



ORIGINAL ARTICLE

# Speech production factors and verbal working memory in children and adults with developmental language disorder

Gerard H. Poll<sup>1</sup>  and Carol A. Miller<sup>2,\*</sup> 

<sup>1</sup>Miami University and <sup>2</sup>Pennsylvania State University

\*Corresponding author. Email: [cam47@psu.edu](mailto:cam47@psu.edu)

(Received 13 July 2019; revised 31 December 2020; accepted 08 January 2021; first published online 18 February 2021)

## Abstract

Verbal working memory (VWM) deficits are common in individuals with developmental language disorder (DLD) but are not well understood. This study evaluated how both memory and language production factors influence VWM performance in children and adults with DLD, focusing on the influence of serial position, phonological activation (PA), and lexical frequency. Participants were 30 children with DLD and 26 with typical language, and 21 adults with DLD and 23 with typical language. The participants completed a listening span task in which they were asked to recall the final words of sentences in sets of increasing size. Responses (dependent variable) were coded as correct, incorrect, or no response. Final words were coded for frequency, serial position within the set, and PA (number of occurrences of the initial phoneme, vowel, and whole word in the task). These variables, along with age and language status, were entered as predictors in mixed-effects multinomial regression models. Extreme serial position, greater PA, and higher frequency reduced incorrect and no responses. These effects were attenuated for the DLD group, and the effect of greater PA varied with set size. The findings suggest that for individuals with DLD, VWM performance is affected by more limited effective language experience and by the dynamic task demands.

**Keywords:** developmental language disorder; listening span; verbal working memory

Verbal working memory (VWM) difficulties are among the most consistent deficits observed in developmental language disorder (DLD; Leonard et al., 2007; Montgomery, 2003), and persist as children get older (Leonard, Ellis Weismer, Weber-Fox, & Miller, 2014). However, the nature of the VWM deficits in DLD is not fully understood. In this study, we investigated factors that contribute to both correct and incorrect responses in a VWM task by children and adults with and without DLD. The listening span task provides an intriguing point of contact between theoretical accounts of DLD, influential constructs from psycholinguistic

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

word production models, and classic constructs of memory theory such as primacy and recency. We applied these combined perspectives to explore whether target word frequency, primacy, and recency contribute to VWM performance for children with DLD. In addition, we add to prior work by considering how these factors affect VWM in adults with DLD and how phonological activation (PA) affects VWM responses for both children and adults.

Approximately 7%–11% of kindergarteners present with a language impairment in the absence of hearing impairment, intellectual disability, social–behavioral disorders, or frank neurological impairment (Norbury et al., 2016; Tomblin et al., 1997). This profile (with normal-range nonverbal IQ) has long been called specific language impairment (Leonard, 2014). Recently, the term DLD has been proposed (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE Consortium, 2017). In the present paper, we use DLD, emphasizing the presence of the disorder over an extended developmental trajectory, and the possibility of co-occurring deficits, for example, emotional disorders, problems with attention, speech, or reading (Bishop et al., 2017).

Longitudinal data suggest that the majority of children with DLD continue to have language difficulties beyond childhood (e.g., Clegg, Hollis, Mawhood, & Rutter, 2005; Johnson et al., 1999; Lee & Tomblin, 2015). Therefore, it is important to describe and explain the trajectory of language impairment through adulthood. By understanding both change and consistency as children with DLD grow up, we will not only be able to provide better diagnosis and treatment but also gain a better understanding of the underlying mechanisms and developmental processes involved in language impairment. The present study focused on how accounts of DLD intersect with speech production and memory mechanisms in VWM tasks.

