Published in final edited form as:

J Speech Lang Hear Res. 2013 December 1; 56(6): 1845–1856. doi:10.1044/1092-4388(2013/12-0233).

Why Words are Hard for Adults with Developmental Language Impairments

Karla K. McGregor, Ulla Licandro, Richard Arenas, Nichole Eden, Derek Stiles, Allison Bean, and Elizabeth Walker

The University of Iowa

Abstract

PURPOSE: To determine whether word learning problems associated with developmental language impairment (LI) reflect deficits in encoding or subsequent remembering of forms and meanings.

METHOD: Sixty-nine 18-25-year-olds with LI or without (ND) took tests to measure learning of 16 word forms and meanings immediately after training (encoding) and 12-hours, 24-hours, and 1-week later (remembering). Half of the participants trained in the morning and half in the evening.

RESULTS: At immediate posttest, those with LI performed more poorly on form and meaning than those with ND. Poor performance was more likely among those with more severe LI. The LI and ND groups demonstrated no difference in remembering word meanings over one week. In both groups, participants who trained in the evening, and therefore slept shortly after training, demonstrated greater gains in meaning recall than those who trained in the morning. In contrast, the LI-ND gap for word form recall widened over the week.

CONCLUSIONS: Some adults with LI have encoding deficits that limit the addition of word forms and meanings to the lexicon. Similarities and differences in patterns of remembering in the LI and ND groups motivate the hypothesis that consolidation of declarative memory is a strength for adults with LI.

Depending on age, context, and symptomology, developmental language impairments (LI) may be diagnosed as specific language impairment or learning disability or any of their subtypes (e.g., expressive specific language impairment, dyslexia). While not denying valid distinctions between specific language impairment and learning disability (Snowling, Bishop, & Stothard, 2000), language pathology is at the core of both (U.S. Department of Education, 2004).

Word learning problems tend to cut across these diagnostic boundaries. In some individuals, these problems represent an area of great weakness. When assessed across 19 domains, eight-year-olds with LI diagnosed as learning disability scored most poorly on a word

Karla McGregor Room 303 SHC Iowa City, IA 52242 karla-mcgregor@uiowa.edu.

Author Notes

Ulla Licandro is now at the Leibniz University of Hannover, Germany; Derek Stiles is now at Rush University; and Allison Bean is now at the Ohio State University. We thank Amanda Berns, Tim Arbisi-Kelm, Alison Bahnsen, Ashley Farris-Trimble, Joanna Lee, Megan Richards, Rachel See, Emily Czerniejewski, Gwyneth Rost, and Katy Mueller for their assistance. Mark Harris at the University of Iowa Student Disabilities Services was instrumental in recruitment of participants. Bob McMurray, Larissa Samuelson, and Bruce Tomblin provided helpful comments on an earlier draft. Portions of this paper were presented at the University of Sydney in February 2011 and the Symposium for Research on Child Language Disorders in Madison, WI in June 2011. The first author gratefully acknowledges the support of NIH-NIDCD 1R21DC009292-01 and a fellowship residency at the Obermann Center of the University of Iowa. Neither funding source played any role in the design, analysis, or writing of this study or in the decision to submit the paper for publication.

learning task (Korkman & Pesonen, 1994). Likewise, children with LI diagnosed as specific language impairment do not learn new words as readily as unaffected peers in response to incidental exposures (Oetting, Rice, & Swank, 1995; Rice, Oetting, Marquis, Bode, & Pae, 1994) or didactic training (Alt & Plante, 2006; Dollaghan, 1987; Gray, 2004).

Verbal memory for known words is poor as well, when attempting to find single words (Bell, McCallum, & Cox, 2003; Faust & Sharfstein-Friedman, 2003; German, 1982; Lahey & Edwards, 1999; McGregor, Newman, Reilly, & Capone, 2002) or recall word lists (Kail & Leonard, 1986; Kikas, Männamaa, Kumari, & Ulst, 2008; Nation, Adams, Bowyer-Crane, & Snowling, 1999). Relative to unaffected peers, children with LI also have more difficulty classifying words into taxonomic categories (Kikas et al., 2008; Siegel, Cook, & Gerard, 1995), guessing words when given their definitions (Männamaa, Kikas, & Raidvee, 2008), and providing definitions when given words (Kikas et al., 2008; Mainela-Arnold, Evans, & Coady, 2010; Marinelle & Johnson, 2002; McGregor, Berns, Owen, Michels, Duff et al., 2012; McGregor et al., 2002). In affected individuals, word knowledge remains deficient relative to peers throughout adolescence (McGregor, Oleson, Bahnsen, & Duff, 2013; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

In the current study we begin to determine why word learning is difficult for people with LI broadly defined. To do so, we trained young adults with or without LI on new word forms and their novel meanings and tested their memory for this new information immediately after training and three additional times over the course of a single week. By focusing on both form and meaning aspects of words as recalled at immediate and subsequent post-training intervals, we hoped to identify the particular challenges that young adults experience when learning new words.

Form or meaning?

Word knowledge is multifaceted. To know a word fully is to know its spoken and written form, its meaning, its morphology, syntax, and pragmatics. The learning of spoken word forms and their meanings has been most studied in the LI population. Nuanced tests demonstrate that it is sometimes difficult for people with LI to encode new word meanings (Alt & Plante, 2006); however, problems learning word forms are more often reported. For example, following a training period wherein new word forms and meanings were presented, children with specific language impairment performed significantly worse than unaffected age mates on measures of form (naming and recognition) and meaning (defining, answering questions about word meaning, and recognition of pictured referents). However, compared to vocabulary mates who were about two years their junior, they performed lower on one of the form measures only. Gray (2004) identified subgroups among children with specific language impairment, those who have difficulty learning both form and meaning and others who seem to have difficulty with form only. Theoretical accounts of specific language impairment assume greater deficits in form than meaning as well (Joanisse, 2004; Ullman & Pierpont, 2005).

Encoding or remembering?

Encoding is the process by which a new memory is formed. Encoding begins when the brain perceives a novel experience and, in the case of encoding a word, that novel experience is a new form, a new meaning, or both. The encoding of word forms and meanings involves activation of the left inferior prefrontal cortex and left medial temporal lobe (Kirchhoff, Wagner, Maril, & Stern, 2000). Activation of the medial temporal lobe is higher in response to fully novel stimuli (e.g., the first presentation of a novel word) than to less novel stimuli (e.g., subsequent presentations). The hippocampus too plays a role in novelty discrimination

and, therefore, plays a critical role in encoding as the greater the novelty, the greater the likelihood of encoding (Habib, McIntosh, Wheeler, & Tulving, 2003).

