Research Article

Distributional Learning in College Students With Developmental Language Disorder

Jessica Hall,^a Amanda Owen Van Horne,^a Karla K. McGregor,^a and Thomas Farmer^a

Purpose: This study examined whether college students with developmental language disorder (DLD) could use distributional information in an artificial language to learn about grammatical category membership in a way similar to their typically developing (TD) peers.

Method: Seventeen college students with DLD and 17 TD college students participated in this task. We used an artificial grammar in which certain combinations of words never occurred during training. At test, participants had to use knowledge of category membership to determine

which combinations were allowable in the grammar, even though they had not been heard.

Results: College students with DLD performed similarly to TD peers in distinguishing grammatical from ungrammatical combinations.

Conclusion: Differences in ratings between grammatical and ungrammatical items in this task suggest that college students with DLD can form grammatical categories from novel input and more broadly use distributional information.

Language contains a vast amount of statistical regularities on multiple levels, and so it is perhaps unsurprising that a number of recent studies have shown a relationship between language ability and the ability to perform statistical learning tasks (Kidd, 2012; Kidd & Arciuli, 2016; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010). Language is composed of categories (phonemes, words) that are organized into hierarchical sequences, which work together to communicate meaning. Statistical learning tasks often allow us to better understand how learners acquire sequential information. Tasks that test acquisition of adjacent and nonadjacent sequential dependencies are often considered analogous to common language learning problems. Attending to the co-occurrences of phonemes and syllables in an artificial language is considered analogous to the task of word segmentation (e.g., Saffran, Aslin, & Newport, 1996). Similarly, learning that A predicts B in AXB paradigms is considered analogous to learning English morphological sequences,

such as *is-verb-ing* (e.g., Gómez, 2002). These tasks have been explored extensively in typical adult and child learners (for a good review, see Aslin & Newport, 2012) and have also been extended to study individuals with developmental language disorder (DLD; e.g., Plante, Gómez, & Gerken, 2002). Findings have generally shown that both adjacent and nonadjacent dependencies are robustly available to typical learners (in infancy: Gómez & Maye, 2005; Saffran, Aslin, et al., 1996; in adults: Newport & Aslin, 2004; Saffran, Newport, & Aslin, 1996). Individuals with DLD are less proficient in learning statistical sequences, suggesting a potential mechanism for the observed deficits in language learning (Hsu & Bishop, 2010; but see studies by Aguilar & Plante, 2014; Bishop & Hsu, 2015, which show preserved learning in nonverbal statistical learning tasks).

However, the literature on how typical learners or individuals with DLD form grammatical categories via statistical information is much sparser. Grammatical categories may be formed via analysis of distribution information, which requires linking the sequential information across utterances to form relevant categories (Abbot-Smith & Tomasello, 2006; Childers & Tomasello, 2001; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Pine, & Baldwin, 1997). Corpus and modeling studies have demonstrated that the consistency of these types of distributional cues makes them useful for identifying categories in English and Dutch (Mintz, Newport, & Bever, 2002; St. Clair, Monaghan, & Christiansen, 2010; Van Heugten & Johnson, 2010). One might predict, then, that difficulty using distributional information could lead to delayed or deficient

^aThe University of Iowa, Iowa City

Amanda Owen Van Horne is now at the University of Delaware, Newark
Karla K. McGregor is now at Boys Town National Research Hospital, Omaha, Nebraska

Thomas Farmer is now at the University of California, Davis

Correspondence to Jessica Hall: jessica-e-hall@uiowa.edu

Editor-in-Chief: Sean Redmond

Editor: Lisa Archibald

Received January 11, 2017

Revision received May 1, 2017

Accepted July 11, 2017

https://doi.org/10.1044/2017_JSLHR-L-17-0013

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

category learning, causing problems in language acquisition. The profile of individuals with DLD suggests that distributional learning may lead to some of the difficulties they face; in this study, we test this hypothesis. In what follows, we briefly review the evidence on statistical learning of sequences in DLD and why the DLD profile might lead one to expect deficits in distributional learning and then discuss how distributional learning differs from the traditional sequential tasks and a task to test distributional learning.

Statistical Learning and Individuals With DLD: Implications and Evidence

Studies with individuals with DLD have shown challenges in learning adjacent and nonadjacent dependencies (see, e.g., a meta-analysis by Obeid, Brooks, Powers, Gillespie-Lynch, & Lum, 2016), and these findings have been used to argue that they have difficulty learning and using language because of an underlying domain-general procedural learning deficit. Individuals with DLD (also known as specific language impairment) are most commonly defined as a group of people who have a developmental, rather than acquired, language learning impairment in the absence of an obvious cause, such as hearing impairment, autism, or cognitive disability (Bishop, Snowling, Thompson, & Greenhalgh, 2017), though those with concomitant disorders such as attention-deficit/hyperactivity disorder and dyslexia are commonly included (but see Redmond, 2005). Table 1 summarizes the extant evidence for impaired statistical learning in individuals with DLD.

Serial reaction time tasks, in which successful participants show evidence of speeded performance for trained versus untrained sequences, provide the most robust evidence of a deficit in statistical learning being linked to grammatical proficiency. A meta-analysis by Lum, Conti-Ramsden, Morgan, and Ullman (2014) found a medium effect size for deficits in serial reaction time tasks in individuals with DLD as compared to typical peers, with differences between studies explained by effects of age and length of the task. In two studies, grammatical proficiency was a better predictor of task performance than

DLD diagnosis or vocabulary skill (Hedenius et al., 2011; Tomblin, Mainela-Arnold, & Zhang, 2007).

Artificial grammar learning studies differ from serial reaction time tasks in that they ask participants to explicitly judge whether a sequence is consistent with a “grammar” the individual was previously exposed to. In artificial grammar learning studies, results show difficulty with linguistic stimuli: Children and adults with DLD perform worse than typical peers in adjacent dependency-based tasks (Evans, Saffran, & Robe-Torres, 2009; Plante et al., 2002) and nonadjacent dependency-based tasks (Grunow, Spaulding, Gómez, & Plante, 2006; Hsu, Tomblin, & Christiansen, 2014). Deficits in sequential or adjacent dependency learning suggest a very broad deficit and imply low-level encoding difficulty, such as difficulty in distinguishing the order of phonemes or morphemes heard within a word, and would be consistent with the observed slower vocabulary development (Gray, 2004; McGregor, Bahnsen, Oleson, & Duff, 2013; Riches, Tomasello, & Conti-Ramsden, 2005). Difficulty learning nonadjacent dependencies may be related to difficulty observed in DLD in producing correct grammatical morphemes (e.g., third-person /s/ predicted by a third-person pronoun, such as “she” as in “she sits/eats/walks”; Bedore & Leonard, 1998; Rice & Wexler, 1996).