### Theoretical Accounts of DLD, Language, and Working Memory

Theoretical approaches to DLD offer varied accounts of VWM limitations. Domain-general approaches propose that deficits in cognitive abilities such as working memory (WM), processing speed, or inhibition control cause, at least in part, the language deficits observed in DLD (e.g., Leonard et al., 2007; Marton, Kelmenson, & Pinkhasova, 2007). By some accounts, limited VWM contributes to the language processing difficulties of children with DLD (e.g., Baddeley, 2003; Montgomery, 2003; Montgomery, Evans, Fargo, Schwartz, & Gillam, 2018). Domain-specific accounts of DLD, in contrast, posit that language deficits are specific to the language system (e.g., Rice & Wexler, 1996; van der Lely, Rosen, & Adlard, 2004). In domain-specific accounts, VWM and language are separable. There are also many domain-general accounts which, despite according a central role to WM deficits in DLD, view WM as distinct from language (e.g., Archibald, 2017; Montgomery, 2003). Some findings from children with DLD, however, suggest that VWM performance is driven by language knowledge (Mainela-Arnold & Evans, 2005; Mainela-Arnold, Evans, & Coady, 2010).

An emergentist account of DLD suggests that language deficits are the consequence of interactions between affected children's processing limitations and the statistical properties of the language (Evans, 2001). This view focuses on the dynamic nature of children's language abilities. For example, language errors

are more prevalent in demanding contexts and less likely in stable, less demanding contexts. Compared to typical peers, children with DLD require more language exposure to learn statistical information critical to language learning, such as the location of word boundaries (Evans, Saffran, & Robe-Torres, 2009). As a result, individuals with DLD gain less from similar experience with language, or, put another way, individuals with DLD may have less effective language experience as compared to peers with typical language. Thus, influences on speech production derived from language experience (e.g., word frequency) may affect the VWM performance of each group differently.

Both *multicomponent* and *emergent* models of VWM have been proposed (Schwering & MacDonald, 2020). Models from both perspectives include mechanisms by which VWM interacts with knowledge of language in long-term memory. Multicomponent models posit that long-term memory is separable from WM and that processing of information is distinct from passive storage (e.g., Baddeley, 2003; Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Case, Kurland, & Goldberg, 1982; Gathercole & Baddeley, 1993). Mechanisms by which language knowledge affects VWM include encoding and redintegration (Hulme et al., 1997; Martin, Lesch, & Barta, 1999; Schweickert, 1993). Encoding, an early stage process where a to-be-remembered word creates a short-term memory trace, is aided by stronger long-term representation of that word (Martin et al., 1999). VWM models where the short-term storage is an activated part of long-term memory make a similar claim for the influence of long-term memory (Cowan, 1988, 1995). Redintegration is a later-stage process whereby a partially degraded memory trace may be reconstructed by accessing the long-term representations of the word (Hulme et al., 1997; Schweickert, 1993).

Emergent models of VWM posit that processing and storage are not different mechanisms and that long-term and short-term memory are not distinct (e.g., Cowan, Li, Glass, & Saults, 2018; Kowialiewski & Majerus, 2018). MacDonald and colleagues (Acheson & MacDonald, 2009; MacDonald & Christiansen, 2002; Schwering & MacDonald, 2020) have argued for a strongly emergent conceptualization where VWM is not a separate system from language. Acheson and MacDonald (2009) suggested that serial position effects, word frequency effects, and the influence of phonological similarity on recall are found in both language production and VWM tasks, and result (emerge) from the same mechanisms.

Encoding, redintegration, and emergent accounts suggest that VWM limitations may involve both memory and language production factors. The influence of language production factors on VWM has been extensively studied in the aphasia literature (e.g., Foygel & Dell, 2000; Martin & Saffran, 1997; Martin et al., 1999), and in typical adults (Hulme, Roodenrys, Brown, & Mercer, 1995; Hulme et al., 1997; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). Similarly, data bearing on the relations between language knowledge, VWM, and the dynamics of task demands in individuals with DLD may help to guide theory development for this population.

