Research Article

Individual and Developmental Differences in Distributional Learning

Jessica Hall,^a Amanda J. Owen Van Horne,^b
Karla K. McGregor,^{a,c} and Thomas A. Farmer^a

Purpose: This study examined whether children and adults with developmental language disorder (DLD) could use distributional information in an artificial language to learn about grammatical category membership similarly to their typically developing (TD) peers and whether developmental differences existed within and between DLD and TD groups.

Method: Sixteen children ages 7–9 with DLD, 26 age-matched TD children, 17 college students with DLD, and 17 TD college students participated in this task. We used an artificial grammar learning paradigm in which participants had to use knowledge of category membership to determine the acceptability of test items that they had not heard during a training phase.

Results: Individuals with DLD performed similarly to TD peers in distinguishing grammatical from ungrammatical combinations, with no differences between age groups. The order in which items were heard at test differentially affected child versus adult participants and showed a relation with attention and phonological working memory as well.

Conclusion: Differences in ratings between grammatical and ungrammatical items in this task suggest that individuals with DLD can form grammatical categories from novel input and more broadly use distributional information. Differences in order effects suggest a developmental timeline for sensitivity to updating distributional information.

How is it that we all can agree that *splorbs* (/splɔrbz/) could be the English name of a newly invented candy but *zplorbs* (/zplɔrbz/) could not? How is it that children fill lexical gaps in endearing utterances such as *I broomed the mess*, and we readily understand what they mean? One evidence-based explanation is that people pay attention to statistical regularities in linguistic input, and these regularities support language learning and processing. Specifically, knowledge of the phonotactics of English is supported by our ability to track the conditional probability that one phoneme comes immediately before or after another (adjacent dependencies; Saffran, Aslin, & Newport, 1996). As conditional probability approaches zero, words such as *zplorbs* will be judged as unacceptable. Similarly, to learn that a plural determiner such as *some* predicts a plural morpheme, such as *-s*, children can attend to the conditional probability that one morpheme predicts the other, even with

intervening material between them (nonadjacent dependencies; Gómez, 2002). *Jon built some tables* is acceptable but *Jon built some table* is not. The latter violates our statistically derived expectation that a plural *-s* typically follows *some*.

Not all statistical patterns in the language are conditional; some are distributional. Our knowledge of grammatical categories is likely supported by our ability to track distributional regularities. Distributional regularity refers to patterns of co-occurrence over sets of exemplars. For example, verbs often follow pronouns (e.g., I, you, he, she, it, we, they) and precede inflectional morphemes (e.g., *-ing*, *-ed*, *-s*), so something sandwiched between the pronoun *I* and the inflection *-ed*—such as *broom* in *I broomed the mess*—can be interpreted as a verb. Evidence that grammatical categories can be informed by distributional information comes from artificial language learning experiments (Reeder, Newport, & Aslin, 2013) and computational modeling and behavioral experiments of novel verb learning (Mintz, 2006).

That statistical learning supports language learning, and processing is suggested by research on individual differences. For example, among typical adults, those who performed better in an artificial grammar learning task with nonadjacent dependencies read object relative sentences more quickly than those who performed worse (Misyak, Christiansen, & Tomblin, 2010). The performance of typical

^aThe University of Iowa, Iowa City

^bThe University of Delaware, Newark

^cBoys Town National Research Hospital, Omaha, Nebraska

Correspondence to Jessica Hall: jessicahall@email.arizona.edu

Editor-in-Chief: Mary Alt

Received November 27, 2017

Revision received February 14, 2018

Accepted March 11, 2018

https://doi.org/10.1044/2018_LSHSS-STLT1-17-0134

Publisher Note: This article is part of the Special Issue: How Statistical Learning Relates to Speech-Language Pathology.

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

4- to 5-year-olds on a statistical learning task with sequentially presented information corresponded with their ability to remember a primed structure (Kidd, 2012). Similarly, 6- to 8-year-olds' performance on a statistical learning task with adjacent dependencies, independent of working memory, predicted accuracy in comprehension of passives and object relative clauses (Kidd & Arciuli, 2016).

In the field of communication sciences and disorders, an increasing number of studies have examined statistical learning in populations such as developmental language disorder (DLD; also known as specific language impairment; Bishop, Snowling, Thompson, & Greenhalgh, 2017), as deficits in statistical learning may relate to deficits observed in language. As we shall see, evidence demonstrating that people with DLD have deficient conditional learning has amassed. Much less is known about their distributional learning.

DLD and Different Types of Statistical Learning

Conditional Statistical Learning

Table 1 provides a summary of statistical learning tasks and findings. With some exceptions (Clark & Lum, 2017; Richardson, Harris, Plante, & Gerken, 2006), the evidence reveals that both adjacent and nonadjacent conditional dependencies represent a challenge for individuals with DLD, and this challenge might account, in part, for the grammatical impairments that characterize DLD. For instance, children with DLD had difficulty in tracking adjacent probabilities in a speech segmentation task, in which the frequency that syllables co-occur could be used to predict word boundaries; they required increased exposure to perform as well as typically developing (TD) children (Evans, Saffran, & Robe-Torres, 2009). Von Koss Torkildsen, Dailey, Aguilar, Gómez, and Plante (2013) found adults with DLD able to learn and generalize adjacent dependencies under a high-variability condition (when there were 24 exemplars of the X syllable in AX strings) but not a low-variability condition (three exemplars). As for nonadjacent dependencies, Grunow, Spaulding, Gómez, and Plante (2006) presented adults with DLD and typical peers with a grammar in which one nonsense word (A) predicted another nonsense word (B) with at least one word in the middle (X). The individuals with DLD could not distinguish between grammatical (AXB) and ungrammatical strings (BXA or AXC) in the artificial grammar, although unaffected adults could. Hsu, Tomblin, and Christiansen (2014) found adolescents with DLD to be deficient at learning nonadjacent dependencies even under conditions of high variability. If statistical learning is at the least helpful for learning grammar, individuals with DLD seem at a disadvantage.

Distributional Statistical Learning

In contrast to the bulk of the findings on conditional statistical learning, our recent work has shown that adults with DLD can use distributional information in an artificial grammar learning task to the same degree as TD adults (Hall, Owen Van Horne, McGregor, & Farmer, 2017). In

that study, we trained participants with an artificial grammar but purposefully left out certain combinations of items during training. Some of these items were consistent with the grammar (i.e., grammatical), and some were not (i.e., ungrammatical). At test, participants rated the untrained grammatical combinations as more acceptable than untrained ungrammatical combinations. The adults with and without DLD performed similarly.