Over time, the newly encoded memory changes. At the simplest level, it may be forgotten or remembered. The most accepted model of memory formation posits that memory can transition over time from short-term to long-term storage and forgetting is much less likely after this transition has been achieved (Nader & Hardt, 2009). Consolidation is a process by which memories enter the long-term store. Via consolidation, memories strengthen and become less vulnerable to decay (Walker, 2005). They may also become more integrated into a network of related memories (Clay, Bowers, Davis, & Hanley, 2007; Dumay Gaskell, 2007). Consolidation is a slow process occurring over minutes, hours, days, or even longer periods depending upon the type of memory being consolidated (Nader & Hardt, 2009; Walker, 2005). According to Walker (2005), consolidation is not dependent upon external experience; rather, consolidation occurs as the brain “replays” newly encoded information over time (see also Wilson & McNaughton, 1994). Whereas the hippocampus provides crucial support to encoding, the neocortex supports consolidation (Davis & Gaskell, 2009; McClelland, McNaughton, & O’Reilly, 1995), and sleep is posited to contribute to the transfer between these complementary systems (Gais & Born, 2004; Ellenbogen, Hu, Payne, Titone, & Walker, 2007).

Consolidated memories are stable, that is, they will show no additional enhancement (or interference). However, once a consolidated memory is retrieved, it becomes labile thus providing an additional opportunity for change. Nader and Hardt (2009) term this “re-consolidation.” Consolidation and reconsolidation are not identical processes. For example, reconsolidation is completed more quickly than consolidation (Alberini, 2005). Also, the precise mechanisms of re-consolidation are a matter of debate. Nevertheless, the functional results are the same. In both cases, one possible result is a stronger memory trace. Reconsolidation is a likely contributor to the well documented “test effect” whereby repeated testing is known to support remembering, even when no feedback is provided to the test taker regarding the accuracy of the retrieved information (Roediger & Karpicke, 2006). Each test requires retrieval, which sets reconsolidation in motion.

It has long been assumed that the process of encoding verbal information is fragile among those affected by LI. Indirect evidence comes from training studies in which word learning improves when affected individuals are provided with more numerous (Gray, 2003; Horohov & Oetting, 2004; Kaplan, Dewey, Crawford, Fisher, 1998; Nash & Donaldson, 2005; Riches, Tomasello, & Conti-Ramsden, 2005) or more salient (Ellis Weismer & Hesketh, 1993; 1996; 1998) exposures to words, thus more useful experiences to support encoding. Evidence also comes from tasks involving recall of familiar but unrelated words. One pertinent example is the work of Nichols, Jones, Roman, Wulfeck, Delis, Reily, and Bellugi (2004). Using the California Verbal Learning Test-Children’s Version (Delis, Kramer, Kaplan, & Ober, 1994), they asked 6- to-14-year-olds with LI diagnosed as specific language impairment to learn a list of words. The list was presented five times with a recall probe administered after each. Compared to unaffected peers, those with LI encoded fewer of the words than their peers by the fifth exposure.

Whether or not people with LI are also poorer at remembering once encoding has been achieved is a matter of debate. In the study described above, Nichols et al. (2004) followed the fifth recall attempt with two additional attempts roughly 5 and 20 minutes later with no additional exposures to the word list in the interim. Recall at the shorter and longer delay was comparable for the LI participants just as for their peers, suggesting no deficits in remembering once the initial level of encoding was considered. However, problems with remembering over a longer time interval were evident in a training study wherein children

with LI and their unaffected peers were exposed to new words, either 3 or 10 times each, in a video story (Rice et al., 1994). Specifically, when given 10 exposures to each word, the children with LI performed as well as their same-age peers on an immediate recognition probe but, when that probe was repeated one-to-three days later, the children with LI demonstrated losses in accuracy relative to the immediate probe whereas their peers demonstrated gains.

Current study

Given these inconsistencies in the literature, our over-arching goal was to determine why words are hard for people affected by LI. Specifically, we asked whether word forms and word meanings were equally challenging learning targets and whether encoding or remembering better differentiated LI from ND learners.

Although not a direct replication, the protocol used here was influenced by Dumay, Gaskell, and colleagues. Dumay, Gaskell, and Feng (2004) trained adults with normal learning abilities on novel word forms and their meanings (e.g., *cathedruke* is a type of vegetable). Semantic priming effects (e.g., the extent to which *cathedruke* reduced reaction times for lexical decisions on *vegetable*) and semantic word association effects (e.g., the frequency with which *cathedruke* elicited a semantically related word association such as *vegetable*) were greater one week later than on training day but training day and 1-day posttests did not differ (Dumay et al., 2004), suggesting that consolidation of word meaning requires more than a single day.

Dumay and Gaskell (2007) taught adults with normal learning abilities novel word forms (in the absence of meaning) via phoneme monitoring. All participants were trained, tested immediately for level of encoding, and then tested again 12- and 24-hours later. The critical manipulation was that half of the participants trained in the evening and half in the morning. In the former, sleep intervened before the 12-hour posttest; in the latter, before the 24-hour posttest. A word recognition task required the participants to determine which of two spoken words was in the training set (e.g., *shadowks* or *shadowkt*) via 2-alternative forced choice (2AFC). A free recall task required them to say as many of the trained word forms as they could remember. Recognition was near ceiling immediately after training and at 12- and 24-hour posttests. Free recall improved significantly by the 12-hour posttest for those trained in the evening and by the 24-hour posttest for those trained in the morning suggesting a shorter time course for consolidation of word form than meaning and demonstrating a role for sleep in promoting consolidation.

In the current study, we devised a protocol that included some of the procedures from Dumay et al., (2004) and some from Dumay and Gaskell (2007) in an effort to examine the learning of both form and meaning targets in young adults with LI. We taught these young adults and then tested their memory four times over the course of a week, immediately post training and 12-hours, 24-hours, and 1-week post training, in an effort to examine problems with encoding as compared to remembering. For half of the participants (those assigned to the p.m. condition), the first interval of sleep occurred between the immediate and 12-hour posttests; for the others (those assigned to the a.m. condition), the first interval of sleep occurred between the 12- and 24-hour posttests (Figure 1).

Questions and predictions

Form or meaning

Relative to unaffected peers, do adults with LI have difficulty learning the *form* of the word or the *meaning* that the word conveys? If both, is one a greater deficit than the other? The

critical comparison is how the LI group performed relative to their peers on the recall and recognition of forms vs. meanings at each posttest. We predicted that both would be deficient. Given more numerous reports of problems with form than meaning in the literature, we predicted a greater deficiency in the learning of forms.

Encoding or remembering

Relative to unaffected peers, do adults with LI have difficulty with initial encoding or with remembering over time? If both, is one a greater deficit than the other? The critical comparison is how the LI group performed relative to their peers on recall and recognition at the immediate posttest vs. subsequent posttests. Given the extant literature, we predicted that both would be problematic. Given more numerous reports of encoding problems in the literature, we predicted a greater deficiency in encoding.

In this protocol, as in the real world, remembering depended upon a number of potential processes including consolidation, retrieval, and reconsolidation. As a preliminary step towards isolating consolidation from the others, we manipulated sleep intervals relative to the 12-hour posttest interval. We asked whether, like their unaffected peers, adults with LI demonstrate consolidation after a period of sleep. The critical comparison is how, relative to their peers, participants with LI perform on recognition and recall at the 12-hour posttest when assigned to the p.m. condition vs. the a.m. condition. If consolidation is intact, those in the p.m. condition should perform better because only they slept in the interval between training at the 12-hour posttest. Given the exploratory nature of this comparison, we did not make any firm predictions.