However, studies that employ different types of statistics or tasks or that tap different modalities show preserved statistical learning ability. Hsu and Bishop (2014) found similar performance by children with and without DLD in a procedural learning task that did not involve sequence learning. A recent study by Clark and Lum (2017) showed similar learning in children with DLD and typical language for higher-order conditional sequences, in which A predicts B, C, or D with equal likelihood, though there were differences in the proficiency with first-order conditional sequences, in which A predicts those items differentially. In the Weather Prediction Task (Knowlton, Squire, & Gluck, 1994), a task that also requires attention to conditional sequences for success, older children with DLD (mean age 11;3 [years;months]) were less skilled than their typical peers (Kemény & Lukács, 2010), but a group of younger children with DLD (mean age 9;4) did not show these differences

Table 1. Findings from studies of statistical learning ability in individuals with developmental language disorder.

Study type	Example	Findings
Serial reaction time	Repeat sequence with response box (1234)	Deficits in children (effect size = .328 in meta-analysis by Lum et al., 2014)
Paired associates	<i>Dugong</i> + picture of dugong	Deficits in children on verbal task but not nonverbal task (Bishop & Hsu, 2015)
Adjacent dependencies	AB	Deficits in children (Evans et al., 2009) and adults (Plante et al., 2002)
Nonadjacent dependencies	AxB	Deficits for adults (Grunow et al., 2006); NS differences in visual task with adults (Aguilar & Plante, 2014); deficits in adolescents (Hsu et al., 2014)
Probabilistic learning	AB AC	Deficits for children mean age 11 (Kemény & Lukács, 2010); no deficits for children with mean age of 9 (Lukács & Kemény, 2014)
Higher-order vs. first-order conditional sequences	Higher order: AB = AC First order: AB > AC	Deficits for first-order but not higher-order conditionals for children (Clark & Lum, 2017)

Note. NS = nonsignificant.

(Lukács & Kemény, 2014). In a nonadjacent dependency-based artificial grammar learning task with visual rather than auditory stimuli, adults with DLD were not significantly worse than the group without impairment (Aguilar & Plante, 2014).

Given these discrepant findings, it is important to consider that (a) different types of statistical learning tasks may depend on different learning mechanisms and (b) developmental effects may be at play when determining strengths and weaknesses in statistical learning in the DLD population. The evidence for a statistical learning deficit in DLD thus far is based entirely on sequential statistics, that is, that a particular item predicts another item. It is thus possible that individuals with DLD show a different profile in their use of distributional learning, as it is not strictly derived from linear sequential information, than is observed in sequential learning.

Distributional Learning and Categorical Information in Language

Distributional learning is based on the ability to notice patterns across different sequential dependencies. For example, distributional information tells us that words that occur after pronouns also commonly occur with the suffix -s. Recognition of these patterns results in the formation of categories—in this case, the category *verb*—and allows for generalizations to be made. On a more fine-grained level, observing that, within this category, certain verbs occur in nonoverlapping contexts (e.g., with and without direct objects) could result in more richly defined subcategories of verbs.

Previous statistical learning studies demonstrated that infants and adults can use distributional information in combination with a salient clue such as morphology or absolute position to learn a novel grammar (Frigo & McDonald, 1998; Gerken, Wilson, & Lewis, 2005; Smith, 1966, 1969). Novel verb-learning studies also demonstrate that children and adults generalize to new contexts when other verbs with similar distributional and semantic properties appear in those contexts (Ambridge & Lieven, 2011; Ambridge, Pine, & Rowland, 2011, 2012). Only recently has it been shown that adults (Reeder, Newport, & Aslin, 2013) and children (Hall, Van Horne, & Farmer, 2017) can form categories from distributional information in the absence of other cues. This type of learning has not yet been tested in individuals with DLD.

Early work by Leonard (1995) clearly demonstrated that individuals with DLD have access to the full range of grammatical categories available to typical language learners. And yet, this does not mean that this domain is free of deficits. Children with DLD produce grammatical morphology at a delayed rate and with a greater dependence on the input in comparison to typically developing (TD) peers for both natural and novel input (Conti-Ramsden & Jones, 1997; Leroy, Parris, & Maillart, 2014; Skipp, Windfuhr, & Conti-Ramsden, 2002). They omit arguments more often and use fewer argument structures than same-age peers (Ebbels, 2005; Grela & Leonard, 1997; King, 2000; King & Fletcher, 1993; Sanz-Torrent, Andreu, Badia, &

Sidera, 2011; Thordardottir & Ellis-Weismer, 2002). They use “general all purpose” verbs at a very high rate, evidence of weak verb learning (Conti-Ramsden & Jones, 1997; Rice & Bode, 1993). Even at older ages (11–15 years), they show errors in the correct use of a verb’s subcategorization frame, using a verb such as “fill” in an utterance where verbs such as “pour” are found, for example, “the lady is filling the sweets into the jar” (Ebbels, 2005). These studies suggest that, at least at certain levels, category learning may be problematic in DLD. An artificial grammar learning task would allow us to examine this type of learning within a short window of time in a controlled experimental setting.

Task That Examines Distributional Learning

The artificial grammar learning task in Reeder et al. (2013) provides an excellent way to test our prediction, as it demonstrated that typical adults can use distributional information alone to form categories and to restrict the way those categories are deployed. Participants were exposed to a speech stream and then asked to judge the grammaticality of familiar and untrained sequences. In the first two experiments of their study, participants could use adjacent dependencies within sequences to determine that a test item was grammatical or ungrammatical. Critically, in the third experiment, success was only possible if the participants induced categories based on distributional rather than sequential information. Novel test items included sequences that contained zero adjacent or nonadjacent dependencies but were grammatical because of the similar distributions of the component words. Participants rated novel grammatical items lower than familiar grammatical items, which they did not do in the first two experiments, but they also rated novel grammatical items as better than the ungrammatical items, thus demonstrating that distributional information affected how they formed categories. Two additional experiments further documented how the amount of exposure to grammatical sequences and gaps influenced how learners extended or constrained grammatical categories within this artificial language. Experiment 3 provided the most realistic task setting, because in reality language learners do not hear every possible combination of words. Instead they must generalize, which is possible through attention to words’ distributions.