Whereas spared distributional learning but deficient conditional learning among people with DLD might seem puzzling, Thiessen (2017) points to a potential explanation. Thiessen (2017) proposed that distributional learning and conditional learning are supported by different neurological mechanisms. Distributional learning relies on noticing similarity among many exemplars, whereas conditional learning requires extracting a conditional association based on the frequency of the co-occurrence of two exemplars (Thiessen, Kronstein, & Hufnagle, 2013). The separate mechanisms hypothesis garners support from developmental studies. Consider that children as young as 8 months use conditional probabilities in speech segmentation tasks (Saffran et al., 1996). In contrast, 18-month-olds use distributional information in an artificial grammar learning task, but 12-month-olds do not (Marchetto & Bonatti, 2013). Evidence for separate mechanisms also comes from computational models that show differences in adeptness at identifying conditional regularities versus distributional regularities (Endress & Bonatti, 2007; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Thiessen & Pavlik, 2013).

In light of the overwhelming evidence of conditional learning deficits among individuals with DLD, it is tempting to view our finding that adults with DLD had no difficulty with distributional learning (Hall, Owen Van Horne, McGregor, & Farmer, 2017) as a third strand of evidence in support of the separate mechanisms hypothesis. However, because our participants were adults, another possibility must be considered. Perhaps their successful learning was attained through compensatory processes they established over the course of development. A comparison of performance with children with DLD would allow us to determine whether distributional learning is spared at both points in development or whether deficits in distributional learning exist early in development and are ameliorated by adulthood.

Although DLD persists into adulthood, children with DLD show growth in language ability over developmental time, often with delays compared to peers. For example, tense morphology acquisition occurs at a delayed rate in DLD (Rice, Wexler, & Hershberger, 1998). Therefore, statistical learning is apt to differ between children with DLD and adults with DLD; the key test is to determine whether they differ to a greater degree than children and adults with TD. A recent meta-analysis of statistical learning in DLD found no influence of age in relation to the amount of deficit; however, this article analyzed data from just 10 studies, none of which tested distributional learning, and the age range of participants was 6–19 years old (Lammertink, Boersma, Wijnen, & Rispens, 2017). A

Table 1. Types of statistical learning studies and findings for individuals with developmental language disorder (DLD).

Statistic	Dependencies	Stimuli	In language	Findings for DLD
Conditional	Adjacent dependencies	A predicts B immediately	<i>Pret+ty</i> more common than <i>ty+ba</i> : word boundary is pretty baby	Deficits for adults (Plante et al., 2002) and children (Evans et al., 2009; Haebig et al., 2017; von Koss Torkildsen et al., 2013)
	Nonadjacent dependencies	A predicts B following intervening X	<i>is</i> predicts <i>-ing</i> after an intervening verb	Deficits for adults (Grunow et al., 2006) and adolescents (Hsu et al., 2014); NS differences in visual task with adults (Aguilar & Plante, 2014)
	Higher order	A predicts D when followed by B, not when followed by C	<i>is</i> predicts <i>-ing</i> after a verb but not after an adjective	No deficits for children (Clark & Lum, 2017)
	First order	A predicts B	Gender suffixes in Russian: <i>knig</i> predicts <i>-oj</i> ; gender suffixes in Spanish: <i>la</i> predicts <i>-a</i> Also like examples from adjacent/nonadjacent above	Deficits for children ages 7–12 (Clark & Lum, 2017; Hsu & Bishop, 2014); preserved learning in adults but no generalization (Richardson et al., 2006)
Distributional	Categorical	A words (but not B words) predict X words	Verbs (but not nouns) can be followed by <i>-ed</i> and <i>-ing</i>	No deficits for adults (Hall, Owen Van Horne, McGregor, & Farmer, 2017)

Note. NS = nonsignificant.

more appropriate test of developmental differences would be to use the same task with child and adult participants with and without DLD.

From our previous work, we had three fourths of the data required for such a test: TD children from Hall, Owen Van Horne, and Farmer (2017) and adults with and without DLD from Hall, Owen Van Horne, McGregor, and Farmer (2017). For the purpose of this special issue on statistical learning in atypical populations, we took the opportunity to add data from children with DLD so that we could directly compare statistical learning between ages and diagnostic groups. The combined data and the new analysis presented here address the question of whether adults and children with and without DLD are equally capable of extracting distributional regularities, a particular kind of statistical pattern, from relatively brief input.

We also take this opportunity to explore additional questions about sensitivity to changing distributional information and the role of attention. Our study with adults with DLD (Hall, Owen Van Horne, McGregor, & Farmer, 2017) showed a large effect of the order in which test items were heard. We often think of test items as being interchangeable, but particularly in artificial grammar learning studies, test items are new examples of the artificial grammar. When the test items are grammatical items, this increases exposure to the target grammar. However, ungrammatical test items may lead to drift, pulling the learner toward different conclusions about the made-up language. This effect of testing order is important to study across developmental groups because differences could indicate increased sensitivity to new information, as the distribution changes across the course of

the testing phase through exposure to ungrammatical phrases. With regard to DLD, being slower to update than peers could lead to difficulty in processing or delayed growth. There is some evidence that this is indeed observable in natural language learning in DLD. Older children with DLD have shown slower retreat from overgeneralization. Both typical children and children with DLD make errors in terms of using a verb such as “fill” in an utterance where verbs such as “pour” are found (e.g., “the lady is filling the sweets into the jar”), but individuals with DLD make more errors at older ages than typical individuals do (Ebbels, 2005). It is generally agreed that the retreat from overgeneralization happens because children eventually notice that they never hear the incorrect sentences and instead should use verbs such as “fill” differently. Finally, we wanted to determine that any developmental differences in distributional learning performance were not simply due to increased attention during training due to more mature executive functioning mechanisms. Therefore, we include a measure of attention during training in our final models to determine its impact on test performance. Given that a number of studies have also shown poorer attention and working memory in DLD, this would also allow us to conclude whether any diagnostic group differences were attributable to poorer attention/memory of novel word forms.

Research Questions and Hypotheses

In this article, we examine three main questions. Our first question is whether individuals with and without DLD

show developmental differences in distributional learning. We test two competing hypotheses:

Hypothesis 1. Distributional learning is preserved in DLD at all ages. Despite deficits in statistical learning performance evident from studies of conditional statistics, we might see preserved learning of distributional information because different mechanisms may be involved. Similar performance by children and adults with and without DLD would support this hypothesis.

Hypothesis 2. Distributional learning is deployed differently at different developmental time points in DLD. To support this hypothesis, we would expect to see that adults with DLD could recognize grammatical and ungrammatical strings, but children with DLD could not.

