological short-term memory, phonotactic probability and long term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 3755–3771.
- Hall, J., McGregor, K. K., & Oleson, J.** (2017). Weaknesses in lexical-semantic knowledge among college students with specific learning disabilities: Evidence from a semantic fluency task. *Journal of Speech, Language, and Hearing Research, 60*, 640–653.
- Hart, B., & Risley, T. R.** (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.
- Horst, J. S., & Samuelson, L. K.** (2008). Fast mapping but poor retention in 24-month-old infants. *Infancy, 13*, 128–157.
- Isaki, E., Spaulding, T. J., & Plante, E.** (2008). Contributions of language and memory demands to verbal memory performance in language-learning disabilities. *Journal of Communication Disorders, 41*(6), 512–530.
- Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W.** (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169.
- Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S.** (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition, 3*(3), 198–206.
- Karpicke, J. D., & Grimaldi, P. J.** (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24*(3), 401–418.
- Karpicke, J. D., & Roediger, H. L.** (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968.
- Kaufman, A. S., & Kaufman, N. L.** (2004). *Kaufman Brief Intelligence Test—Second Edition*. Bloomington, MN: Pearson, Inc.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysaert, M.** (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*, 978–990.

- 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.
- Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2015). Verbal declarative memory impairments in specific language impairment are related to working memory deficits. *Brain and Language*, 142, 76–85.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- McGregor, K., Arbisi-Kelm, T., & Eden, N. (2017). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology*, 19(1), 43–57.
- McGregor, K. K., Langenfeld, N., Horne, S., Oleson, J., Anson, M., & Jacobson, W. (2016). The university experiences of students with learning disabilities. *Learning Disabilities Research & Practice*, 31(2), 90–102.
- McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, 56(6), 1845–1856.
- Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, 14(2), 225–229.
- Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & Cognition*, 37(8), 1077–1087.
- Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders*, 36(3), 221–231.
- Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and specific language impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology*, 19(1), 78–94.
- Morice, R., & McNicol, D. (1985). The comprehension and production of complex syntax in schizophrenia. *Cortex*, 21, 567–580.
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3, 41.
- Newman, L., Wagner, M., Knokey, A.-M., Marder, C., Nagle, K., Shaver, D., . . . Schwarting, M. (2011). *The post-high school outcomes of young adults with disabilities up to 8 years after high school* (A report from the National Longitudinal Transition Study-2 (NLTS2), NCSER 2011-3005). Retrieved from <http://www.nlts2.org/reports>
- Nichols, S., Jones, W., Roman, M. J., Wulfeck, B., Delis, D. C., Reilly, J., & Bellugi, U. (2004). Mechanisms of verbal memory impairment in four neurodevelopmental disorders. *Brain and Language*, 88(2), 180–189.
- Nippold, M. A., & Schwarz, I. E. (2002). Do children recover from specific language impairment? *Advances in Speech-Language Pathology*, 4(1), 41–49.
- Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words* (Research on Speech Perception, Progress Report No. 10). Bloomington, IN: Indiana University, Psychology Department, Speech Research Laboratory.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 3.
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1051.
- Plante, E., Ramage, A. E., & Magloire, J. (2006). Processing narratives for verbatim and gist information by adults with language learning disabilities: A functional neuroimaging study. *Learning Disabilities Research & Practice*, 21(1), 61–76.
- Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research*, 53(2), 414–429.
- Poll, G. H., Miller, C. A., & van Hell, J. G. (2016). Sentence repetition accuracy in adults with developmental language impairment: Interactions of participant capacities and sentence structures. *Journal of Speech, Language, and Hearing Research*, 59(2), 302–316.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803.
- Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mester & B. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford, UK: Elsevier.
- Sanford, C., Newman, L., Wagner, M., Cameto, R., Knokey, A. M., & Shaver, D. (2011). *The post-high school outcomes of young adults with disabilities up to 6 years after high school: Key findings from the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2011-3004). Washington, D.C.: National Center for Special Education Research.
- Sheng, L., Byrd, C. T., McGregor, K. K., Zimmerman, H., & Bludau, K. (2015). List memory in young adults with language learning disability. *Journal of Speech, Language, and Hearing Research*, 58(2), 336–344.
- Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, 49(6), 1175–1192.
- Tomblin, J. B., Freese, P. R., & Records, N. L. (1992). Diagnosing specific language impairment in adults for the purpose of pedigree analysis. *Journal of Speech and Hearing Research*, 35(4), 832–843.
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40(6), 1245–1260.
- Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences*, 28(1), 51–64.
- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
- Whitehouse, A. J., Watt, H. J., Line, E. A., & Bishop, D. V. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders*, 44(4), 511–528.