Questions and Predictions

This work by Reeder and colleagues (2013) provided inspiration for a series of studies that we have undertaken. An adaptation of Experiment 3 by Hall, Van Horne, and Farmer (2017) showed that TD children could also learn grammatical categories through distributional information and thus is a viable mechanism for language acquisition. In this study, we aim to determine whether college students with DLD differ from same-age peers in their ability to use distributional information to form grammatical categories in an artificial grammar learning task adapted from Experiment 3 of Reeder et al. (2013). Namely, we test whether

they discriminate novel grammatical items from ungrammatical items to a similar degree. Second, we are interested in individual differences in distributional learning ability and how they contribute to success in the task within and across groups. We examined grammatical proficiency given previous links between grammar and statistical learning ability (Hedenius et al., 2011; Tomblin et al., 2007). We hypothesize that if college students with DLD use distributional information differently, they will show difficulty forming grammatical categories within the experiment. Furthermore, we hypothesize that differences in grammatical ability, specifically, predict differences in the ability to form generalizations.

Method

Participants

We recruited 34 participants, ages 18–24 years, 17 of whom were identified as having DLD according to a test battery developed by Fidler, Plante, and Vance (2011). The participants with DLD responded to an advertisement requesting volunteers who “struggle with spoken or written language.” Although there are no agreed-upon criteria for identifying adults with DLD, the Fidler et al. (2011) battery was selected because of its good sensitivity and specificity (80% and 87%, respectively; see Table 6 of the original study) in identifying childhood DLD and included a modified version of the Token Task (Morice & McNicol, 1985) and a spelling task. It is well documented that deficits in older children with DLD are more prominent in writing than oral measures (Dockrell & Connelly, 2009; Mackie & Dockrell, 2004; Windsor, Scott, & Street, 2000), and so it is not surprising that a spelling measure is useful for identification. A weighted composite measure of these two language tasks was used to verify group membership. Per the procedure documented in Fidler et al. (2011), all participants with DLD scored in the positive range and all TD participants scored in the negative range. Participant groups were matched by age, gender, and college enrollment, all attending either a midsize Midwestern university or a community college. All participants had normal hearing, scored above 85 on the Kaufman Brief Intelligence Test–Second Edition, Matrices subtest (Kaufman & Kaufman, 2004), and had no history of acquired neurological disorders. Two of the participants in the DLD group reported a diagnosis of language impairment, seven reported reading impairment, two reported both language and reading impairments, three reported writing impairments, one reported attention-deficit disorder, and two reported no diagnoses. None of the TD participants reported any history of these or other diagnoses relevant to DLD. Regarding diagnoses, we did not base group membership on previous diagnosis because of the noted underidentification (Tomblin et al., 1997), and we did not exclude students with attention-deficit/hyperactivity disorder or dyslexia, as these conditions are often comorbid with DLD (Catts, Adlof, Hogan, & Weismer, 2005; Mueller & Tomblin, 2012), and individuals with DLD often receive a variety of labels (Reilly,

Bishop, & Tomblin, 2014). All participants also completed the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007) to document their vocabulary skills and a nonstandardized nonword repetition task (Dollaghan & Campbell, 1998) as a measure of phonological working memory. Groups differed on all standardized assessments, which are summarized in Table 2. All participants also participated in a study of semantic fluency and two other studies focused on vocabulary learning (Hall, McGregor, & Oleson, 2017; McGregor, Arbisi-Kelm, & Eden, 2017; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017), and results supported common findings that the DLD group performs more poorly on language-learning tasks than the group with typical language. See McGregor, Gordon, et al. (2017) for a more thorough discussion of participants in this study.

Stimuli

The grammar of the artificial language used in this study was adapted from Reeder et al. (2013). The language included five categories of words, which appeared in sentences in the order of (Q)AXB(R), in which Q and R words were optional. Each category contained two or three words. The A, X, and B words are shown in Figure 1, and all training and test items are listed in the Appendix. Stimuli construction was inspired by Reeder et al. (2013), with adaptations made so that the task could be used with children. We summarize differences at the end of this Method section.

Training

Participants heard 12 of 27 possible combinations of AXB. Within the training sentences, only a subset of As appeared with each X, and a subset of Xs appeared with each B. Figure 1 gives an illustration of lexical dependencies for the training, with absent dependencies indicated

Table 2. Participant demographic and test means and standard deviations by diagnostic group (DLD and TD).

Demographic/ test	DLD	TD	<i>p</i>	<i>d</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		
Age	20.6 (1.1)	21.1 (1.8)	.42	–0.28
Education (years)	13.6 (1.1)	14.0 (1.7)	.50	–0.24
Token	35.4 (5.1)	40.1 (3.0)	< .01	–1.12
Spelling	4.5 (3.0)	11.8 (2.4)	< .01	–2.69
PPVT-4	199.4 (11.2)	206.5 (6.7)	.03	–0.77
KBIT-2	98.6 (10.7)	110.1 (13.0)	< .01	–0.97
NWR	85.7 (5.1)	90.6 (3.4)	< .01	–1.13

Note. Scores on the Token Task and spelling task are raw counts of items correct out of 44 and 15, respectively. Scores on the Peabody Pictures Vocabulary Test–Fourth Edition (PPVT-4) are raw scores. Scores on the Kaufman Brief Intelligence Test–Second Edition, nonverbal subtest (KBIT-2) are standard scores with a normative mean of 100 and a standard deviation of 15. Scores on nonword repetition (NWR) are a percentage of phonemes correct out of 96. DLD = developmental language disorder; TD = typically developing.

Figure 1. Linear dependencies during training.

by dashes. During training, participants heard each of the 12 AXB combinations in three different sentence types, and this corpus of 36 sentences was heard three times for a total of 108 trials. The training generally took 10–15 min and was sufficient for 7-year-old children to learn the grammar (Hall, Van Horne, & Farmer, 2017).

Participants completed a one-back task during training, using two keyboard buttons to indicate if the current sentence was the same as in the trial immediately prior, which was true for 35 trials. Data from individuals who scored less than 60% accuracy on the one-back task would be discarded, but all participants performed above this threshold. The one-back task provided a measure of attention during training.