Our second question is whether children with DLD are slower to update language representations. This would be supported by different effects of test item order between diagnostic groups and ages. Our final question is whether distributional learning is supported by attention and memory mechanisms that mature with development. Evidence would comprise a relationship between a task measuring attention during training and the ability to recognize grammatical and ungrammatical strings.

Method

Participants

There were 34 adult 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). These adults were originally recruited for studies of word learning, 33 in McGregor, Arbisi-Kelm, and Eden (2017) and one in McGregor (2017) and details about their recruitment and testing can be found there. Subsequently, they participated in Hall, Owen Van Horne, McGregor, and Farmer (2017), and the data reported there are the data we analyzed for the current article.

There were 46 child participants ages 7–9, 20 of whom were classified as having DLD through a score of 95 or below on the Structured Photographic Expressive Language Test–Third Edition (SPELT-3; Dawson, Stout, & Eyer, 2003), as recommended by Perona, Plante, and Vance (2005). All but two children in the TD group participated in Hall, Owen Van Horne, and Farmer (2017), and the data reported there are the data we analyzed for the current article. Six children in the DLD group participated in this study after completing a treatment study on morpheme production (Owen Van Horne, Curran, Larson, & Fey, 2018; Owen Van Horne, Fey, & Curran, 2017).

Independent sample *t* tests were conducted to ensure groups were well matched on age; Cohen's *d* is reported in Table 2. All participants had normal hearing per self-report (adults) or screening (children; American Speech-Language-Hearing Association, 1997), had normal or corrected vision per parent or self-report, scored above 85 on the Kaufman Brief Intelligence Test–Second Edition matrices subtest (Kaufman & Kaufman, 2004), and had no history of acquired

neurological disorders by parent or self-report. None of the participants in the child or adult TD groups reported any history of acquired neurological disorders or any other diagnoses of cognitive or language impairment. We did not exclude participants in the DLD group with attention-deficit/hyperactivity disorder or dyslexia, as these conditions are often comorbid with DLD (Catts, Adlof, Hogan, & Weismer, 2005; Mueller & Tomblin, 2012), and furthermore, individuals with DLD often receive a variety of labels (Reilly, Bishop, & Tomblin, 2014). In addition, all participants completed the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007) to document their vocabulary skills. Diagnostic groups differed on all standardized assessments, which are summarized in Table 2 along with demographic information.

Stimuli

We used the artificial grammar from Reeder et al. (2013) and adapted stimuli for use with children. Differences between our stimuli and the original study are described at the end of the Method section. The grammar consisted of five category types, with two to three words per category. Category membership determined the order in which words could appear during training. Training sentences minimally contained A, X, and B words, in that order, with Q and R words appearing optionally at the beginning and end of the sentence. Q and R words prevented participants from deducing category membership by position alone (i.e., certain words always occurring at the beginning or end of a sentence). Training and test items can be found in the Appendix.

Training

During training, participants heard a corpus of 36 sentences, derived from 12 combinations of A, X, and B words, occurring in three of four possible sentence types, QAXB, AXB, AXBR, or QAXBR. For example, the AXB combination *flairb tomber bleggin* also appeared in *klidum flairb tomber bleggin frag* (QAXBR) and *spad flairb tomber bleggin* (QAXB). Each sentence was heard three times for a total of 108 trials. The task tests category learning by exposing participants to only a subset of grammatically possible combinations. The shared distribution of category items is a cue to listeners as to which combinations are grammatically possible. The grammar was such that words with similar distributions (e.g., appearing after a certain *X* word or before a certain *R* word) belonged to the same category, and category membership cued which combinations of words were possible. If learners were sensitive to these shared distributions, they would be able to distinguish which combinations were possible while only being trained on a subset of the corpus. For example, one of the A words, *flairb*, never appears with one of the X words, *zub*, during the training. However, *flairb* appears with *tomber*, and other A words that appear with *tomber* also appear with *zub*. Figure 1 displays combinations of A (in black), X (in dark gray), and B (in light gray) heard during training, with excluded combinations

Table 2. Participant demographic and test means (standard deviations) by age group (children and adults) and diagnostic group (developmental language disorder [DLD] and typically developing [TD]), in which Cohen's *d* represents effect size of group differences for each measure for participants included in analyses.

Group		<i>n</i>	Age (years)	Education (years)	KBIT-2 SS	PPVT-4 Raw	SPELT-3 SS	Spelling	Token
Children	DLD	16	7.7 (.7)	15.2 (2.1)	109.4 (13.7)	131.8 (13.3)	85.0 (10.3)		
	TD	24	7.9 (.7)	18.4 (3.4)	117.0 (10.3)	154.4 (12.5)	111.4 (7.0)		
<i>d</i>			0.23	1.11**	0.62	1.75***	2.99***		
Adults	DLD	17	20.6 (1.1)	13.6 (1.1)	98.6 (10.7)	199.4 (11.2)		4.5 (3.0)	35.4 (5.1)
	TD	17	21.1 (1.8)	14.0 (1.7)	110.1 (13.0)	206.5 (6.7)		11.8 (2.4)	40.1 (3.0)
<i>d</i>			0.28	0.24	0.97**	0.77*		2.69***	1.12**

Note. Education is reported for children's mothers and adult participants. Scores on the KBIT-2 Nonverbal subtest and the SPELT-3 are standard scores with a normative mean of 100 and a standard deviation of 15. Scores on the PPVT-4 are raw scores. Scores on the Spelling and Token tasks are raw counts of items correct out of 15 and 44, respectively. KBIT-2 = Kaufman Brief Intelligence Test–Second Edition; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; SPELT-3 = Structured Photographic Expressive Language Test–Third Edition; SS = standard score.

* $p < .05$. ** $p < .01$. *** $p < .0001$.

indicated by dashed lines. The training took 10–15 min and was sufficient for 7-year-old TD children to learn the grammar (Hall, Owen Van Horne, & Farmer, 2017).

To ensure attention during training, participants completed a 1-back task in which they indicated by button press if the current sentence repeated the sentence in the trial immediately prior, which was true for 35 trials. A priori it was determined that data from individuals who scored less than 60% accuracy on the 1-back task would be discarded. We were able to retain data from all participants, except for three children in the DLD group and one additional participant who qualified as DLD but refused all tasks and withdrew from the study. Data from these four participants are excluded from analyses and the demographic data reported in Table 2. This resulted in data from only 16 children with DLD being employed in all subsequent analyses reported here.