## Listening span

In order to evaluate WM, researchers often use complex span tasks in which the participant must remember stimuli while also performing a processing task

(Jarrold, 2017; Marton, Eichorn, Campanelli, & Zakarias, 2016). One type of complex span is a listening span task (an auditory version of the reading span task, e.g., Just & Carpenter, 1992). The participant hears a sentence and judges whether it is true (Gaulin & Campbell, 1994). After a set of sentences, the participant is asked to recall the final word of each sentence (or the *target words*). The number of sentences in a recall set increases as the task progresses.

Listening span tasks engage the speech and language processing system (Allen & Hulme, 2006). Other complex span tasks use closed sets (digits and letters); therefore, the possibilities for recalling incorrect items are limited. In contrast, listening span requires the storage and recall of words, creating an opportunity for recall of incorrect words that may share properties with target words. Listening span has been used to investigate VWM in children with DLD (Ellis Weismer, Evans, & Hesketh, 1999; Mainela-Arnold & Evans, 2005; Mainela-Arnold et al., 2010; Marton & Eichorn, 2014; Marton et al., 2007; Montgomery & Evans, 2009). Children with DLD consistently recall fewer words despite demonstrating comprehension of the distractor sentences that is comparable to peers with typical language (TL). Some of these studies (Ellis Weismer et al., 1999; Marton & Eichorn, 2014; Marton et al., 2007) have included error analyses, comparing group means of different types of errors. Mainela-Arnold and Evans (2005) followed the emergent account of VWM (MacDonald & Christiansen, 2002) to investigate effects of word frequency and serial position on word recall in children with DLD. In the present study, we built on this body of previous work by considering additional predictor variables, predicting different types of responses, applying different statistical models, and including adults with DLD. Our aims are to better specify how VWM memory limitations may result from differences in how individuals with DLD respond to lexical and memory factors, and how those response profiles change with development.

## Influences on verbal WM performance

### **Serial position effects**

A word's position in a list affects how likely it is to be recalled (e.g., Glenberg et al., 1980; Greene, 1986; Page & Norris, 1998; Sheng, Byrd, McGregor, Zimmerman, & Bludau, 2015; Tan & Ward, 2000). Words presented early or late in the list are recalled more accurately than those in the middle. These effects are called *primacy* and *recency*, respectively. In a listening span task, the lists of words to be recalled are interspersed with distractor sentences; however, serial position effects are found in tasks that include distractors (e.g., Glenberg et al., 1980).

Individuals with DLD may not differ qualitatively from TL peers in serial position effects. Both young adults (Sheng et al., 2015) and children (Majerus et al., 2009) with DLD showed similar primacy and recency effects to TL peers when recalling lists. Mainela-Arnold and Evans (2005) found recency but not primacy effects in children with and without DLD; however, they stated that the recency effect seemed "somewhat heightened" in the DLD group. Gillam, Cowan, and Marler (1998), however, found that children with DLD benefited less from recency than a control group. Several accounts of serial position effects have been proposed. Acheson and MacDonald (2009) suggested that an account invoking the temporal

distinctiveness of words at the beginning and ending of a list is most compatible with integration of language production with WM. Others have proposed that serial order of memoranda provides context cues, which make early and late items in a list more distinct and more likely to be selected from competing candidates for recall (Burgess & Hitch, 2006; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016). Gillam and colleagues (1998) suggested that children with DLD may less effectively encode incoming phonological information into short-term memory, attenuating their ability to benefit from recency. Regardless of the precise mechanism, if serial position increases activation of a word, it could affect recall responses in a VWM task.

### **Phonological activation**

Here, we follow the broad outlines of Levelt's speech production model (Levelt, Roelofs, & Meyer, 1999) involving semantic, lexical, and phonological levels of representation, but assume that activation can feed back from the phonological level "up" to the lexical level, as suggested by Foygel and Dell (2000). In this type of word production model, a meaning is intended. A lexical item is selected to express that meaning, and then a phonological form is retrieved to instantiate the lexical item. Speech production models identify factors influencing whether speakers produce intended and unintended words; these same factors are likely to bear on production of correct and incorrect responses in VWM tasks.