Test

In the test phase, participants heard 36 AXB sentences of three types: (a) nine “grammatical” sentences heard during training; (b) nine novel “grammatical” sentences, that is, sentences that were not heard during training but were consistent with the “grammar” of the artificial language (these dependencies are indicated by dashed lines in Figure 1); and (c) 18 “ungrammatical” sentences. Familiar and novel sentences were each heard twice, whereas ungrammatical sentences were heard only once to avoid familiarization, for a total of 54 trials. Bigram frequencies for each item were calculated, and the average associative chunk strength, as defined in Ellefson and Christiansen (2000), was not different, $t(1,219.71) = -0.59$, $p = .55$, across novel ($M = 4.83$) and ungrammatical ($M = 4.94$) items. Participants rated test items by pressing one of two buttons to indicate whether a sentence was “something the alien would say” and then moved a slider scale to indicate confidence (see Ambridge, Pine, Rowland, & Young, 2008). The scale provided increased power to detect variability in learning. Participants performed an apple/pear discrimination task that had the same button choice and scale to familiarize them with test response measures before beginning the experiment.

Procedure

E-Prime (Schneider & Zuccoloto, 2007) was used to deliver all three parts and perform data collection. During

training, participants watched videos of cartoon aliens saying the sentences on a computer. Instructions only directed participants on how to perform the one-back task. Each training trial consisted of a video of an alien “speaking” a sentence. Participants performed the one-back task on every trial, except the first: They were to press the red button if an alien said the same thing as the alien before it and to otherwise press the green button (stickers over keyboard buttons D and K). Four-second video clips that included depictions of aliens’ attempts to fix their ship also played intermittently between trials to keep participants’ interest.

Immediately following training, participants performed the test. Each test trial consisted of an audio recording of an alien speaking the sentence. Participants were instructed to listen to the sentence and determine if it was something the aliens would say. Examiners reminded participants of the earlier apple/pear task and told them to press the red button to indicate a negative response and green for a positive response, and then to place the marker on the slider scale to indicate the certainty of their choice. The scale was colored red and green to correspond with the colors of the buttons.

Changes to the original Reeder task were as follows. Stimuli were recorded in child-directed tone and pitch by a female adult not familiar with the study. The training phase was presented in video format with a one-back task rather than an audio stream to make it more engaging and to have a measure of attention for child participants. We used one version of the grammar since Reeder et al. (2013) found no differences for the version used. Reeder et al. (2013) included nine AXB combinations in the training for Experiment 3. The previous study tested only AXA and BXB ungrammatical sentences; in our experiment, we included these types as well as four other combinations, AXR, BXA, QXB, and XBA, to exclude the possibility that lower ratings for ungrammatical items were due to hearing two items of the same category in one sentence and to determine whether participants could discriminate based on co-occurrence information. We used three instances of each sequence, a slightly shorter exposure than in Reeder et al. (2013), in which participants heard sentences four times each. Reeder et al. (2013) used a 5-point Likert scale only at test, whereas we employed the button press and slider scale responses.

Analysis

We used the lme4 package (Bates, Maechler, Bolker, & Walker, 2016) and the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016) in R version 3.1.3 (R Core Team, 2015) to run linear mixed effects models to explore several comparisons of interest. To answer our first question of whether college students with DLD can use distributional information similarly to typical peers to form grammatical categories, we ran a model with scale rating as the dependent variable and item type (familiar, novel, ungrammatical), group membership (DLD, TD), and their interaction as main effects. Responses for novel versus ungrammatical items is the critical comparison, as it is evidence that participants formed grammatical categories. We also compared familiar to ungrammatical and familiar to novel to determine whether groups differed in their memory for trained items. Ungrammatical and TD served as the reference categories unless otherwise specified. To simplify findings, we report only scale rating, as we verified that the binary measure showed similar patterns of results through a model with the binary measure as a predictor and the visual analog scale as the dependent variable. The binary measure significantly predicted the analog scale rating, $t(31.84) = 14.68, p < .0001$, with no effect of group, $t(30.5) = -0.38, p = .71$, nor an interaction, $t(31.7) = 0.59, p = .56$, indicating that both groups utilized the scale equally well. We converted scale ratings to z scores, as in Reeder et al. (2013), to control for variable use of the scale by participants. Z -score scale ratings are the dependent variable for all analyses.

To address concerns that ungrammatical test items may have interfered with responses to later grammatical items, we ran a second model that included the order in which familiar and novel items were heard, first or second, with first serving as the reference category.

To test our second question about the role of individual differences, we removed group from the model and tested the contributions of three covariates: (a) accuracy on the one-back test during training as a measure of attention, (b) performance on the nonword repetition task as a measure of phonological working memory, and (c) performance on the modified Token Task as a measure of grammatical proficiency. These variables were centered to minimize multicollinearity. We used an individual differences approach rather than dividing the groups into grammatically impaired/not impaired to increase power.

Results

No participant scored below 60% on the one-back task during training, and so data from all participants were included. Groups did not significantly differ from each other in mean one-back accuracy (DLD, $M = 92.7\%$, $SE = 0.02$; TD, $M = 94.5\%$, $SE = 0.02, p = .49$).

First, we considered the question of whether individuals with DLD and TD individuals could utilize distributional information similarly. Random intercepts for subject

and item and a random subject slope for item type were the maximal random effects structure supported by the data. Log likelihood ratios and Akaike information criterion (AIC) comparison showed that only the random intercept for item was needed, with subject variance likely accounted for by standardization of the scale. We found evidence of learning: All participants distinguished novel items from ungrammatical, $t(67.74) = 2.20, p = .03$, and familiar items from ungrammatical, $t(68.54) = 3.58, p < .001$. Estimated effect size for novel versus ungrammatical is calculated by finding the correlation between the fitted and observed values of the model (Xu, 2003), $\Omega^2 = .27$. This indicated that both groups of participants were able to distinguish between grammatical and ungrammatical strings. A stricter test would be to find that the familiar and novel items did not differ because this would indicate that participants were not simply memorizing strings or parts of strings they had heard during training. To compare ratings for familiar and novel items, we reset the reference variable to familiar and found that differences were not significant, $p = .12$, suggesting that participants' performance reflects generalization beyond just memorization of specific forms heard during training. In answer to our primary question of whether group performance differed, we found no main effect of group, $p = .19$, but the interaction between diagnostic group and the difference between novel and ungrammatical ratings was marginal, $t(1,779) = -1.81, p = .07$. To further explore this difference, we conducted single sample t tests of ratings by item type, which confirmed that mean ratings for the DLD group for all item types were different from zero (familiar and novel items were above zero and ungrammatical items were below zero), indicating that individuals were actively considering the novel items as acceptable. For the TD group, only ratings for familiar and ungrammatical items were different from zero. This suggests that, as a group, the TD participants were ambivalent about how to rate novel items. Figure 2 illustrates mean ratings by item type for each group. Fourteen of 17 TD participants and all 17 participants with DLD followed a pattern of higher ratings for novel grammatical items than ungrammatical.