Test

The test phase consisted of 36 AXB sentences of three types: “familiar” sentences consisting of nine of the 12 AXB combinations heard during training, “novel” sentences that consisted of an additional nine AXB combinations

never heard during training but consistent with the rules of the artificial grammar, and “ungrammatical” sentences that violated the rules of the artificial grammar (i.e., were not AXB combinations but were instead AXA, BXB, AXR, BXA, QXB, and XBA). Each of the 18 ungrammatical sentences was unique to reduce drift during testing, whereas the nine familiar and nine novel sentences were heard twice, for a total of 54 trials. The Appendix lists all training and test items. Average associative chunk strength, which is the average of the frequency during training of the three component dependencies of every test item (AX, XB, and AXB; Knowlton & Squire, 1994), provides a measure of familiarity. Novel and ungrammatical test items could have similar associative chunk strength because each contained one familiar bigram and one unfamiliar bigram. We ran a mixed-effects model with associative chunk strength as the dependent variable, a random effect of item, and item type as a predictor. Chunk strength did not differ between novel ($M = 4.83$) and ungrammatical ($M = 4.94$) items, $t(32.95) = -0.26$, $p = .80$.

Participants indicated the acceptability of a test item by button press and by moving a slider scale to indicate confidence (see Ambridge, Pine, Rowland, & Young, 2008). The scale provided increased power to detect variability in

Figure 1. Depiction of artificial grammar. The words in the darkest boxes belong to “A” category, words in medium gray belong to “X” category, and words in lightest gray belong to “B” category. Solid lines indicate combinations of words heard in training items (e.g., *flairb lapal mawg*); dashed lines indicate grammatical combinations never heard during training that occurred in novel test items (e.g., *flairb zub mawg*). From Hall, Owen Van Horne, McGregor, and Farmer (2017). Copyright © 2017 by the American Speech-Language-Hearing Association. Reprinted with permission.

learning. Prior to the experiment, participants performed an apple/pear visual discrimination task that had the same button choice and scale to familiarize them with response measures.

Procedure

We used E-Prime to deliver the experiment and perform data collection. In the training phase, participants were instructed to watch videos of cartoon aliens saying the sentences on a laptop. They were told 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). Participants performed the 1-back task on every trial, except the first. The training also included eleven 4-s nonverbal video clips of the aliens doing various activities. These played intermittently between trials to keep participants' interest.

The test phase immediately followed training. Participants were instructed to listen to each sentence and indicate if it was something the aliens would say by pressing the red button for no and the green button for yes and then moving a slider scale with corresponding colors to specify their certainty.

Changes to the original Reeder task are as follows. Stimuli were recorded in child-directed tone and pitch by an adult female not familiar with the study. The training phase was presented in video format with a 1-back task to make it more engaging for child participants. This additionally served as a measure of attention. We used only one version of the grammar because 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, whereas we have included 12. 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. We demonstrate in Hall, Owen Van Horne, and Farmer (2017) that participants rate ungrammatical items equally, with the exception of AXR items. 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, Mächler, Bolker, & Walker, 2015) and the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) in R version 3.4.3 (R Core Team, 2017) to run linear mixed-effects models to explore comparisons of interest. Comparison of Akaike information criteria (AIC) was used to select the best-fit random effects structures. We converted scale ratings to *z* scores as in Reeder et al. (2013) to control for variable use of the scale

by participants. To simplify findings, *z*-score scale ratings are the dependent variable for all analyses.

To answer our first question about the role of developmental differences within and between diagnostic groups, we constructed a model with scale rating as the dependent variable and item type (familiar, novel, ungrammatical), diagnostic group membership (DLD, TD), age group (adult, child), and their interactions as fixed effects. Responses for novel versus ungrammatical items are the critical comparison for formation of categories as they are 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 items, the TD group, and the adult group served as the reference categories, unless otherwise specified. For this and the following models, we used AIC comparison to determine the best-fit fixed-effects structure. To address our second research question, whether age and diagnostic groups differ in how they update information during test, we ran a second model that included the order in which test items were heard. To confirm that adult data were not masking child diagnostic group differences, we also examined the child data set on its own, asking the same questions described previously. No differences from the larger analysis were identified, and thus, this is not reported here. Finally, to address our third research question about the role of attention and phonological working memory, we tested accuracy on the 1-back test during training as a measure of attention. This variable was centered to minimize multicollinearity. This model excluded diagnostic group because (a) we expected attention to influence performance equally across groups and (b) diagnostic group membership was associated with this factor, as reported below.

Results

For all models, AIC comparison determined that the best-fit models included only a random item intercept.

1-Back Performance

First, we examined 1-back performance to determine whether this measure differed across diagnostic group and age. Recall that we excluded three individuals (all children with DLD) performing with less than 60% accuracy as a control to ensure at least a minimal attention to the artificial grammar learning task. Even with these three children excluded, linear regression showed main effects and an interaction between age and diagnostic groups in predicting mean 1-back accuracy. Table 3 reports mean 1-back accuracy for diagnostic and age groups. Adults were more accurate than children, $t = -15.87$, $p < .0001$, and TD groups were more accurate than DLD groups, $t = -4.83$, $p < .0001$. The gap between diagnostic groups differed by age, $t = -7.51$, $p < .0001$. Unpacking the interaction by dividing into adult and child age groups revealed that the gap between diagnostic groups was larger for children than for adults, diagnostic group effect in children, $B = -.06$,

Table 3. Mean accuracy and standard deviation on the 1-back task during training by diagnostic (developmental language disorder [DLD] and typically developing [TD]) and age groups (children and adults).

Group	Mean accuracy	SD
Children with DLD	83%	8%
TD children	89%	9%
Adults with DLD	93%	7%
TD adults	94%	7%

$t = -14.71$, $p < .0001$, and in adults, $B = -.02$, $t = -5.32$, $p < .0001$. We return to these differences in our final analysis.

Developmental Differences

Next, we directly compared test performance by adults and children within and across diagnostic groups. The ability to distinguish novel and ungrammatical items was the critical test of category formation from distributional information. Figure 2 displays results. There were no main effects or interactions with diagnostic or age group, $ps > .1$. Standardized ratings for novel items were higher than ungrammatical items, $t(57) = 2.64$, $p = .01$, indicating that both DLD and TD groups formed grammatical categories. Familiar items were also rated higher than ungrammatical items, $t(57) = 4.42$, $p > .0001$, but did not differ from novel items, $p = .11$, indicating that both novel grammatical and familiar grammatical items were considered equally good. These results

suggest that individuals with DLD can use distributional information similarly to same-age peers across development. Recall that the results of the model hold with and without adults included, indicating that this finding is robust across developmental time.