Method

Participants

Participants were 69 18-25 year olds from the Midwest of the United States. Of the 69 participants, 37 had no current or previous diagnosis (ND); 32 had current diagnoses of LI via self-report. Those with LI were asked to specify their LI diagnosis if known. Fourteen listed learning disability only; 6 listed reading disability only; 4 listed both learning disability and reading disability; 1 listed language disorder; 1 listed dysnomia; and 6 did not specify.

We could document the diagnoses of 25 of the 32 LI participants from other sources. We recruited 16 participants with LI from the Office of Students with Disabilities on their college campuses. Their diagnoses had been documented to verify their need for classroom accommodations, a process that involved administration of the Wechsler Individual Achievement Test (Wechsler, 1992) or the Woodcock Johnson Psychoeducational Battery (Woodcock, McGrew, & Mather, 2001) to establish academic impairment as well as the Wechsler Adult Intelligence Scale-III (Wechsler, 1997) to establish average or above IQ. We recruited nine of the 32 participants with LI from the Tomblin epidemiologic project. Therefore, their history of LI had been documented via the EpiSLI diagnostic system (Tomblin, Records, Zhang, 1996), which also profiled according to poor standardized language scores but IQ within normal limits.

We administered the Test of Adolescent and Adult Language-4 (TOAL4, Hammill, Brown, Larsen, & Wiederholt, 2007), a modified version of the Token Test of Language Comprehension (Morice & McNicol, 1985), the Peabody Picture Vocabulary Test-IV (Dunn & Dunn, 2007), the Expressive Vocabulary Test (Williams, 2007), the nonverbal matrices of the Kaufman Brief Intelligence Test (KBIT, Kaufman & Kaufman, 2004), and a pure tone audiometric screening at .5, 1, 2, and 4 kHz at 25dB bilaterally to all participants. All participants had to pass the audiometric test.

To enroll, those with unconfirmed diagnoses had to earn at least two scores that fell more than 1 standard deviation below the mean on the subtests of the TOAL4 to establish a language problem and score no more than 1 standard deviation below the mean on the KBIT to ensure that any language problem was not part of a broader intellectual deficit. We compared this unconfirmed group to the other two groups and found differences only in education (the EpiSLI group had 2 fewer years on average than the unconfirmed group, $t = 2.63$, $df = 14$, $p = .02$) and KBIT standard scores (those referred from the Office of Students with Disabilities averaged 107, the unconfirmed cases averaged 97, $t = 1.98$, $df = 21$, $p = .06$). The unconfirmed cases did not differ from the other LI subgroups on any standardized language measure. The TOAL4 subtests represent a mix of oral and written tasks. The majority of the LI group (28) scored below the mean on a mix of oral and written tasks; 3 on written tasks only; and 1 on oral tasks only.

Table 1 summarizes demographics and test scores and compares the LI and ND groups on each. Note that the LI group consistently scored lower on all measures of language. Also, although each participant scored within one standard deviation of the mean on the KBIT, the LI group still scored significantly lower than the ND group. This is a frequently noted characteristic of the LI phenotype (Tomblin, Smith, & Zhang, 1997).

Because attention deficit/hyperactivity disorder (ADHD) is frequently comorbid with LI (Cantwell & Baker, 1991) and because inattention could adversely affect learning, we administered the Conners Adult ADHD Rating Scales-Screening Version (Conners, Erhardt, & Sparrow, 1999). Four of the 32 participants with LI (12.5%) and two of the 37 unaffected participants (5%) met clinical cutoffs for ADHD. Fisher's exact test revealed no significant difference in rate of elevated symptoms in the two groups of participants upon enrollment in the study, $p_{1-tailed} = .27$, $p_{2-tailed} = .41$.

To better control the effects of sleep intervals in relation to the learning interval, we asked participants to refrain from napping during the day of each visit and to keep a log of the amount of nighttime sleep they received the night before each visit. Participants recorded time to bed, time to sleep, number of awakenings during the night and their length, and the time awake in the morning to allow objective estimates of total sleep. They also made qualitative judgments about the amount and quality of sleep and about how they felt at the time of the visit.

The sleep logs revealed similar amounts of sleep between diagnostic groups with the LI group averaging 6.6 hours of sleep per night ($SD = 71.75$) and the ND group averaging 6.5 hours ($SD = 65.69$), $t = .48$, $df = 67$, $p = .63$. The range across individuals was great with as little as 2.75 hours and as much as 10.13 hours reported. The two groups made identical subjective judgments: judging their sleep as "sort of enough," the quality of their sleep as somewhere between "ok" and "good," and their current feeling as "a little tired."

Materials

The stimuli were 32 word triplets each of which included a disyllabic, monomorphemic English word (e.g., *army*) and two pseudowords that diverged from the English word at the final syllable (e.g., *armo* and *armu*) (Supplemental Appendix A). To ensure familiarity, the English words were found in childhood reading materials (Moe, Hopkins, & Rush, 1982). The pseudowords had the same stress pattern as the English words that served as their base. For each triplet, one pseudoword served as a novel word to be trained and the other as an untrained foil in the word subtest of the 2AFC test.

To give meaning to the 32 novel words, each was randomly assigned to a fantasy referent (Supplemental Appendix B). The referents, created by combining two animate objects (e.g.,

a pony and a snake) or inanimate objects (e.g., a phone and a baseball bat), were depicted in black line drawings. A semantic neighbor was also drawn for each referent by combining the same base category (e.g., pony or phone) with a different object (e.g., shark or bone). The neighbors served as untrained foils in the referent subtest of the 2AFC test. To ensure familiarity, the base categories were selected because their names appear in children's books (Moe et al., 1982).

The 32 novel words and referents were divided into two sets of 16 so that eight items in each set referred to an animate object and eight referred to an inanimate object. One of the two sets was trained and selection of the training set was counterbalanced across participants.

Procedure

Each participant first completed an enrollment visit during which they gave informed consent and took standardized tests of language, nonverbal cognition, hearing, and ADHD.

Participants were randomly assigned to a training condition. Those assigned to the a.m. condition began their training visit in the morning; those in the p.m. condition began in the evening. As a result 16 people from the LI group and 18 people from the ND group were in the a.m. condition and 16 people from the LI group and 19 people from the ND group were in the p.m. condition. During the training visit, each participant was interviewed about sleep function, trained on the new words and referents, and then tested immediately via free recall, lexical decision, 2AFC recognition, and word association (in that order). In subsequent visits these four tests were repeated but training was not.

Training—Participants viewed and listened to a computerized training script presented in 12 blocks. In each block, each of the 16 novel referents was displayed on the screen for 7 seconds. During this time, the participant heard the associated novel word four times within a 2-sentence description. The description always comprised the semantic category name and an illustrative sentence (Supplemental Appendix B). To ensure that the participant was actively engaged, a question about either the sounds in the novel word (e.g., “does it start with /o/?”) or the physical features of its referent (e.g., “does it have eyes?”) was presented after a third of all items (4 times for each item across 12 blocks). Half of the questions pertained to the word and half to the referent. No feedback on accuracy was provided. This procedure was repeated for a total of 12 blocks, yielding 12 exposures to each referent and verbal description and 48 exposures to each word form (4 per description × 12 blocks). The order of presentation was randomized within blocks and the order of blocks was counterbalanced across participants.