The order in which items were heard was important to consider because exposure to ungrammatical items during the test could have degraded performance on all items by muddying the representation of the grammar. We added the order items were heard to the model. A main effect of order heard showed that grammatical items heard the second time were rated lower than they were the first time, $t(50.0) = -3.3909, p = .0013$. Figure 3 displays the effects of item type and order on ratings, and model results are reported in Table 3.¹

¹We performed a follow-up analysis using Bayesian methods and the glmer2stan package (McElreath, 2013) and rstan package (Stan Development Team, 2016) in R version 3.4.3 (R Core Team, 2017). The results from this analysis supported those from the linear mixed effects model, with significant effects of item type and order and null findings for group and any interaction with group, and thus, more detailed results are not reported.

Figure 2. Boxplots demonstrating mean standardized scale rating for each participant by diagnostic group (DLD = developmental language disorder; TD = typically developing) for each item type (familiar, novel, and ungrammatical).

Figure 3. Ratings by item type (familiar, novel, and ungrammatical) and the order items were heard by diagnostic group (DLD = developmental language disorder; TD = typically developing).

Table 3. Results of the mixed effects model showing the influence of item type, diagnostic group, and order on ratings.

Factor	Variance	SD	β	SE	p
Random factors					
Item	0.15	0.38			
Fixed factors					
(Intercept)			-.40	0.10	< .01
Group (reference category = TD)			-.09	0.07	.19
Item type (reference category = ungrammatical)					
Familiar			.91	0.16	< .01
Novel			.76	0.16	< .01
Group $\times$ Item Type					
DLD $\times$ Familiar			.10	0.10	.33
DLD $\times$ Novel			.18	0.10	.07
Order			-.46	0.14	< .01

Note. TD = typically developing; DLD = developmental language disorder.

Finally, to answer our question of how individual differences contributed to statistical learning performance, we added one-back accuracy during training, nonword repetition performance, performance on the Token Task, and each variable's interaction with item type. We entered variables individually, as well as all at once, per the recommendation to use the maximal structure in Barr, Levy, Scheepers, and Tily (2013). Log likelihood ratio tests and AIC comparison again indicated that only a random item intercept was needed. None of the variables or interactions significantly predicted ratings, no matter how they were entered into the model, all $ps > .1$. This finding held whether or not order was included in the model.

Discussion

Participants with DLD performed similarly to typical participants in an adaptation of Experiment 3 from Reeder et al. (2013), a distributional learning artificial grammar task. Although we used a different scale in our outcome measure, our results follow the pattern from Reeder et al. (2013) in that participants were able to distinguish novel grammatical sentences from ungrammatical sentences. From this, we can conclude that participants in the DLD group formed grammatical categories similarly to typical peers. Findings from the follow-up analysis suggested that measures of attention, phonological working memory, and grammatical proficiency were not predictive of statistical learning ability.

The results of this study suggest intact procedural learning in the DLD population, at least in young adulthood. On the face of it, these results would appear to be in conflict with a procedural learning deficit account of DLD. Certainly, interpretation would be more straightforward if clear group differences in performance or correlations with language ability were observed. That said, this finding is not out of line with a theory of procedural learning deficit in DLD if (a) that deficit is strictly related to sequential statistics; (b) as a developmental disorder, the profile changes with age; or (c) our measures of language

were not sensitive enough to meaningfully correlate with this task. It is possible that the nonsequential aspects of the task supported participants' success. Hsu and Bishop (2014) also found evidence of preserved nonsequential learning in children with DLD. This is not altogether clear, however, because it seems that attention to sequential statistics (i.e., tracking which bigrams occur) would drive the nonsequential learning (A words appear before X words and thus represent a category that is distinct from words that appear in other positions). Future studies that compare sequential and nonsequential learning within the same group could clarify this finding. Although we expected a deficit because of childhood difficulties related to grammatical category formation observed in DLD (Conti-Ramsden & Jones, 1997; Leroy et al., 2014; Skipp et al., 2002), adults in this population have not been reported as having deficits in the area of grammatical categories or in the use of subcategorization frames, though studies of adolescents and adults show continued difficulty with tense and agreement (Wetherell, Botting, & Conti-Ramsden, 2007) and complex language (Nippold, Mansfield, Billow, & Tomblin, 2008, 2009). Finally, the Token Task was chosen primarily for group identification and does not directly test grammatical category knowledge or proficiency.

Although it is unusual to observe equivalence between a DLD and typical language group in language-based statistical learning tasks, other reports of successful statistical learning among people with DLD corroborate our findings (Augilar & Plante, 2014; Gabriel, Maillart, Guillaume, Stefaniak, & Meulemans, 2011; Gabriel, Stefaniak, Maillart, Schmitz, & Meulemans, 2012). It is unlikely that the good outcome reflected low task demand. Individual performance on the task in this study varied, even within the typical population, suggesting that neither ceiling nor floor performance explains the results (see Figure 2 for illustration of spread of performance). It is possible that if we shortened or lengthened exposure, differences across groups may have surfaced. Changes in exposure duration can have consequences for category formation via distributional learning, as shown in Experiments 4 and 5 in Reeder et al. (2013).

Future work might examine whether group differences might be observed via different manipulations, such as distributions that induce the existence of subcategories in typical learners.