We added order to our model to determine the effects of updating information during the testing phase. Model results are provided in Table 4 and illustrated in Figure 3. The effect of item type remained, and there were significant two-way interactions between item type and order, familiar, $t(145) = -4.07$, $p < .0001$, novel, $t(136) = -3.25$, $p = .001$, item type and age group, $t(405) = -2.21$, $p = .03$, and three-way interactions among item type, order, and age group, and item type, order, and diagnostic group. The three-way interaction between item type, order, and age group shows a larger order effect for familiar items, which was in turn larger for the adult group, $t(4048) = 2.48$, $p = .01$. The interaction between order and item type is unsurprising because familiar and novel items repeated and ungrammatical items did not, but it suggests that adult learners, at least, are updating and tempering their ratings for what is acceptable in the language as they are exposed to more ungrammatical items. The three-way interaction between item type, order, and diagnostic group shows a larger order effect for familiar items for the TD group, $t(4048) = 2.02$, $p = .04$. The three-way interactions suggest that adults are better than children and individuals with TD are better than individuals with DLD at remembering specific items from the training early on during the test and become worse later with interference from other items.

Figure 2. Mean standardized rating for each item type (familiar, novel, ungrammatical) by age and diagnostic group (typically developing [TD], developmental language disorder [DLD]). Error bars represent standard error.

Table 4. Results of the mixed-effects model showing the influence of item type, age group, diagnostic group (typically developing [TD], developmental language disorder [DLD]), and order on ratings.

Factor	Variance	SD	β	SE	p
Random factors					
Item	0.06	0.26			
Fixed factors					
(Intercept)			−0.42	0.18	.02
Item type (reference category = ungrammatical)					
Familiar			1.30	0.23	< .01
Novel			0.91	0.23	< .01
Age group (reference category = adult)			−0.06	0.15	.68
Diagnostic group (reference category = TD)			−0.08	0.17	.65
Order			0.004	0.006	.51
Age group × Item type					
Child × Familiar			−0.45	0.20	.03
Child × Novel			−0.07	0.20	.74
Diagnostic group × Item type					
DLD × Familiar			−0.31	0.22	.15
DLD × Novel			0.04	0.22	.86
Order × Item type					
Order × Familiar			−0.03	0.007	< .01
Order × Novel			−0.02	0.006	< .01
Age group × Diagnostic group			−0.09	0.23	.69
Order × Diagnostic group			0.0004	0.005	.94
Age group × Order			−0.003	0.005	.49
Age group × Diagnostic group × Item type					
Child × DLD × Familiar			0.01	0.30	.68
Child × DLD × Novel			−0.28	0.30	.35
Age group × Diagnostic group × Order			0.0007	0.007	.92
Age group × Order × Item type					
Child × Order × Familiar			0.02	0.006	.01
Child × Order × Novel			0.007	0.006	.28
Diagnostic group × Order × Item type					
DLD × Order × Familiar			0.01	0.007	.04
DLD × Order × Novel			0.004	0.007	.56
Age group × Diagnostic × Order × Item type					
Child × DLD × Order × Familiar			−0.01	0.009	.31
Child × DLD × Order × Novel			0.004	0.009	.66

Attention as a Predictor of Learning

To examine the contributions of these differences in 1-back accuracy on distributional learning, we included interactions between item type, age group, order, and 1-back accuracy. Table 5 gives the model results, and Figure 4 displays the influence of training accuracy on ratings of grammatical items. Note that only grammatical items are included in the figure because order did not affect ratings for ungrammatical items, as reported in the second model. For visualization purposes, we divided test items into items the first time they were heard (the left side of the figure) and the second time they were heard (the right side of the figure), though order is included as a continuous variable in our statistical model. As observed in previous models, the main effect of item type and the interaction between item type and order were maintained, $ps < .001$, as well as the three-way interaction between item type (familiar), order, and age group, $p = .04$. In addition, we observed a main effect of 1-back accuracy, $t(3994) = 2.70$, $p = .007$, and marginal main effects of age group, $p = .05$, and order, $p = .08$. There was an interaction between 1-back accuracy and age group, $t(3994) = -2.63$, $p = .009$, an interaction between 1-back accuracy and order, $t(3994) = -3.01$, $p = .003$, an interaction between age

group and order, $t(3994) = -2.06$, $p = .04$, and a three-way interaction between these variables, $t(3994) = 2.42$, $p = .02$, such that better performance on the 1-back task during training resulted in higher ratings for items heard earlier rather than later during test. This effect was driven by adults, suggesting that adults with better attention during training showed increased sensitivity to updating information during test. The lack of interaction between item type and 1-back accuracy is evidence that attention during training did not substantially contribute to participants' ability to distinguish novel and ungrammatical items, which was the crucial test of grammatical category learning.

Discussion

In this study, we compared the distributional learning of children and adults with and without DLD. The key contribution of our results is that we see developmental change but no diagnostic differences in distributional learning in individuals with and without DLD. We could not have determined this without using the same task for child and adult groups and conducting a direct four-way statistical comparison of learning at different ages and between

Figure 3. Influence of test item order as it interacted with item type (familiar, novel, ungrammatical) for child and adult participants, across diagnostic groups (typically developing [TD], developmental language disorder [DLD]) on ratings.

diagnostic groups. Casual comparison of ratings between groups and age levels may have led to the erroneous conclusion of developmental and group differences in acceptability ratings for novel items, particularly. It is only by directly comparing groups that we can confidently conclude that children and adults with and without DLD do not differ in the

formation of grammatical categories from distributional information alone and that differences only occur in performance over the length of the testing period.

We tested two hypotheses regarding the developmental timescale of distributional learning in DLD: (a) distributional learning is preserved across age in DLD or (b) children

Table 5. Results of the mixed-effects model showing the influence of 1-back accuracy during training as interacting with item type, age group, and order on ratings.