Tests—We administered four tests with no feedback on accuracy during each visit. Of interest here were the 2AFC recognition, free recall, and word association tests. A lexical decision test in which participants made timed lexical decisions (“yes” for English words; “no” for non-English pseudowords) in response to 128 items that included the 32 novel words from sets A and B and their 32 English neighbors is not considered here as the number of participants did not yield enough power to produce interpretable results.

2AFC: There were three 2AFC recognition tasks used to measure learning of forms, referents, and their linkage. In 2AFC-word, participants heard each novel word (e.g., *armo*) paired with its untrained novel lexical neighbor (e.g., *armu*) separated by an ISI of 500 ms and they pressed a button to indicate which was familiar. In 2AFC-referent, they saw each novel referent (e.g., snake-pony) paired with its untrained semantic neighbor (e.g., shark-pony) presented one on the left and one on the right of the computer screen and, again, pressed a button to indicate which was familiar. These remained on the screen until the

button press. In 2AFC-link, for half of the items, participants heard a word (e.g., *armo*) and pressed a button to indicate whether that word named the pictured referent on the left (e.g., a snake-pony) or right (e.g., a beaver-turtle) of the screen, with the left and right pictures being randomly paired animate or inanimate items from the training set. For the other half, the participants saw a pictured referent (e.g., beaver-turtle) and pressed a button to indicate whether that picture went with the first word (e.g., *buckedge*) or second word (e.g., *partrip*) they heard. The timing parameters paralleled those in the word and referent subtests. There were four versions of each of the three subtests that varied in order of item presentation within and between pairs. Assignment of versions to test period was counterbalanced across participants. Responses were scored for accuracy.

Free recall: Free recall was used to measure learning of word forms. In free recall, participants had 2 minutes to recall orally as many of the trained words as possible. The dependent variable of interest was number of words recalled. Words only needed to be recognizable as one of the trained words to be counted; that is, they did not have to be produced with 100% accuracy. Working from digital audio, two investigators, blind to condition assignment, independently transcribed 22% of the data. Their point-to-point agreement on which word was being attempted was 100%. After establishing this level of agreement, one of the investigators transcribed the remaining data and coded it for words recalled.

Word association: The word association test was used to measure learning of word meanings. Trained words were presented via an audiorecording and the participant said the first word that came to mind after each. There were four versions of the script that varied in order of item presentation. Assignment of versions to tests was counterbalanced across participants. Two investigators, blind to condition assignment, independently scored 640 responses (approximately 15% of the data) for number of correct semantic relationships (words that related to the category as presented in both the training script and the pictured referent [e.g., pony or horse for *armo*] or words that related to the subcategory as presented in the pictured referent only [e.g., snake or rattler for *armo*]). They disagreed on 7 responses for a point-to-point agreement of 98.9%. After establishing agreement, one investigator scored the remaining data.

Results

Encoding

2AFC—Immediately after training in both LI and ND groups, performance was significantly above chance for recognition of words, referents, and their links, $ps < .0001$; in fact, performance neared ceiling (Figure 2). Given ceiling level scores, we applied an arcsine transformation to improve normality and homogeneity of variance. Inspection of histograms revealed improved normality but the transformation yielded homogeneous variances for the recognition of links only, Levene's $F = 1.90$, $p = .30$; not words, Levene's $F = 9.49$, $p = .003$, or referents, Levene's $F = 5.94$, $p = .02$. Therefore, we conducted a 2 (language diagnosis) $\times$ 2 (training condition) between-subjects ANOVA and used the transformed values of link scores only as a dependent variable. The result was a main effect of language diagnosis, $F(1,65) = 5.55$, $p = .02$, $\text{partial } \eta^2 = .08$, with the ND group outperforming the LI group. Eight of the 32 participants with LI (25%) and 5 of the 37 participants with ND (14%) fell farther than 1 SD below the ND group mean. There were no effects of training condition, $F(1,65) = 2.42$, $p = .12$, and no diagnosis $\times$ training interactions, $F(1,65) < 1$.

Free recall—We conducted a 2 (language diagnosis) $\times$ 2 (training condition) between-subjects ANOVA with number of word forms recalled at the immediate posttest as the dependent variable. There was a main effect of language diagnosis, $F(1,65) = 7.49, p = .008$, $\text{partial } \eta^2 = .10$, as the ND group out-performed the LI group with means of 6.08 ($SD = 3.54$) and 3.94 ($SD = 2.93$), respectively (Figure 3, top). Eleven of the 32 participants with LI (34%) and 5 of the 37 participants with ND (14%) fell farther than 1 SD below the ND group mean. There were no effects of training condition, $F(1,65) < 1$, and no diagnosis $\times$ training interactions, $F(1,65) = 2.23, p = .14$.

Word Association—We conducted a 2 (language diagnosis) $\times$ 2 (training condition) between-subjects ANOVA with number of semantic responses at the immediate posttest as the dependent variable. There was a main effect of language diagnosis, $F(1,65) = 6.7, p = .01$, $\text{partial } \eta^2 = .09$, with the ND group out-performing the LI group (Figure 3, bottom). On average, the ND group gave 6.65 ($SD = 5.65$) correct semantic word associations to trained words at immediate posttest and the LI group gave 3.72 ($SD = 4.13$). Ten of the 32 participants with LI (31%) and 5 of the 37 participants with ND (14%) fell farther than 1 SD below the mean of the ND group. There were no effects of training condition, $F(1,65) = 1.23, p = .27$, and no diagnosis $\times$ training interactions, $F(1,65) < 1$.

Remembering

Because the two diagnostic groups differed in encoding on the immediate 2AFC, free recall, and word association tests, they differed in the extent of potential gains or losses in accuracy that they could demonstrate over time. Therefore, we examined change over time on these tests via normalized difference scores (Ebbels, van der Lely, & Dockrell, 2007). For each test we first compared the normalized difference scores of the LI and ND groups assigned to p.m. and a.m. conditions. When there were significant effects of diagnostic group or training, we then compared difference scores to zero to determine whether they represented significant changes from the immediate posttest performance.

2AFC—Changes in memory for words and referents were minimal with proportion of correct responses ranging from .88 to 1 for both groups and both training conditions across test periods. Although significantly above chance ($ps < .01$), scores were slightly below ceiling for link recognition. Therefore, we analyzed normalized difference scores for link recognition only. A 2(diagnostic group) $\times$ 2(training condition) $\times$ 3(test period) mixed ANOVA with repeated measures on the final variable revealed no main effects of diagnosis, $F(1,65) < 1$, or test, $F(2, 130) > 1$; and no interactions. The effect of training condition was marginal, $F(1,65) = 3.34, p = .07$, with a trend towards better performance in the a.m. condition.