It is worth considering whether equivalent task performance from the DLD and TD groups could reflect alternative means of solving the task. Memorization is one possibility. However, novel and ungrammatical items had similar associative chunk strength, and given the distinction in ratings for these item types, it is unlikely that participants in either group were using memory of individual items to perform the task. Furthermore, nonword repetition was not a significant predictor of performance in our model, which is evidence that working memory was not a significant factor for either group. We cannot rule out the possibility that groups relied on different cognitive resources to perform the task, which could explain null findings for a relationship with grammar ability, but the number of participants and test items prevented a complicated multifactorial model in which we could measure a large number of influential forces. That the interaction between group and item type neared significance and that the TD group's standardized rating of novel items did not differ from zero both suggest TD participants were more conservative in their generalization, perhaps, than the DLD group, whose ratings for novel items were significantly above zero. Because none of our measures of individual differences showed an interaction with item type, we cannot say specifically what might attribute to this pattern.

Regarding individual differences, the sample chosen could have influenced results. We did not find support for a specific relationship between grammatical ability and statistical learning, in contrast to earlier findings (Hedenius et al., 2011; Tomblin et al., 2007). The sample consisted of college students, and we know from longitudinal studies that individuals with DLD are not as academically successful as their peers. College students will necessarily, then, be the higher performing individuals within the population, with language experience that is likely more skilled than those not attending college. If we had tested a group with a wider range of abilities, we might have seen differences. However, other research groups have found differences in statistical learning and vocabulary abilities in comparisons of college students with and without DLD (Grunow et al., 2006; Plante et al., 2002; Sheng, Byrd, McGregor, Zimmerman, & Bludau, 2015), as well as in this sample (Hall, McGregor, & Oleson, 2017; McGregor, Arbisi-Kelm, & Eden, 2017; McGregor, Gordon, et al., 2017). The Token Task may not have had sufficient spread to adequately show the relationship or may not have been the most sensitive test for measuring grammatical proficiency. That many in the DLD group had similar Token Task scores to those in the TD group supports this latter possibility (though it is important to note that Token Task scores did significantly distinguish groups). This work provides a rationale for collecting a larger, wider sample in a future study that employs more specified measures for detecting cognitive or language abilities that contribute to task performance.

One final phenomenon of note is the difference in ratings the first and second time they are heard for all participants. Statistical learning tasks often repeat test items to gain statistical power, but this may be at the detriment of participant learning. Participants who were most sensitive to the changes in the distributional input during exposure at test could have shown decreased ratings in the second half of the test as the distributional input changed due to the addition of ungrammatical items in the input and thus affected learning and generalizations. It would be useful if future studies examined performance across the test as well as overall performance to determine the effects of testing. Another possible explanation of order effects is that, with time, participants might have started applying overt strategies to solve the task and were thus less successful in the second half, as explicit strategies seem to interfere in other statistical learning tasks (Fletcher et al., 2005; Howard & Howard, 2001). Future studies should include implicit measures of learning to avoid interference from explicit strategies that participants may be using. Isbilen, McCauley, Kidd, and Christiansen (2016) offer an excellent implicit test for statistical learning, used with typical adults, which involves repetition of test strings rather than judgment. It remains to be seen whether this test can be used with people with DLD, given that they consistently show difficulty with nonword repetition tasks compared to peers (Estes, Evans, & Else-Quest, 2007).

Overall, participants across groups showed successful use of distributional information from novel input to form categories. Individual differences in task performance did not seem to be attributable to language abilities. Future work should examine both the developmental aspects of category formation via distributional learning and the role that manipulation of various aspects of the input plays in the success of individuals with a wide range of language abilities and experiences. Certain types of statistical learning tasks show improvement with age, such as the serial reaction time task (Lum et al., 2014; Thomas et al., 2004), though other tasks show decrement (Janacek, Fiser, & Nemeth, 2012), suggesting that it is critical to more fully understand the role that development plays in the performance of individuals with DLD.

Acknowledgments

National Institute on Deafness and Other Communication Disorders Grant 5R01DC011742 supported the efforts of Karla K. McGregor. We thank Nichole Eden, Tim Arbisi-Kelm, Danielle Reese, and Dan Plebanek for their assistance with data collection.

References

- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23(3), 275–290.
- Aguilar, J. M., & Plante, E. (2014). Learning of grammar-like visual sequences by adults with and without language-learning

- disabilities. *Journal of Speech, Language, and Hearing Research*, 57(4), 1394–1404.
- Ambridge, B., & Lieven, E. V.** (2011). *Child language acquisition: Contrasting theoretical approaches*. New York, NY: Cambridge University Press.
- Ambridge, B., Pine, J. M., & Rowland, C. F.** (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics*, 22(2), 303–323.
- Ambridge, B., Pine, J. M., & Rowland, C. F.** (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, 123(2), 260–279.
- Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R.** (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition*, 106(1), 87–129.
- Aslin, R. N., & Newport, E. L.** (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21(3), 170–176.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bates, D., Mächler, M., Bolker, B., & Walker, S.** (2016). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-12. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Bedore, L. M., & Leonard, L. B.** (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 41(5), 1185–1192.
- 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, 3.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T.** (2017). Phase 2 of CATALISE: A multinational and multi-disciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58, 1068–1080.
- Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E.** (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, 48(6), 1378–1396.
- Childers, J. B., & Tomasello, M.** (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, 37(6), 739–748.
- Clark, G. M., & Lum, J. A.** (2017). First-order and higher order sequence learning in specific language impairment. *Neuropsychology*, 31(2), 149–159.
- Conti-Ramsden, G., & Jones, M.** (1997). Verb use in specific language impairment. *Journal of Speech, Language, and Hearing Research*, 40(6), 1298–1313.
- Dockrell, J. E., & Connelly, V.** (2009). The impact of oral language skills on the production of written text. In V. Connelly, A. Barnett, J. Dockrell, & A. Tolmie (Eds.), *BJEP monograph series II: Part 6 Teaching and learning writing* (pp. 45–62). Leicester, United Kingdom: The British Psychological Society.
- Dollaghan, C., & Campbell, T. F.** (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136–1146.
- Dunn, L. M., & Dunn, D. M.** (2007). *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4)*. Minneapolis, MN: Pearson Assessments.
- Ebbels, S. H.** (2005). *Argument structure in specific language impairment: From theory to therapy* (Doctoral dissertation). University of London.
- Ellefson, M. R., & Christiansen, M. H.** (2000). The evolution of subadjacency without universal grammar: Evidence from artificial language learning. In J. L. Dessalles & L. Ghadakpour (Eds.), *The evolution of language: 3rd International Conference* (pp. 98–101). Paris, France: Ecole Nationale Supérieure des Telecommunications.
- Estes, K. G., Evans, J. L., & Else-Quest, N. M.** (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177–195.
- Evans, J. L., Saffran, J. R., & Robe-Torres, K.** (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 321–335.
- 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.
- Fletcher, P. C., Zafiris, O., Frith, C. D., Honey, R. A. E., Corlett, P. R., Zilles, K., & Fink, G. R.** (2005). On the benefits of not trying: Brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cerebral Cortex*, 15(7), 1002–1015.
- Frigo, L., & McDonald, J. L.** (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language*, 39(2), 218–245.
- Gabriel, A., Maillart, C., Guillaume, M., Stefaniak, N., & Meulemans, T.** (2011). Exploration of serial structure procedural learning in children with language impairment. *Journal of the International Neuropsychological Society*, 17(2), 336–343.
- Gabriel, A., Stefaniak, N., Maillart, C., Schmitz, X., & Meulemans, T.** (2012). Procedural visual learning in children with specific language impairment. *American Journal of Speech-Language Pathology*, 21(4), 329–341.
- Gerken, L., Wilson, R., & Lewis, W.** (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32(2), 249–268.
- Gómez, R. L.** (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436.
- Gómez, R., & Maye, J.** (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7(2), 183–206.
- 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.
- Grela, B. G., & Leonard, L. B.** (1997). The use of subject arguments by children with specific language impairment. *Clinical Linguistics & Phonetics*, 11(6), 443–453.
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E.** (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, 39(2), 158–170.
- 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(3), 640–653.
- Hall, J., Van Horne, A. J. O., & Farmer, T. A.** (2017). Distributional learning aids linguistic category formation in school-age children. *Journal of Child Language*. Advance online publication. <https://doi.org/10.1017/S0305000917000435>
- Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Verissimo, J., Dye, C. D., ... Ullman, M. T.** (2011). Grammar predicts procedural learning and consolidation deficits in children with specific