Factor	Variance	SD	β	SE	p
Random factors					
Item	0.07	0.26			
Fixed factors					
(Intercept)			−0.63	0.17	< .01
Item type (reference category = ungrammatical)					
Familiar			1.22	0.22	< .01
Novel			1.01	0.21	< .01
Age group (reference category = adult)			0.25	0.13	.05
Order			0.09	0.005	.08
1-back accuracy (centered)			3.09	1.15	< .01
Age group × Item type					
Child × Familiar			−0.40	0.17	.02
Child × Novel			−0.21	0.17	.21
Order × Item type					
Order × Familiar			−0.02	0.006	< .01
Order × Novel			−0.02	0.006	< .01
1-back accuracy × Item type					
1-back accuracy × Familiar			−1.53	1.51	.31
1-back accuracy × Novel			−1.34	1.49	.36
Age group × Order			−0.008	0.004	.04
Age group × 1-back accuracy			−3.78	1.44	< .01
Order × 1-back accuracy			−0.10	0.03	.003
Age group × Order × Item type					
Child × Order × Familiar			0.01	0.005	.04
Child × Order × Novel			0.01	0.005	.07
Age group × Order × 1-back accuracy			0.11	0.04	.02
Age group × 1-back accuracy × Item type					
Child × 1-back accuracy × Familiar			2.41	1.90	.21
Child × 1-back accuracy × Novel			2.40	1.86	.20
Order by 1-back accuracy × Item type					
Order × 1-back accuracy × Familiar			0.03	0.05	.49
Order × 1-back accuracy × Novel			0.05	0.05	.27
Age group × Order × 1-back accuracy × Item type					
Child × Order × 1-back accuracy × Familiar			−0.02	0.06	.72
Child × Order × 1-back accuracy × Novel			−0.06	0.06	.28

with DLD will show poorer learning than same-age peers. The former was supported. We found that individuals with DLD formed grammatical categories based on distributional information similarly to their TD peers, with no differences across age groups. Clark and Lum (2017), Hsu and Bishop (2014), Ullman and Pierpont (2005), and others have suggested that certain types of statistics may be problematic and other types may be more easily learned for individuals with DLD. Hsu and Bishop (2014), in particular, find deficits for procedural learning of sequence-specific information in children with DLD, but no deficits in a non-sequential task. It is possible that the nonsequential nature of distributional statistics enabled learning. A finding that participants with DLD could use distributional statistics despite well-documented difficulty with conditional statistics is also in line with the hypothesis of different mechanisms supporting different types of learning (Thiessen, 2017; Thiessen et al., 2013), and thus possibly different mechanisms are being impaired or preserved in individuals with DLD. Because success in this distributional learning task requires only that participants form and recall category rules, it is possible that the task plays upon a relative strength in DLD. A study that directly compared performance between

tasks that test distributional and sequence-specific information by the same children, with other variables such as exposure and stimuli characteristics controlled for, would best determine whether the distributional nature of this task specifically enabled learning over and above any other variable.

The finding that the children with and without DLD do not differ is perhaps tempered by the fact that we had to exclude three children, all of whom had DLD, who failed the 1-back task. Of these children, one had the lowest PPVT-4 and SPELT-3 scores, one had the lowest Kaufman Brief Intelligence Test–Second Edition score, and one was in the bottom five for standard PPVT-4 score and one of the youngest participants. The child with DLD who could not complete the artificial grammar learning task and withdrew from the study also had low standardized test scores relative to the DLD group. Although we found that including the three excluded participants in the models did not result in a significant main effect of or interaction with group, it remains possible that these children represent a subset of individuals with DLD for whom distributional learning is also challenging. Visual comparison of the data from the three excluded children reveals no clear pattern, though all differed widely from group results: one participant rated

Figure 4. Influence of 1-back accuracy during training and order items are heard on ratings for adult and child groups, grammatical items only. For visualization purposes, we divided test items into items the first time they heard (the left side of the figure, labeled “1st”) and the second time they are heard (the right side of the figure, labeled “2nd”), though order is included as a continuous variable in our statistical model.

novel items very low compared to the other item types, the second participant rated novel and ungrammatical items similarly, and the third rated novel items much higher than the other items and higher than any other participant group-rated novel items. It would be worthwhile to develop artificial grammar learning tasks that do not require explicit tests of attention and learning to determine whether this subset is a true characterization of the larger DLD population.

Our second research question was whether children with DLD would be slower to update language representations. Individuals across diagnostic groups showed a similar effect of test item order, in that all participants rated grammatical items heard early in the test higher than those heard later. The deterioration of ratings for grammatical items over the course of testing could be due to exposure to ungrammatical items during the test phase, which ultimately changes the distribution of the grammar. Although adults did not show better learning than children in either group, adults showed a larger effect of order for familiar test items than children. We found a similar effect for the TD group over the DLD group, though there was no three-way interaction between age group, diagnostic group, and order. The differences observed likely reflect differences in memory and interference effects among groups. They support Gómez's (2017) hypothesis that adults excel at remembering specific exemplars and children tend to abstract away from exemplars,

even though that was specifically referring to children under the age of 2.

Finally, our third research question was whether developmental differences in distributional learning can be attributed to the ability to attend to or encode linguistic information. First, we found differences in 1-back performance between diagnostic groups and between age groups and a gap in performance between TD and DLD groups that was larger for child groups. It is unsurprising that adults had better attention and/or memory than children during the 1-back task, given that memory is a cognitive resource that improves between childhood and adulthood (e.g., Gathercole, 1998; Luna, Padmanabhan, & O'Hearn, 2010). We also might have expected diagnostic differences given that the 1-back task is essentially a receptive version of a nonword repetition task (tapping phonological working memory), which is well known to be difficult for individuals with DLD, particularly in childhood.

However, we did not find evidence that attention during training affected the ability to form grammatical categories from distributional information. It is remarkable that the child TD and DLD groups do not differ from each other in category learning performance, given the striking differences in auditory attention. This suggests that, for the implicit learning of categories, heightened attention is not necessary, consistent with findings from other statistical

learning tasks with children (Finn, Lee, Kraus, & Hudson Kam, 2014; Janacek, Fiser, & Nemeth, 2012). However, children from both diagnostic groups may not yet have had the attention skills to update category membership, which could translate to some rigidity in the use of syntactic forms compared with more flexibility in adults.

We did see that 1-back accuracy affected the strength of the order effect. For the adults, higher 1-back scores predicted higher ratings for grammatical items at the beginning of the testing phase with a decline over time. This supports the previous finding that adults have increased sensitivity to updating the distributional information at the test phase of the experiment, which might specifically be attributed to the ability to attend to and encode phonological information. An alternative interpretation is that adults are better than children at remembering the items heard during training, and their memory fades rapidly as they hear more items not from the training. However, we saw a similar pattern of falling ratings for novel items, which are not heard during training, suggesting this phenomenon is not just an effect of memory. A final possible interpretation is that the order effect could be attributable to motivation: Adult participants who are the most attentive at training are also the most attentive at test and therefore show the most robust sensitivity to changes in the grammar. 1-Back task accuracy suggests that it is easy and straightforward for adults and more challenging for children, and as such, other factors may have contributed to child accuracy beyond attention (i.e., impulse control or a stronger yes-bias). Although it is difficult to say with certainty why adults showed order effects, we can conclude with confidence that our results indicate sensitivity to distributional information is related to developmental stage, but not diagnostic category.