Free recall—A 2(diagnostic group) $\times$ 2(training condition) $\times$ 3(test period) mixed ANOVA with repeated measures on the final variable and a normalized difference score based on number of words correctly recalled as the dependent variable revealed no main effects of diagnosis, $F(1,65) < 1$, or training condition, $F(1,65) < 1$. There was also no interaction between training and test conditions; however, consistent with previous reports (Dumay & Gaskell, 2007), those assigned to the p.m. condition performed numerically better at the 12-hour posttest than those assigned to the a.m. condition. There was a main effect of test, $F(2,130) = 10.92, p < .0001$, $\text{partial } \eta^2 = .14$, and this was qualified by a test $\times$ diagnostic group interaction, $F(2,130) = 4.22, p = .02$, $\text{partial } \eta^2 = .06$ (Figure 4). A Bonferroni post hoc test revealed that the ND cohort made significantly larger gains in number of words recalled at the 24- and 1-week posttest than at the 12-hour posttest. One sample t-tests comparing difference scores to zero (the immediate posttest baseline) revealed significant gains at the 243hour posttest ($t = 2.30, df = 36, p = .03, d = .38$) and the 1-week

posttest ($t = 3.06$, $df = 36$, $p = .004$, $d = .50$) for the ND group. Although the gain at the 24-hour posttest was significantly greater than zero for the LI group, $t = 2.15$, $df = 31$, $p = .04$, $d = .34$, the changes over baseline at 12-hour, 24-hour, and 1-week posttests did not vary significantly one from the other.

Word association—A 2(diagnostic group) $\times$ 2(training condition) $\times$ 3(test period) mixed ANOVA with repeated measures on the final variable revealed a main effect of test, $F(2,124) = 4.97$, $p = .008$, $\text{partial } \eta^2 = .07$, such that change over baseline was larger at the 24-hour- and 1-week posttests than at the 12-hour posttest, $ps < .03$ (Figure 5). There was a main effect of training condition, $F(1,62) = 8.11$, $p = .006$, $\text{partial } \eta^2 = .12$, such that the p.m. condition demonstrated greater gains over baseline than the a.m. condition. One sample t-tests comparing difference scores to zero (the immediate posttest baseline) revealed significant gains at the 12-hour posttest ($t = 3.24$, $df = 33$, $p = .003$, $d = .54$), 24-hour posttest ($t = 4.43$, $df = 33$, $p < .0001$, $d = .76$) and 1-week posttest ($t = 4.56$, $df = 33$, $p < .0001$, $d = .76$) for those in the p.m. condition. In contrast, none of the gains made by those in the a.m. condition were greater than zero, $ps > .35$. There was no main effect of diagnostic group, $F(1,62) < 1$, and no interactions.

Individual differences—In summary, immediately after training, young adults with LI recognized fewer links between words and their referents, recalled fewer word forms, and made fewer semantic associations than their ND peers. Unlike their ND peers, they also failed to improve word form recall over the course of the week. To further explore the weaknesses demonstrated by the LI group, we correlated the number of word forms recalled during free recall with the number of semantic responses given during word association, both from the immediate posttest, to determine whether those who were poorer at encoding word forms were also those who were poorer at encoding word meanings. The result confirmed a moderate relationship, $r = .40$, $p = .02$. The amount of change in word form recall by the end of the week was correlated with the encoding of word meaning at the immediate pretest, $r = .36$, $p = .04$ but not with the encoding of word form at the immediate pretest, $p > .5$. It was also the case that those who were poorer at encoding word forms tended to score lower on the PPVT-IV, $r = .46$, $p = .007$, and the modified Token Test, $r = .36$, $p = .04$. Those who were poorer at encoding word meanings tended to score (marginally) lower on the TOAL4, $r = .34$, $p = .06$.

Discussion

As a group, young adults with LI presented with word learning deficits. Given the same amount of training as their unaffected peers, they encoded less information about word forms, word meanings, and the linkage of form to meaning. They did not fail to encode—their near-ceiling level performance on the recognition tests and their above floor-level performance on the two production tests prove this—but they encoded less. The gap between LI and ND groups was moderate in effect size for the encoding of both form and meaning.

Once initial levels of encoding were taken into account, the LI and ND groups demonstrated no differences in remembering word meanings or word form-to-meaning links over the course of one week. However, this was not the case for word form recall. Over the week, those with LI demonstrated fairly stable levels of form recall relative to the immediate posttest but those with ND demonstrated significant gains. Therefore, the gap between LI and ND in the learning of word forms, which was already of moderate size upon encoding, grew larger over time.

Overall then, in this particular protocol, the young adults with LI presented with word learning deficits characterized by poor encoding of both form and meaning and poor memory for form over time. However, these deficits were not characteristic of all participants with LI. Depending on the test, 25-to-34% of the participants with LI fell lower than 1 standard deviation from the ND mean on the immediate posttests. This rate is similar to the rate of poor word learners among preschoolers with SLI (Gray, 2004). Within the LI group, those with poorer encoding of meaning also tended to have poorer encoding of form and less gain in memory for form over time. They also tended to have lower scores on standardized tests of language suggesting that word learning deficits are more likely among those with more severe LI. It is worth pointing out then, that the current sample, comprised of those who were managing their LI successfully enough to attend college, might yield an underestimate of the rate of word learning deficits in the broader population of young adults with LI.

Why should it be that the LI group lagged farther and farther behind their ND peers in the learning of word form, but not word meaning, over time? Because the learning and use of rule-governed aspects of phonological form involves procedural memory (Gupta & Tisdale, 2009; Ullman & Pierpoint, 2005), the greater problem with form is consistent with accounts that posit procedural learning deficits as the core of LI (Ullman & Pierpoint, 2005). Tests that elicited stronger performances in the current study likely by-passed weaknesses in procedural memory by minimizing retrieval demands (the 2AFC tests) or tapping declarative memories (the word association test).

Baddeley's work offers an alternative, but not mutually exclusive, explanation. According to Baddeley (Baddeley & Hitch, 1974; Baddeley, 2000) working memory is partitioned into the phonological loop, the visual spatial sketch pad, episodic buffer, and the central executive. These temporary memory stores feed and interface with long term memory and the phonological loop in particular is hypothesized to support long-term memory of language. People with LI have deficits in the phonological loop (Gathercole & Baddeley, 1990) and these are so pervasive as to make nonword repetition (a test that taxes the phonological loop) a useful clinical tool, especially for younger children with LI (Conti-Ramsden, 2003). Tests that elicited stronger performances in the current study likely by-passed weaknesses in the phonological loop by minimizing phonological retrieval demands (the 2AFC tests) or tapping memories that are supported, in part, by the visual spatial sketch pad (the word association test).

However, these explanations alone fail to account for the finding that the size of the gap between LI and ND groups was equally large for meaning as for form at the *immediate* posttest. At the immediate posttest, but not the subsequent ones, the LI group performed as poorly on the word association test (our measure of meaning) as on the free recall test (our measure of form).

A third potential explanation emerges from the current data. First consider the possibility that the gains in form recall demonstrated by the ND group might, in large part, reflect encoding in response to repeated test exposures. In particular, the word association test required the examiner to present each of the trained words to the participant so that he or she could provide a word association; therefore, each word association test was an opportunity for additional encoding of form. The word forms were also presented in the form subtest of the 2AFC test. In contrast, the opportunities for encoding of meaning during the posttests were limited to the pictures of referents in the 2AFC test. No spoken definitions or associated words were provided. If gains in form over time were heavily dependent upon encoding but gains in meaning were not and if, as previous and current evidence suggests,

people with LI are poor encoders, then it follows that the LI-ND gap in form, but not meaning, would grow larger over the course of the week.