- language impairment. *Research in Developmental Disabilities*, 32(6), 2362–2375.
- Howard, D. V., & Howard, J. H.** (2001). When it does hurt to try: Adult age differences in the effects of instructions on implicit pattern learning. *Psychonomic Bulletin & Review*, 8(4), 798–805.
- Hsu, H. J., & Bishop, D. V.** (2010). Grammatical difficulties in children with specific language impairment: Is learning deficient? *Human Development*, 53(5), 264–277.
- Hsu, H. J., & Bishop, D. V.** (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental Science*, 17(3), 352–365.
- Hsu, H. J., Tomblin, J. B., & Christiansen, M. H.** (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, 5, 175.
- Isbilen, E., McCauley, S. M., Kidd, E., & Christiansen, M. H.** (2016). *Testing implicit-statistical learning implicitly*. Poster presented at the Fifth Implicit Learning Seminar, Lancaster, England.
- Janacek, K., Fiser, J., & Nemeth, D.** (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, 15(4), 496–505.
- Kaufman, A. S., & Kaufman, N. L.** (2004). *K-BIT-2: Kaufman Brief Intelligence Test—Second Edition*. Circle Pines, MN: American Guidance Service.
- Kemény, F., & Lukács, A.** (2010). Impaired procedural learning in language impairment: Results from probabilistic categorization. *Journal of Clinical and Experimental Neuropsychology*, 32(3), 249–258.
- Kidd, E.** (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, 48(1), 171–184.
- Kidd, E., & Arciuli, J.** (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, 87(1), 184–193.
- King, G.** (2000). Verb complementation in language impaired school age children. In M. Aldridge (Ed.), *Child language*. Clevedon, United Kingdom: Multilingual Matters.
- King, G., & Fletcher, P.** (1993). Grammatical problems in school-age children with specific language impairment. *Clinical Linguistics & Phonetics*, 7(4), 339–352.
- Knowlton, B. J., Squire, L. R., & Gluck, M. A.** (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, 1(2), 106–120.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B.** (2016). *lmerTest: Tests in linear mixed effects models. R package version 2.0-32*. Retrieved from <https://cran.r-project.org/web/packages/lmerTest/index.html>
- Leonard, L. B.** (1995). Functional categories in the grammars of children with specific language impairment. *Journal of Speech and Hearing Research*, 38(6), 1270–1283.
- Leroy, S., Parisse, C., & Maillart, C.** (2014). Le manque de généralisation chez les enfants dysphasiques: une étude longitudinale. *ANAE: Approche Neuropsychologique des Apprentissages chez l'Enfant*, 26(131), 357–365.
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M.** (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30(2), 333–370.
- Lieven, E. V., Pine, J. M., & Baldwin, G.** (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24(1), 187–219.
- Lukács, A., & Kemény, F.** (2014). Domain-general sequence learning deficit in specific language impairment. *Neuropsychology*, 28(3), 472–483.
- Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T.** (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, 51, 1–10.
- Mackie, C., & Dockrell, J. E.** (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language, and Hearing Research*, 47(6), 1469–1483.
- McElreath, R.** (2013). *glmer2stan: RStan models defined by glmer formulas. R package version 0.995*.
- 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., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J.** (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research*, 60, 2891–2905.
- McGregor, K., Oleson, J., Bahnsen, A., & Duff, D.** (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders*, 48(3), 307–319.
- Mintz, T. H., Newport, E. L., & Bever, T. G.** (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26(4), 393–424.
- Misyak, J. B., & Christiansen, M. H.** (2012). Statistical learning and language: An individual differences study. *Language Learning*, 62(1), 302–331.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B.** (2010). Frontiers: On-line individual differences in statistical learning predict language processing. *Frontiers in Language Sciences*, 1, 31.
- Morice, R., & McNicol, D.** (1985). The comprehension and production of complex syntax in schizophrenia. *Cortex*, 21(4), 567–580.
- Mueller, K. L., & Tomblin, J. B.** (2012). Examining the comorbidity of language disorders and ADHD. *Topics in Language Disorders*, 32(3), 228–246.
- Newport, E. L., & Aslin, R. N.** (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B.** (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech-Language Pathology*, 17(4), 356–366.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B.** (2009). Syntactic development in adolescents with a history of language impairments: A follow-up investigation. *American Journal of Speech-Language Pathology*, 18(3), 241–251.
- Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A.** (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in Psychology*, 7, 1245.
- Plante, E., Gómez, R., & Gerken, L.** (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, 35(5), 453–462.
- R Core Team.** (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- R Core Team.** (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Redmond, S. M.** (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics*, 19(2), 109–127.

- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66(1), 30–54.
- Reilly, S., Bishop, D. V., & Tomblin, B. (2014). Terminological debate over language impairment in children: Forward movement and sticking points. *International Journal of Language & Communication Disorders*, 49(4), 452–462.
- Rice, M. L., & Bode, J. V. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language*, 13(37), 113–131.
- Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39(6), 1239–1257.
- Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research*, 48(6), 1397–1411.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621.
- Sanz-Torrent, M., Andreu, L., Badia, I., & Sidera, F. (2011). Argument omissions in preschool Catalan and Spanish speaking children with SLI. *Infancia y Aprendizaje*, 34(1), 49–66.
- Schneider, E., & Zuccoloto, A. (2007). E-prime 2.0 [Computer software]. Pittsburg, PA: Psychological Software Tools.
- 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.
- Skipp, A., Windfuhr, K. L., & Conti-Ramsden, G. (2002). Children's grammatical categories of verb and noun: A comparative look at children with specific language impairment (SLI) and normal language (NL). *International Journal of Language & Communication Disorders*, 37(3), 253–271.
- Smith, K. H. (1966). Grammatical intrusions in the free recall of structured letter pairs. *Journal of Verbal Learning and Verbal Behavior*, 5(5), 447–454.
- Smith, K. H. (1969). Learning Co-occurrence restrictions: Rule induction or rote learning? *Journal of Verbal Learning and Verbal Behavior*, 8(2), 319–321.
- Stan Development Team. (2016). *RStan: The R interface to Stan. R package version 2.14.1*. Retrieved from <http://mc-stan.org/>
- St. Clair, M. C. S., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, 116(3), 341–360.
- Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: An fMRI study of children and adults. *Journal of Cognitive Neuroscience*, 16(8), 1339–1351.
- Thordardottir, E. T., & Weismer, S. E. (2002). Verb argument structure weakness in specific language impairment in relation to age and utterance length. *Clinical Linguistics & Phonetics*, 16(4), 233–250.
- Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, 3(4), 269–293.
- 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.
- Van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, 63(2), 197–209.
- Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22), 3527–3541.
- Wetherell, D., Botting, N., & Conti-Ramsden, G. (2007). Narrative in adolescent specific language impairment (SLI): A comparison with peers across two different narrative genres. *International Journal of Language & Communication Disorders*, 42(5), 583–605.
- Windsor, J., Scott, C. M., & Street, C. K. (2000). Verb and noun morphology in the spoken and written language of children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, 43(6), 1322–1336.

Appendix (p. 1 of 2)

List of All Training Items, Listed by AXB Combination Type and Sequence Type.

Item	Type	Sequence
daffin tomber bleggin	AXB1	AXB
daffin tomber bleggin gentif	AXB1	AXBR
klidum daffin tomber bleggin ^a	AXB1	QAXB
spad daffin tomber bleggin gentif	AXB1	QAXBR
daffin tomber mawg frag	AXB2	AXB
klidum daffin tomber mawg gentif	AXB2	QAXBR
daffin zub bleggin gentif	AXB3	AXBR
klidum daffin zub bleggin	AXB3	QAXB
spad daffin zub bleggin frag	AXB3	QAXBR
daffin zub fluggit	AXB4	AXB
daffin zub fluggit gentif	AXB4	AXBR
spad daffin zub fluggit	AXB4	QAXB
flairb lapal fluggit	AXB5	AXB
flairb lapal fluggit frag	AXB5	AXBR
spad flairb lapal fluggit frag	AXB5	QAXBR
flairb lapal mawg	AXB6	AXB
flairb lapal mawg frag	AXB6	AXBR
klidum flairb lapal mawg	AXB6	QAXB
flairb tomber bleggin	AXB7	AXB
klidum flairb tomber bleggin frag	AXB7	QAXBR
spad flairb tomber bleggin	AXB7	QAXB
flairb lapal fluggit	AXB8	AXB
flairb lapal fluggit frag	AXB8	AXBR
spad flairb lapal fluggit frag	AXB8	QAXBR
glim lapal fluggit	AXB9	AXB
klidum glim lapal fluggit	AXB9	QAXB
spad glim lapal fluggit gentif	AXB9	QAXBR
glim lapal mawg	AXB10	AXB
glim lapal mawg gentif	AXB10	AXBR
spad glim lapal mawg	AXB10	QAXB
glim zub bleggin	AXB11	AXB
glim zub bleggin frag	AXB11	AXBR
klidum glim zub bleggin frag	AXB11	QAXBR
glim zub fluggit	AXB12	AXB
glim zub fluggit gentif	AXB12	AXBR
klidum glim zub fluggit	AXB12	QAXB

^aThis item should have been *klidum daffin tomber mawg* and was intended to be an AXB2 item.

Appendix (p. 2 of 2)

List of All Test Items, Listed by Combination Type.

Grammatical		Ungrammatical	
Item	Type	Item	Type
daffin tomber mawg	AXB2	daffin zub flairb	AXA
daffin zub bleggin	AXB3	flairb lapal glim	AXA
daffin zub fluggit	AXB4	glim zub daffin	AXA
flairb lapal fluggit	AXB5	daffin tomber frag	AXR
flairb lapal mawg	AXB6	flairb zub gentif	AXR
flairb tomber bleggin	AXB7	glim lapal frag	AXR
glim lapal fluggit	AXB9	mawg bleggin flairb	BBA ^a
glim lapal mawg	AXB10	mawg lapal daffin	BXA
glim zub bleggin	AXB11	bleggin zub glim	BXA
daffin lapal bleggin	novel AXB	fluggit tomber flairb	BXA
daffin lapal mawg	novel AXB	bleggin lapal fluggit	BXB
daffin tomber fluggit	novel AXB	mawg zub bleggin	BXB
flairb tomber fluggit	novel AXB	fluggit tomber mawg	BXB
flairb zub bleggin	novel AXB	spad lapal fluggit	QXB
flairb zub mawg	novel AXB	spad tomber bleggin	QXB
glim tomber bleggin	novel AXB	klidum zub mawg	QXB
glim tomber fluggit	novel AXB	tomber mawg glim	XBA
glim zub mawg	novel AXB	zub fluggit daffin	XBA

Note. Combination type for familiar items refers back to the AXB combination as listed in the earlier part of the Appendix.

^aBBA occurred as a mistake; this item was intended to be a BX A item.