Foygel and Dell (2000) described how word production errors can arise as a result of spreading activation, which occurs within and between levels of their speech production models (from the lexical level to the phonological level, and back "up" to the lexical level). Within the phonological level, the models distinguish between word onsets, vowels, and codas, a distinction that we follow in our analyses. Multiple words and sounds may be activated during the process of production, creating an opportunity for errors if a word other than the target is more highly activated.

Phonological activation may also facilitate word recall due to compressibility. Studies of nonverbal WM (Chekaf, Gauvrit, Guida, & Mathy, 2018; Mathy, Chekaf, & Cowan, 2018) show that longer sequences of items can be recalled if their complexity can be simplified by identifying features shared among the items. In a verbal complex span task, phonological activation may enhance encoding and thus the ability to recognize phonological patterns within and among target words, aiding compressibility. Recognizing such phonological patterns in complex span tasks may require the ability to effectively switch attention between the processing task and compression (to aid storage). Studies of typical adults have found that compression processes may be fairly automatic (Mathy et al., 2018) whereas they may be more demanding of attentional resources for children with DLD (Montgomery et al., 2018).

Evidence regarding the structure of the word production system in individuals with DLD is limited. Several studies indicate that the underlying architecture is similar for children with and without DLD, although systems of children with DLD may not operate as efficiently (e.g., Brooks, Seiger-Gardner, Obeid, & MacWhinney, 2015; Mainela-Arnold, Evans, & Coady, 2008, 2010; Seiger-Gardner & Brooks, 2008; Seiger-Gardner & Schwartz, 2008). Brooks et al. (2015) found that children with DLD were more time limited in their ability to use phonological priming to

support word production compared to typical peers. Mainela-Arnold et al. (2008) found that lexical access in children with DLD was more vulnerable to competition from other words.

### **Frequency effects**

Word frequency is influential in many tasks requiring the processing and production of words (e.g., Gagnon, Schwartz, Martin, Dell, & Saffran, 1997; Leonard, Nippold, Kail, & Hale, 1983). Higher frequency of a word in a given language tends to facilitate its production (Jescheniak & Levelt, 1994). Higher frequency words are more likely to be recalled by both children with DLD and children with TL (Leonard et al., 1983; Mainela-Arnold & Evans, 2005; Mainela-Arnold et al., 2010). For typical young adults, Hulme et al. (1997) found that word spans are greater for high-frequency words, and Allen and Hulme (2006) found that higher frequency words are more likely to be recalled from a list. Furthermore, among responses that were not correct, there were fewer omission errors and more phonological approximation errors for high-frequency words. The effect of word frequency on recall has been attributed to reintegration. High-frequency words are thought to be more easily retrieved to support restoration of degraded memory traces (Hulme et al., 1997; Schweickert, 1993). Given its role in speech production and memory, word frequency must be considered as a predictor of listening span performance.

### **Integrated effects of factors: An example**

To make the contributions of these influences on VWM performance more concrete, consider the word *fly*. In our data, it was one of the words children were most likely to produce in error. In the Competing Language Processing Test (CLPT; Gaulin & Campbell, 1994), used in the present study, *fly* first occurs in the second set as a target word in *Trains can fly*. It occurs again in the seventh set in *Birds can fly*, and in the eleventh set in *Airplanes can fly*. Thus, *fly* is activated as a target word multiple times during the task. In addition, a word beginning with /f/ occurs as the first word in a distractor sentence four times during the task. Thus, feedback from the phonological level to the lexical level of the Foygel and Dell (2000) models would tend to increase the activation of *fly*. Although it does not have a consistently early or late serial position within sets, the repeated occurrence of *fly*, its phonological similarity to other words, and its relatively high frequency among the top one-third most frequent words in the task may conspire to keep *fly* activated enough to be produced in place of other target words, yielding an intrusion error.