Relation to Previous Findings and Ideas for Future Research

A previous study of grammatical category learning in participants with DLD showed worse generalization by the DLD group relative to TD peers (Richardson et al., 2006). In our study, we see that individuals with DLD generalize as well as TD peers. Training and test items were quite different across studies. The large number of exposures to a small training set in Richardson et al. (2006) may have perhaps hindered generalization. Reeder et al. (2013) found decreased acceptance of novel items by adults when they lengthened the exposure to the grammar (Experiment 4 in the original article), which they reasoned was because participants began to assume that if they had not yet heard a combination after a longer time, it was because it was not grammatical in the language. A follow-up study that manipulates length of exposure to the same grammar used in this study would allow us to say with certainty whether participants with DLD restrict generalizations to the same degree as TD peers.

Future studies might isolate the aspects of the task that facilitate learning, because a better understanding of

what makes a grammar easy to learn versus difficult for this population could help us control the linguistic environment in treatment for more effective outcomes. For example, the items within categories in this grammar shared no morphological or phonological features to indicate membership. Perhaps this variability aided learning. Factors such as this could be tested in future artificial grammar learning tasks. This research would then inform the design of treatment studies to test the validity of findings in real-world contexts: for example, a treatment study that compares generalization of syntactic frames for a set of target words belonging to the same grammatical category but do not share morphophonological features against a set of words that share morphophonological features.

Conclusions and Implications

This study is the first to show intact distributional learning in children with DLD, with similar performance to age-matched TD peers. It is also the first to show evidence of developmental effects in distributional learning across both groups with DLD and groups with typical language development. These findings, taken in light of an extensive literature showing poorer learning in child and adult groups for conditional statistical learning tasks, support hypotheses of different mechanisms responsible for different types of statistical learning. The finding of a relationship with measures of attention to phonological information for adult groups provides support for statistical learning to play an influential role in language acquisition.

Clinical Relevance

Understanding how individuals with DLD learn grammatical categories has implications for therapy in terms of how clinical input is structured, as Plante and Gómez (2018) put forth in their discussion of principles of learning. Although more specific connections to clinical practice must be speculative at this point, some possible links arise. For example, our finding that category learning from distributional information is a strength rather than a weakness in this population suggests that presenting targets in therapy might best be structured through a distributional set of stimuli rather than individual exemplars. One example of a treatment approach that does this can be found in Hadley and colleagues' work on unique syntactic types. They have suggested that focusing on training a variety of sentence subjects and a variety of predicates may both diversify the syntactic frames that children use and improve grammatical morpheme accuracy (Hadley, McKenna, & Rispoli, 2018; Hadley, Rispoli, & Holt, 2017). Focusing on the category of sentence subject rather than the nonadjacent dependencies associated with agreement marking may capitalize on children's ability to use distributional learning. This suggestion may be somewhat controversial because training morphemes as a category may dilute the total dose for that form and potentially slow progress (Leonard, Camarata, Brown, & Camarata, 2004; Leonard, Camarata, Pawłowska,

Brown, & Camarata, 2006, 2008; Warren, Fey, & Yoder, 2007) or make it less easy for children to identify and acquire the highly regular form (Regularity Principle; Plante & Gómez, 2018). It is notable that the approach promoted by Hadley and colleagues uses full lexical items rather than pronouns. Pronouns are highly frequent and repetitive, whereas lexical items are more likely to have a high degree of variability within the category. The degree of variability may shift the focus from learning the most frequent regular forms to identifying common factors that the category membership share and thereby acquiring grammatical categories such as sentence subject or agent.

Fey, Long, and Finestack (2003) have similarly argued that clinicians should focus on intermediate goals. They suggest that this might include writing clinical goals with grammatical categories (such as tense morphemes in general) as targets rather than particular morphemes (such as past tense *-ed*). Similarly, planning treatment so that it focuses on generalization might be important. Owen Van Horne et al. (2017) demonstrated that training past tense *-ed* with lexical exemplars that are less likely to be heard with *-ed* in everyday speech made it more likely that children would use *-ed* with a wide variety of verbs. However, follow-up analyses also showed that gains from intervention that focused only on past tense *-ed* did not lead to gains in other tense-related morphemes (Owen Van Horne et al., 2018). The computational model in which this work is grounded (Li & Shirai, 2000) would suggest that training the morphemes as a category, rather than as individual exemplars, is more likely to lead to broad-based change. Future work is required to better understand how to best capitalize on preserved distributional learning in therapy, particularly given that language likely draws on both distributional and conditional learning.

Acknowledgments

This work was supported by NIH-NIDCD F31DC015370 awarded to the first author and NIH-NIDCD 5R01DC011742 awarded to the third author. We thank Tim Arbisi-Kelm, Caitie Hilverman, Sarah O'Neill, and Elissa Newport for their help with stimuli creation and Nichole Eden, Danielle Reese, and Dan Plebanek for their assistance with data collection.

References

- 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., 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.
- American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening*. Retrieved from <http://www.asha.org/policy/GL1997-00199/>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- 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(10), 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.
- Clark, G. M., & Lum, J. A. (2017). First-order and higher order sequence learning in specific language impairment. *Neuropsychology*, 31(2), 149–159.
- Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *SPELT-3: Structured Photographic Expressive Language Test*. DeKalb, IL: Janelle Publications.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4)*. Austin, TX: Pearson Assessments.
- Ebbels, S. H. (2005). *Argument structure in specific language impairment: From theory to therapy* (Doctoral dissertation, University of London).
- Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, 105(2), 247–299.
- 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.
- Fey, M. E., Long, S. H., & Finestack, L. H. (2003). Ten principles of grammar facilitation for children with specific language impairments. *American Journal of Speech-Language Pathology*, 12(1), 3–15.
- Fidler, L. J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments [published erratum in: *American Journal of Speech-Language Pathology* 2013; 22: 577]. *American Journal of Speech-Language Pathology*, 20(1), 2–13.
- Finn, A. S., Lee, T., Kraus, A., & Hudson Kam, C. L. (2014). When it hurts (and helps) to try: The role of effort in language learning. *PLoS One*, 9(7), e101806.
- Gathercole, S. E. (1998). The development of memory. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39(1), 3–27.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436.
- Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B*, 372(1711), 20160054.
- 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.
- Hadley, P. A., McKenna, M. M., & Rispoli, M. (2018). Sentence diversity in early language development: Recommendations for target selection and progress monitoring. *American Journal of Speech-Language Pathology*, 27, 553–565. https://doi.org/10.1044/2017_AJSLP-17-0098
- Hadley, P. A., Rispoli, M., & Holt, J. K. (2017). Input subject diversity accelerates the growth of tense and agreement: Indirect benefits from a parent-implemented intervention. *Journal of Speech, Language, and Hearing Research*, 60(9), 2619–2635.
- Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 58(11), 1251–1263.