Next, note that, in the ND group, memory for meaning improved over the course of the week but improvements were significant only for those assigned to the p.m. training condition. Consider the possibility that gains in meaning demonstrated by the ND group might, in large part, reflect consolidation in declarative memory facilitated by sleep closely following the period of encoding. This was, in fact, the conclusion reached by Gais, Lucas, and Born (2006) who found sleep close to encoding to be critical to retention of English-German translation equivalents (words that share the same meaning). Note also that the participants with LI, who reported no differences in amount or quality of sleep relative to their peers, replicated the ND patterns of growth on the word association test. They also increased semantic responding over time but only if assigned to the p.m. condition. If gains in meaning over time were heavily dependent upon critically timed sleep-based consolidation of declarative memory but gains in form were not, and if people with LI have robust sleep-based consolidation in the declarative system, then it follows that the LI-ND gap in meaning would not grow larger over the course of the week.

In summary, one explanation that accounts for the patterns in the current data set is that, growth in memory for word forms depended less on sleep-based consolidation of declarative memory (a hypothesized strength for the LI group) and more on encoding from repeated exposures (a documented weakness for the LI group). This explanation is perfectly compatible with the others. The procedural deficit and phonological loop deficit hypotheses emphasize the type of memory that is problematic whereas the third hypothesis emphasizes the process of memory. The process of encoding is deficient.

Open questions

At least two open questions are ripe for investigation. First, in the repeated testing approach used to track changes over time, remembering was likely supported by consolidation, retrieval, and reconsolidation. Enhancements likely reflected practice as well; as the participants became very familiar with the test protocol, stimuli, and response types required, their performance improved. In other words, more tests allowed more practice of the tests, a benefit typically explained as a case of transfer-appropriate processing (Morris, Brandsford, & Franks, 1977). In the future, it would be useful to measure consolidation alone by eliminating test intervals between training and the retention interval of interest. Doing so would provide a crucial test for the hypothesis generated here; namely, that consolidation of declarative memory is a relative strength for young adults with LI. Moreover, this disentangling would allow a cleaner comparison of consolidation in the declarative and procedural memory systems, a step motivated by evidence that the consolidation of procedural memory is impaired in children with LI who have grammatical deficits (Hedenius, Persson, Tremblay, Adi-Japha, Verissimo, et al., 2011).

Second, the finding that gains in memory for word meaning were facilitated when sleep closely followed encoding is relatively novel. But, given that we found the same pattern in two independent groups, those with and without LI, and that Gais and colleagues (2006) reached the same conclusion, it is one to be considered seriously. Others (Clay et al., 2007; Dumay et al., 2004) have reported consolidation-based gains in meaning over the course of a week as characteristic of an entire group of participants but they did not explore individual differences within the group and whether these differences relate to the timing of sleep relative to encoding. The hypothesis that sleep should be critically timed to encoding to achieve optimal memory is an exciting one to pursue as it could have implications for training and study schedules.

Conclusions

The similarities and differences in patterns of change in participants with and without LI led us to hypothesize that consolidation of declarative memory is a relative strength for young adults with LI. Nevertheless, many of these young adults find it difficult to learn new words. Given the current evidence, we conclude that this difficulty arises because encoding deficits limit the addition of word forms and meanings to the long-term lexicon.

References

- Alt M, Plante E. Factors that influence lexical and semantic fast mapping of young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2006; 49:941–954.
- Baddeley AD. The episodic buffer: A new component for working memory? *Trends in Cognitive Sciences*. 2000; 4:417–23. [PubMed: 11058819]
- Baddeley, AD.; Hitch, GJ. Working memory. In: Bower, GH., editor. *The psychology of learning and motivation*. Vol. 8. Academic Press; New York: 1974. p. 47-90.
- Bell SM, McCallum RS, Cox EA. Toward a research-based assessment of dyslexia: Using cognitive measures to identify reading disabilities. *Journal of Learning Disabilities*. 2003; 36:505–516. [PubMed: 15493433]
- Clay F, Bowers JS, Davis CJ, Hanley DA. Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007; 33:970–976.
- Conners, CK.; Erhardt, D.; Sparrow, E. *Conners' adult ADHD rating scales (CAARS)*. Multi-Health Systems; North Tonawanda, N.Y.: 1999.
- Conti-Ramsden G. Processing and linguistic markers in young children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2003; 46:1029–1037.
- Davis MH, Gaskell MG. A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society*. 2009; 364:3773–3800.
- Delis, DC.; Kramer, JH.; Kaplan, E.; Ober, BA. *The California Verbal Learning Test-Children's Version*. Psychological Corporation; San Antonio, TX: 1994.
- Dollaghan C. Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Disorders*. 1987; 52:218–222. [PubMed: 3455444]
- Dunn, LM.; Dunn, DM. *Peabody Picture Vocabulary Test, Fourth Edition*. Pearson; San Antonio, TX: 2007.
- Dumay N, Gaskell MG. Sleep-associated changes in the mental representation of spoken words. *Psychological Science*. 2007; 18:35–39. [PubMed: 17362375]
- Dumay, N.; Gaskell, MG.; Feng, X. A day in the life of a spoken word. In: Forbus, K.; Gentner, D.; Regier, T., editors. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*. Erlbaum; Mahwah, NJ: 2004. p. 339-344.
- Ebbels S, van der Lely H, Dockrell J. Intervention for verb argument structure in children with persistent SLI: A randomized control trial. *Journal of Speech, Language, and Hearing Research*. 2007; 50:1330–1349.
- Ellenbogen J, Hu P, Payne JD, Titone D, Walker MP. Human relational memory requires time and sleep. *Proceedings of the National Academy of Science USA*. 2007; 104:7723–7728.
- Ellis Weismer S, Hesketh LJ. The influence of prosodic and gestural cues on novel word acquisition by children with specific language impairment. *Journal of Speech and Hearing Research*. 1993; 39:177–190. [PubMed: 8820709]
- Ellis Weismer S, Hesketh LJ. Lexical learning by children with specific language impairment: Effects of linguistic input presented at varying speaking rates. *Journal of Speech and Hearing Research*. 1996; 39:177–190. [PubMed: 8820709]
- Ellis Weismer S, Hesketh LJ. The impact of emphatic stress on novel word learning by children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1444–1458.