### **No response and uncertainty**

The largest category of errors across all age and clinical groups is failure to produce a word at all, or “no response” (Marton et al., 2007). In the context of mechanisms of activation and competition, presumably no response occurs when none of the candidate words has reached a threshold of activation necessary for production. When the activation of one or more words hovers around threshold levels, the speaker faces uncertainty and must choose whether to respond, possibly in error, or to

withhold response. Uncertainty engages metacognitive monitoring and is demanding of WM resources (Coutinho et al., 2015). When the size of a listening span set is within the VWM capacity of an individual, there is little or no uncertainty and most responses are correct. As the task continues, uncertainty grows, because the number of memoranda increases, taxing VWM capacity, and more and more words are activated, increasing the number of competitors. At this point, incorrect responses have the most opportunities to “win” the competition. As set size increases further and VWM capacity is exceeded, again there is little or no uncertainty but now the individual is certain that they do not remember the target word; therefore, no response becomes more prevalent.

A way to better understand how the limitations of individuals with DLD affect VWM task performance is to evaluate factors influencing the kinds of response, both incorrect words and no responses. We sought to understand whether speech production factors affected both children and adults, and whether the influence of these factors differed for individuals with and without DLD.

### Questions and Predictions

1. Do children and adults with DLD differ in the probability of correct recall versus no responses and incorrect responses on a listening span verbal WM task compared to their same-age TL peers? In line with previous findings, we predict that children with DLD will perform more poorly than their TL peers; however, less is known about the VWM performance of adults with DLD. The persistence of DLD into adulthood suggests that adults with DLD will also perform more poorly than their TL peers.

For the following questions we assume that language representation in long-term memory affects VWM performance (Acheson & McDonald, 2009; Mainela-Arnold & Evans, 2005; Martin & Saffran, 1997; Roodenrys et al., 2002). Therefore, we expect serial order and PA effects to be weaker in children than in adults, and in DLD groups compared to TL groups, on the assumption that children and individuals with DLD have less effective language experience, and therefore weaker language representations. However, there is some evidence to predict that weaker representations may exaggerate the effect of high-frequency words (e.g., Mainela-Arnold & Evans, 2005). We expect that serial order, PA, and frequency effects may not be uniform across correct responses, no responses, and incorrect responses.

2. Does serial order of memoranda predict the probability of correct recall, no responses, and incorrect responses, and does it interact with age and clinical status? We expect serial order effects to be present for adults and children with and without DLD (Majerus et al., 2009; Sheng et al., 2015), but to be weaker for children and for individuals with DLD (Gillam et al., 1998).
3. Does PA predict the probability of correct recall, no responses, and incorrect responses for children and adults at different set sizes, and does it interact with clinical status? Correct recall should be more likely for target words with greater PA. PA may interact with group. If individuals with DLD have more limited inhibition control and more difficulty resolving competition, they will

be more likely to produce words that are more phonologically activated, either correct target items or incorrect intrusion errors (Marton et al., 2007). PA may also interact with age. The models of word production that we based our analyses on were generated to account for adult behavior; it remains to be seen if the phonological properties we coded influence children similarly. Finally, the effects of PA may vary by set size. If PA is influential as the participant approaches their capacity limit, it is likely to be less influential as their capacity is exceeded, when they become certain that they do not recall many of the target words.

4. Does word frequency predict the probability of correct recall, no responses, and incorrect responses, and does it interact with age and clinical status? While frequency effects are expected for typical adults (Hulme et al., 1997; Roodenrys et al., 2002), there have been mixed results on whether frequency effects are similar for individuals with DLD and their TL peers (Leonard et al., 1983; Mainela-Arnold & Evans, 2005; Mainela-Arnold et al., 2010). Frequency effects may differ for children compared to adults, as children have had less exposure to words overall than adults, and individuals with DLD may be affected differently due to less effective language experience.