- Hall, J., Owen Van Horne, A. J., & Farmer, T. A. (2017). Distributional learning aids linguistic category formation in school-age children. *Journal of Child Language*, 45(3), 717–735.
- Hall, J., Owen Van Horne, A. J., McGregor, K. K., & Farmer, T. A. (2017). Distributional learning in college students with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 60(11), 3270–3283.
- 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, 1–10.
- 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). *Kaufman Brief Intelligence Test*. Hoboken, NJ: Wiley.
- 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.
- Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 79.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.
- Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 60(12), 3474–3486.
- Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research*, 47(6), 1363–1379.
- Leonard, L. B., Camarata, S. M., Pawłowska, M., Brown, B., & Camarata, M. N. (2006). Tense and agreement morphemes in the speech of children with specific language impairment during intervention: Phase 2. *Journal of Speech, Language, and Hearing Research*, 49(4), 749–770.
- Leonard, L. B., Camarata, S. M., Pawłowska, M., Brown, B., & Camarata, M. N. (2008). The acquisition of tense and agreement morphemes by children with specific language impairment during intervention: Phase 3. *Journal of Speech, Language, and Hearing Research*, 51(1), 120–125.
- Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin, Germany: Walter de Gruyter.
- Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, 72(1), 101–113.
- Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, 67(3), 130–150.
- McGregor, K. K. (2017, June). *Closing the word-learning gap: Extra encoding opportunities equate long-term outcomes for students with and without developmental language disorders*. Paper presented at the Symposium on Research in Child Language Disorders, Madison, Wisconsin.
- McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2017). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology*, 19(1), 43–57.
- Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsch-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 31–63). New York, NY: Oxford University Press.
- 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.
- Mueller, K. L., & Tomblin, J. B. (2012). Examining the comorbidity of language disorders and ADHD. *Topics in Language Disorders*, 32(3), 228–246.
- Owen Van Horne, A. J., Curran, M. K., Larson, C., & Fey, M. E. (2018). Effects of a complexity-based approach on generalization of past tense *-ed* and related morphemes. *Language, Speech, and Hearing Services in Schools*, 49(3S), 681–693. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0142
- Owen Van Horne, A. J., Fey, M., & Curran, M. (2017). Do the hard things first: A randomized controlled trial testing the effects of exemplar selection on generalization following therapy for grammatical morphology. *Journal of Speech, Language, and Hearing Research*, 60(9), 2569–2588.
- Perona, K., Plante, E., & Vance, R. (2005). Diagnostic accuracy of the Structured Photographic Expressive Language Test—Third Edition (SPELT-3). *Language, Speech, and Hearing Services in Schools*, 36(2), 103–115.
- Plante, E., & Gómez, R. (2018). Learning without trying: The clinical relevance of statistical learning. *Language, Speech, and Hearing Services in Schools*, 49(3S), 710–722. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0131
- 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. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- 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., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 41(6), 1412–1431.
- Richardson, J., Harris, L., Plante, E., & Gerken, L. (2006). Sub-category learning in normal and language learning-disabled adults: How much information do they need? *Journal of Speech, Language, and Hearing Research*, 49(6), 1257–1266.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B*, 372(1711), 20160049.
- Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B*, 372(1711), 20160056.

- Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, 139(4), 792–814.
- Thiessen, E. D., & Pavlik, P. I. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, 37(2), 310–343.
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41(3), 399–433.
- von Koss Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research*, 56(2), 618–629.
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Developmental Disabilities Research Reviews*, 13(1), 70–77.

Appendix (p. 1 of 2)

Training Items

daffin tomber bleggin^a
daffin tomber bleggin gentif^a
klidum daffin tomber bleggin^a
spad daffin tomber bleggin gentif^a
daffin tomber mawg frag
klidum daffin tomber mawg gentif
daffin zub bleggin gentif
klidum daffin zub bleggin
spad daffin zub bleggin frag
daffin zub fluggit
daffin zub fluggit gentif
spad daffin zub fluggit
flairb lapal fluggit
flairb lapal fluggit frag
spad flairb lapal fluggit frag
flairb lapal mawg
flairb lapal mawg frag
klidum flairb lapal mawg
flairb tomber bleggin
klidum flairb tomber bleggin frag
spad flairb tomber bleggin
flairb tomber mawg^a
klidum flairb tomber mawg gentif^a
spad flairb tomber mawg^a
glim lapal fluggit
klidum glim lapal fluggit
spad glim lapal fluggit gentif
glim lapal mawg
glim lapal mawg gentif
spad glim lapal mawg
glim zub bleggin
glim zub bleggin frag
klidum glim zub bleggin frag
glim zub fluggit^a
glim zub fluggit gentif^a
klidum glim zub fluggit^a

^aThe AXB combination in these items did not appear in test items.

Appendix (p. 2 of 2)

Test Items

Grammatical		Ungrammatical	
Familiar	Novel		Subtype
flairb lapal fluggit	flairb tomber fluggit	daffin zub flairb	AXA
glim lapal mawg	daffin tomber fluggit	flairb lapal glim	AXA
flairb tomber bleggin	daffin lapal bleggin	glim zub daffin	AXA
daffin tomber mawg	glim tomber fluggit	daffin tomber frag	AXR
flairb lapal mawg	glim tomber bleggin	flairb zub gentif	AXR
daffin zub fluggit	daffin lapal mawg	glim lapal frag	AXR
glim lapal fluggit	flairb zub mawg	mawg bleggin flairb	BXA ^a
glim zub bleggin	glim zub mawg	bleggin zub glim	BXA
daffin zub bleggin	flairb zub bleggin	fluggit tomber flairb	BXA
		mawg lapal daffin	BXA
		bleggin lapal fluggit	BXB
		fluggit tomber mawg	BXB
		mawg zub bleggin	BXB
		klidum zub mawg	QXB
		spad lapal fluggit	QXB
		spad tomber bleggin	QXB
		tomber mawg glim	XBA
		zub fluggit daffin	XBA

^aThis item was intended as XBA item and occurred by mistake; results do not differ when this item is excluded from analyses.