- Faust M, Sharfstein-Friedman S. Naming difficulties in adolescents with dyslexia: Application of the tip-of-the-tongue paradigm. *Brain and Cognition*. 2003; 53:211–217. [PubMed: 14607150]
- Gais S, Born J. Declarative memory consolidation: Mechanisms acting during human sleep. *Learning and Memory*. 2004; 11:679–685. [PubMed: 15576885]
- Gais S, Lucas B, Born J. Sleep after learning aids memory recall. *Learning & Memory*. 2006; 13:259–62. [PubMed: 16741280]
- Gathercole SE, Baddeley AD. Phonological memory deficits in language disordered children: Is there a connection? *Journal of Memory and Language*. 1990; 29:336–360.
- German DJ. Word-finding substitutions in children with learning disabilities. *Language, Speech, and Hearing Services in Schools*. 1982; 13:223–230.
- Gray S. Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language, and Hearing Research*. 2003; 46:56–67.
- Gray S. Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research*. 2004; 47:1117–1132.
- Gupta P, Tisdale J. Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*. 2009; 61:481–502.
- Habib R, McIntosh AR, Wheeler MA, Tulving E. Memory encoding and hippocampally-based novelty/familiarity discrimination networks. *Neuropsychologia*. 2003; 41:271–279. [PubMed: 12457753]
- Hammill, DD.; Brown, VL.; Larsen, SC.; Wiederholt, JL. *Test of Adolescent and Adult Language - Fourth Edition (TOAL-4)*. Pearson; San Antonio, TX: 2007.
- Hedenius M, Persson J, Tremblay A, Adi-Japha E, Veríssimo J, Dye CD, Alm P, Jennishce M, Tomblin JB, Ullman MT. Grammar predicts procedural learning and consolidation deficits in children with Specific Language Impairment. *Research in Developmental Disabilities*. 2011; 32:2362–2375. [PubMed: 21840165]
- Horohov JE, Oetting JB. Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics*. 2004; 25:43–65.
- Joanisse MF. Specific language impairments in children: Phonology, semantics, and the English past tense. *Current Directions in Psychological Science*. 2004; 13:156–160.
- Kail, R.; Leonard, LB. *ASHA Monographs 25*. American Speech-Language-Hearing Association; Rockville, MD: 1986. Word-finding abilities in language-impaired children.
- Kaplan BJ, Dewey D, Crawford SG, Fisher GC. Deficits in long-term memory are not characteristic of ADHD. *Journal of Clinical and Experimental Neuropsychology*. 1998; 20:518–528. [PubMed: 9892055]
- Kaufman, AS.; Kaufman, NL. *The Kaufman Brief Intelligence Test-2*. AGS Publishing; Circle Pines, MN: 2004.
- Kikas E, Männamaa M, Kumari V, Ulst T. The relationships among verbal skills of primary school students with specific learning disabilities and a typically developing comparison group. *International Journal of Disability, Development, and Education*. 2008; 55:315–329.
- Kirchhoff BA, Wagner AD, Maril A, Stern CE. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience*. 2000; 20:6173–6180. [PubMed: 10934267]
- Korkman M, Pesonen AE. A comparison of neuropsychological test profiles of children with attention deficit hyperactivity disorder and/or learning disorder. *Journal of Learning Disabilities*. 1994; 27:383–392. [PubMed: 8051511]
- Lahey M, Edwards J. Naming errors of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 1999; 42:195–205.
- Mainela-Arnold E, Evans JL, Coady J. Explaining lexical-semantic deficits in specific language impairment: The role of phonological similarity, phonological working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research*. 2010; 53:1742–1756.
- Männamaa M, Kikas E, Raidvee A. The effect of testing condition on word guessing in elementary school children. *Journal of Psychoeducational Assessment*. 2008; 26:16–26.

- Marinelle S, Johnson C. Definitional skill in school-age children with specific language impairment. *Journal of Communication Disorders*. 2002; 35:241–259. [PubMed: 12064786]
- McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychology Review*. 1995; 102:419–57.
- McGregor KK, Berns AJ, Owen AJ, Michels SA, Duff D, Bahnsen AJ, Lloyd M. Associations between syntax and the lexicon among children with or without ASD and language impairment. *Journal of Autism and Developmental Disorders*. 2012; 42:35–47. [PubMed: 21365354]
- McGregor KK, Newman RM, Reilly RM, Capone N. Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2002; 45:998–1014.
- McGregor KK, Oleson J, Bahnsen A, Duff D. Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language and Communication Disorders*. 2013 DOI: 10.1111/1460-6984.12008.
- Moe, AJ.; Hopkins, KJ.; Rush, RT. The vocabulary of first grade children. Thomas; Springfield, IL: 1982.
- Morice R, McNicol D. The comprehension and production of complex syntax in Schizophrenia. *Cortex*. 1985; 21:567–580. [PubMed: 4092484]
- Morris CD, Brantford JD, Franks JJ. Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*. 1977; 16:519–533.
- Nader K, Hardt O. A single standard for memory: The case for reconsolidation. *Nature Reviews: Neuroscience*. Mar.2009 10:224–234.
- Nash M, Donaldson M. Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*. 2005; 48:4–93458.
- Nation K, Adams JW, Bowyer-Crane CA, Snowling MJ. Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*. 1999; 73:139–158. [PubMed: 10328862]
- Nichols S, Jones W, Roman MJ, Wulfeck B, Delis DC, Reilly J, Bellugi U. Mechanisms of verbal memory impairment in four neurodevelopmental disorders. *Brain and Language*. 2004; 88:180–189. [PubMed: 14965540]
- Oetting J, Rice M, Swank L. Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*. 1995; 38:443–445.
- Rice ML, Oetting JB, Marquis J, Bode J, Pae S. Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*. 1994; 37:106–122. [PubMed: 8170118]
- Riches NG, Tomasello M, Conti-Ramsden G. Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research*. 2005; 48:1397–1411.
- Roediger HL III, Karpicke JD. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*. 2006; 1:181–210.
- Siegel J, Cook R, Gerard J. Differential categorization of words by learning disabled, gifted, and nonexceptional students. *Perceptual and Motor Skills*. 1995; 81:242–250.
- Snowling MJ, Bishop DVM, Stothard SE. Is pre-school language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry*. 2000; 41:587–600. [PubMed: 10946751]
- Stothard SE, Snowling MJ, Bishop DVM, Chipchase BB, Kaplan CA. Language impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*. 1998; 41:407–418.
- Tomblin JB, Records N, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*. 1996; 39:1284–1294. [PubMed: 8959613]
- Tomblin JB, Smith E, Zhang X. Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders*. 1997; 30:325–344. [PubMed: 9208366]
- Ullman MT, Pierpont EI. Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*. 2005; 41:399–433. [PubMed: 15871604]

- U.S. Department of Education. Building the Legacy: IDEA 2004. 2004. Retrieved from <http://idea.ed.gov/explore/view/p/%2Croot%2Cstatute%2CI%2CA%2C602%2C30%2C>
- Venkateswaran S, Shevell M. The case against routine electroencephalography in specific language impairment. *Pediatrics*. 2008; 122:e911–e916. [PubMed: 18794194]
- Walker MP. A refined model of sleep and the time course of memory formation. *Behavioral and Brain Sciences*. 2005; 28:51–104. [PubMed: 16047457]
- Wechsler, D. Wechsler Individual Achievement Test. Psychological Corporation; San Antonio, TX: 1992.
- Wechsler, D. Wechsler Adult Intelligence Scale-III (WAIS-III). Psychological Corporation; San Antonio, TX: 1997.
- Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. *Science*. 1994; 265:676–679. [PubMed: 8036517]
- Williams, KT. Expressive Vocabulary Test, Second Edition. AGS Publishing; Circle Pines, MN: 2007.
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson Psychoeducational Battery III. Riverside Publishing; Itasca, IL: 2001.