## Method

### *Participants*

The study included 56 children (mean age 10 years) and 44 adults (mean age 22 years) with DLD or TL and whose first language was English. Participants with a history of autism, intellectual disability, hearing loss, significant neurological injury, or cerebral palsy were excluded. All participants passed a hearing screening at 25 dB HL at the speech frequencies. Data from the children with DLD were previously reported in Miller and Wagstaff (2011). The children with TL were drawn from a participant pool that has been reported in Mainela-Arnold, Misra, Miller, Poll, and Park (2012) and Poll et al. (2013). Data from the adults with DLD and TL have been reported in Poll, Miller, and van Hell (2015, 2016), and Poll, Watkins, and Miller (2014).

### *Child sample*

The child sample is summarized in Table 1. Thirty clinically referred children were classified as having DLD by five language measures. Receptive vocabulary was assessed with the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), and expressive vocabulary was assessed with either the Expressive Vocabulary Test (Williams, 1997) or the picture vocabulary subtest of the Woodcock-Johnson Tests of Achievement (3rd ed.; Woodcock, McGrew, & Mather, 2001). Each vocabulary measure yielded a standard score ( $M = 100$ ,  $SD = 15$ ). Receptive and expressive syntax were assessed using the Concepts and Following Directions and Formulating Sentences subtests, respectively, of the Clinical Evaluation of Language Fundamentals (CELF-4; Semel, Wiig, & Secord, 2003). The subtests yielded scaled scores ( $M = 10$ ,  $SD = 3$ ). The fifth language measure was the nonword



**Table 1.** Means (standard deviations) for child sample

Measure	Group	
	DLD	TL
<i>N</i>	30	26
Age (years)	10.3 (1.2)	10.5 (1.7)
PIQ (standard score)	92.4 (12.1)	108.2 (13.6)
CELF-4 C & FD (scaled score)	5.5 (2.7)	10.9 (1.6)
CELF-4 FS (scaled score)	8.1 (2.8)	12.3 (2.3)
Expressive vocabulary (standard score)	84.9 (9.1)	–
Receptive vocabulary (standard score)	92.2 (11.5)	–
Nonword Rep (proportion correct)	0.72 (0.08)	–
CLPT recall (proportion correct)	0.55 (0.11)	0.66 (0.15)

Note: DLD, developmental language disorder. CELF-4, Clinical Evaluation of Language Fundamentals (4th ed.). C & FD, Concepts and Following Directions. FS, Formulated Sentences. CLPT, Competing Language Processing Test.

repetition test (NRT; Dollaghan and Campbell, 1998). Norms are not available for the NRT, but based on previous research (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000) a cutoff was set at 75% phonemes correct. Children were classified as having DLD if they scored 1 *SD* below the mean (or below the NRT cutoff) on at least two of the five measures, or if they scored 2 *SD* below the mean on at least one measure other than the NRT. Two children met criteria on all five measures, 5 children met criteria on four measures, 6 children met criteria on three measures, and 14 children met criteria on two measures. Three children qualified on the basis of one measure; these 3 children received a scaled score of 3 or less on Concepts and Following Directions subtest. Low scores were observed on all tests. There were 25 scores below cutoff on Concepts and Following Directions, 13 below cutoff on Formulating Sentences, 15 below cutoff on expressive vocabulary, eight below cutoff on receptive vocabulary, and 18 below cutoff on NRT. Performance IQ (PIQ) was measured using the Abbreviated Battery of the UNIT (Bracken & McCallum, 1998). All participants in the DLD group had a  $PIQ \geq 72$  (25 out of 30 had a  $PIQ \geq 85$ ).

Children in the TL comparison group were recruited from the community. The 26 children were selected from a larger pool to form a sample similar in age to the group with DLD. They completed the Concepts and Following Directions and Formulating Sentences subtests of the CELF-4, receiving scaled scores of 7 or higher (within 1 *SD* of the mean or higher). PIQ was measured using the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). All participants in the group with TL had a  $PIQ \geq 77$  (25 out of 26 had  $PIQ \geq 89$ ). The two groups differed significantly on PIQ,  $t(50) = 4.6, p < .001$ , although the comparison is questionable, as different tests were used for the groups.

#### Adult sample

Adults were recruited at postsecondary schools and from a database of participants who had been recruited for studies of DLD in Iowa. All participants had PIQs of 75

**Table 2.** Means (standard deviations) for adult sample

Measure	Group	
	DLD	TL
<i>N</i>	21	23
Age (years)	22.4 (2.0)	21.5 (1.8)
PIQ (standard score)	97.5 (7.9)	113.7 (10.0)
CELF WD (raw score)	23.0 (7.7)	37.4 (4.8)
CLPT recall (proportion correct)	0.71 (0.11)	0.88 (0.09)
Running span (maximum list recall)	3.19 (0.62)	4.19 (0.73)

Note: DLD, developmental language disorder. PIQ, performance IQ. CELF-4, Clinical Evaluation of Language Fundamentals. WD, word definitions. CLPT, Competing Language Processing Test.

or above as measured by the Wechsler Adult Intelligence Scale (Wechsler, 1997). Characteristics of the adult groups are summarized in Table 2. The mean PIQ differed between groups,  $t(42) = 5.92, p < .001$ .

Participants meeting screening criteria were classified as having DLD or TL by history and by testing. Those with a positive history of language difficulties (diagnosis of DLD, spoken grammar difficulties, or reading comprehension difficulties) were eligible for the group with DLD; those with a negative history were eligible for the group with TL. Testing combined the Modified Token Test (Morice & McNicol, 1985), a 15-word spelling task, and word definitions from the CELF-4 (standard scores; Semel et al., 2003) as outlined in Fidler, Plante and Vance (2011). Their process had a sensitivity (78%) and specificity (83%), the best accuracy of then known approaches for identifying adults with DLD. Scores were entered into a discriminant function. Those with results in the positive range who also had a positive history were classified as DLD; those with results in the negative range with a negative history were classified as having TL.

In both adults and children, the mean PIQ score was significantly lower in the DLD groups. Such differences are frequently found between samples of individuals with DLD and TL (Fidler, Plante, & Vance, 2011; Gallinat & Spaulding, 2014). When PIQ is an inherent characteristic of a disorder, statistically controlling it complicates rather than clarifies explanation, as discussed by Dennis et al. (2009). In the present study, we restrict generalizations to a phenotype similar to that of our sample.

### Measures and procedures

#### Verbal Working Memory Task

The CLPT (Gaulin & Campbell, 1994) was used to assess VWM. The CLPT is a listening span test that requires the participant to listen to sets of simple sentences (e.g., *Sugar is sweet; Apples are square*) and judge the truth of each sentence by responding “yes” or “no.” The participant is then asked to recall the last word of each sentence in the set (referred to as *target words*). Set size increases from one to six sentences. Practice items were included in determining the PA of target words

because all items affected the level of PA for target words later in the task. Thirty-three percent of target words were verbs (all uninflected), 24% were adjectives, and 43% were nouns, and of the nouns, 75% were regular plurals, 5% were irregular plurals, and 20% were mass nouns. The stimuli were recorded by a female speaker on a Marantz PMD650 minidisc recorder using a head-mounted microphone.

The CLPT was presented at a comfortable loudness from a digitized file under headphones for adults, and using the device's speakers for children. Although the CLPT was designed for use with children, in this study variability in recall performance among the adults was adequate for analyses to be conducted. The truth judgment portion of the task is intended to be easy for all participants; both adults and children averaged 96%–99% correct.