	Day 1 for a.m. training condition	Day 2 for a.m. training condition	Day 7 for a.m. training condition
morning	Trained on new word forms and their meanings Immediate posttest of form and meaning recall and recognition	24 hour posttest of form and meaning recall and recognition	1 week posttest of form and meaning recall and recognition (scheduled at any time of day)
evening	12-hour posttest of form and meaning recall and recognition		
	Day 1 for p.m. training condition	Day 2 for p.m. training condition	Day 7 for p.m. training condition
morning		12-hour posttest of form and meaning recall and recognition	1 week posttest of form and meaning recall and recognition (scheduled at any time of day)
evening	Trained on new word forms and their meanings Immediate posttest of form and meaning recall and recognition	24 hour posttest of form and meaning recall and recognition	

Figure 1.
The experimental schedule.

Figure 2.
Recognition of word referents, word forms, and their links immediately after training.

Figure 3. Number of words recalled and number of semantic responses provided immediately after training.

Figure 4. Change in memory for word forms over time as measured by the free recall task. Asterisks indicate significant change over the immediate posttest with $p < .05$.

Figure 5. Change in memory for word meanings over time as measured by the word association task. Asterisks indicate significant change over the immediate posttest with $p < .05$.

Table 1

Participants' demographics and test scores by group.

Descriptive variable	LI		ND		$t_{(df=67)}$	p	d
	M	SD	M	SD			
Age in years	21.6	1.72	21.2	2.2	<1	.34	
Education in years	14.4	1.88	14.6	1.89	<1	.59	
TOAL4 written composite percentile	12	5.16	63	4.61	-12.60	<0001	-3.11
TOAL4 oral composite percentile	33	8.32	54	6.30	-12.12	<0001	-2.90
PPVT-IV standard score	99	11.69	114	10.39	-5.92	<0001	-1.42
EVT standard score	96	16.08	120	9.56	-7.72	<0001	-1.83
Modified Token Test raw score	38	4.52	43	3.10	-5.96	<0001	-1.45
K-BIT standard score	102	12.07	110	12.82	-2.68	.009	-.65

TOAL4 = Test of Adolescent and Adult Language-fourth edition; PPVT-IV = Peabody Picture Vocabulary Test-fourth edition; EVT = Expressive Vocabulary Test; K-BIT = Kaufman Brief Intelligence Test.

Appendix A

Word stimuli.

Set A			Set B		
English Word	Trained Neighbor	Untrained Neighbor	English Word	Trained Neighbor	Untrained Neighbor
army	ˈarmʊ	ˈarmu	basket	ˈbæskəl	ˈbæskəm
bucket	ˈbʌkɪdʒ	ˈbʌkəv	butter	ˈbʌtəp	ˈbʌtæg
finger	ˈfɪŋgəp	ˈfɪŋgəs	captain	ˈkæptɪdʒ	ˈkæptɪk
garage	ˈgərɒk	ˈgərɒm	pumpkin	ˈpʌmpkət	ˈpʌmpkəs
honey	ˈhʌnə	ˈhʌnoʊ	music	ˈmjuzɪb	ˈmjuzən
cabin	ˈkæbɪb	ˈkæbɪf	ocean	ˈoʊʃək	ˈoʊʃəl
winter	ˈwɪntæg	ˈwɪntən	oven	ˈʌvəd	ˈʌvək
closet	ˈklɒzəm	ˈklɒzæg	pocket	ˈpɒkəm	ˈpɒkər
office	ˈɒfɪd	ˈɒfɪk	partidge	ˈpɑrtɪp	ˈpɑtrəm
lettuce	ˈlɛtəv	ˈlɛtəl	kitchen	ˈkɪtʃət	ˈkɪtʃəf
machine	məˈʃɪg	məˈʃɪz	river	ˈrɪvəd	ˈrɪvəs
movie	ˈmuvu	ˈmuvoʊ	stomach	ˈstʌməs	ˈstʌməb
lady	ˈleɪdə	ˈleɪdaɪ	ticket	ˈtɪkəm	ˈtɪkəf
pillow	ˈpɪlu	ˈpɪli	tennis	ˈtɛnɪb	ˈtɛnɪʃ
puzzle	ˈpʌzəm	ˈpʌzəv	cartoon	ˈkɑrtuk	ˈkɑrtus
city	ˈsɪtʊ	ˈsɪtə	woman	ˈwʊməl	ˈwʊmət

Appendix B

Referent Stimuli and Training Script.

Set A			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
pony			Armo An armo is a type of pony. The armo ate hay.
turtle			Buckedge A buckedge is a type of turtle. The buckedge moved slowly.
sheriff			Fingep A fingep is a type of sheriff. The fingep arrested the criminal.
phone			Garak A garak is a type of phone. The garak rang incessantly.
penguin			Huna A huna is a type of penguin. The huna swam in the icy water.

Set A			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
chicken			<p>Kabib A kabib is a type of chicken. The kabib ran from the farmer.</p>
angel			<p>Winteg A winteg is a type of angel. The winteg flew through the clouds</p>
wagon			<p>Klazem A klazem is a type of wagon. The klazem was hard to pull</p>
ladder			<p>Ofid An ofid is a type of ladder. The ofid was dangerous to climb.</p>
banana			<p>Letev A letev is a type of banana. The letev was not yet ripe.</p>
planet			<p>Mashig A mashig is a type of planet. The mashig was invaded by aliens.</p>
robot			<p>Muvu A muvu is a type of robot. The muvu was state-of-the-art.</p>

Set A			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
bicycle			Leida A leida is a type of bicycle. The leida had ten speeds.
castle			Pilu A pilu is a type of castle. The pilu was drafty in the winter
clown			Puzum A puzum is a type of clown. The puzum was a bit freaky.
elephant			Sito A sito is a type of elephant. The sito charged the jeep.
Set B			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
dragon			Baskel A baskel is a type of dragon. The baskel frightened the prince
cookie.			Buttep A buttep is a type of cookie. The buttep was homemade

Set A			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
monkey			Kaptidge A kaptidge is a type of monkey. The kaptidge climbed a tree.
cat			Pumpkit A pumpkit is a type of cat. The pumpkit purred loudly.
baby			Musib A musib is a type of baby. The musib napped every afternoon.
dog			Oshik An oshik is a type of dog. The oshik barked at the moon.
rabbit			Oved An oved is a type of rabbit. The oved hopped through the garden.
spider			Pockem A pockem is a type of spider. The pockem gave the girl the creeps
monster			Partrip A partrip is a type of monster. The partrip was kinder than he looked

Set A			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
pizza			Kitchet A kitchet is a type of pizza. The kitchet is a date-night favorite.
paddle			Rived A rived is a type of paddle. The rived fell off the canoe.
carrot			Stomas A stomas is a type of carrot. The stomas was not ripe enough to eat
rocket			Tickem A tickem is a type of rocket. The tickem left for outer space.
flower			Tennib A tennib is a type of flower. The tennib blooms only in the early spring.
ambulance			Cartook A cartook is a type of ambulance. The cartook transported the patient.

Set A			
English Word	Target Semantic Neighbor	Untrained Semantic Neighbor (Foil)	Training Script
hammer			<p>Womal A womal is a type of hammer. The womal was the wrong size for the carpenter's hand</p>