Participant responses to the CLPT were recorded during the task. An audio recording of the task was used for ensuring the accuracy of the response record. Minor morphological variations of target words were accepted as correct, for example “wheel” for “wheels.” In recording incorrect word productions, we classified those that were semantic in origin versus phonological. Finding a very small proportion that were semantic errors (6%), we focused on phonological factors in response errors.

#### *Serial position*

Target words occurred in sets requiring from one to six target word recall responses. We represented primacy and recency of target words by coding first words as “1,” second words as “2” and more interior words as “3.” Final words in sets were coded as “5” and penultimate words as “4.” Sets with three items were coded “1, 3, 5” and sets with four items were coded “1, 2, 4, 5.” To understand serial position effects, we contrasted coded two variables to evaluate effects of serial position. The first contrasted extreme position (1, 2 or 4, 5) to interior positions (3); the other contrasted recency (4, 5) to primacy (1, 2). Where participant responses did not clearly align with target words, we eliminated the data from analyses of serial position.

#### *Phonological activation*

To represent the level of phonological activation for each target word, we summed the number of times that the word's phonological elements had been encountered prior to the point in the CLPT where the participant was to recall that target. In models of speech production (Foygel & Dell, 2000), activation is a function of the excitation of the word's phonological form or excitation of the critical segments of the word. For each point where a participant was to produce a target word, we counted the number of times that the participant heard the entire target word, its initial phoneme, or stressed vowel previously in the task. We added instances when the participant uttered the word; these values could vary across individuals. For example, when “feet” is a target word to be recalled in a set, there was 1 prior instance of hearing the entire word, 7 prior instances of hearing the initial “f” phoneme, 11 instances of hearing the /i/ vowel, and no instances when the participant had uttered “feet” previously. We therefore entered a phonological activation level of 19 for that target word. To support computerized counting of phonemes, we

recoded the CLPT words into the CMU Phonetic Dictionary form (see <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>). Both authors independently calculated phonological activation levels; discrepancies were resolved by consensus.

### *Frequency*

To represent the frequency of each target word in the CLPT, we entered the Log of the SUBTLex corpus frequency from the English Lexicon Project (Balota et al., 2007). This is a measure of the frequency of occurrence based on the subtitles for movies. Frequencies were obtained for lexemes, and were not constrained by syntactic category.

## **Analysis approach**

### *Item-level analyses*

Our research questions address how characteristics of target words affect item-level response types. As participants provided responses for sets of target words, we made adjustments to account for any ambiguity in the alignment of response types to target words. In cases where participants made correct responses, those were aligned to the matching target word. If a set contained only no-response errors or one error, then the error responses were aligned to the target words, for which there was no correct response. For sets with multiple incorrect word responses, or a mix of incorrect word and no response errors, then the measures of target word characteristics for those responses were recoded to the mean of the target word measures for the incorrect responses in the set. For 94% of all items, the alignment of the response was clearly aligned to a target word. The same process was used for PA and frequency.

### *Bayesian mixed-effects modeling*

We conducted item-level analyses of the dependent variable, response type, coded as correct, incorrect word, or no response. Predictors were the target word serial position, PA, and frequency and participant language ability group (DLD or TL) and age group (adult or child). To account for repeated responses by participant and by item, we used mixed-effects regression models including random effects for participants and items.

We used Bayesian regression models to complete the analyses. Bayesian models are recommended for obtaining unbiased parameter estimates for categorical, non-normal, dependent variables, particularly when the data are unbalanced (more no responses than incorrect word responses; von der Malsberg, 2016; Zhao, Staudenmayer, Coull, & Wand, 2006). Bayesian analysis involves selecting the probability model, computing the posterior distribution, and determining the fit and convergence of the models (Nalborczyk, Batailler, Loevenbruck, Vilain, & Burkner, 2019). The probability model was a multinomial logit-link function, similar to logistic regression, available in the MCMCglmm package (Hadfield, 2010) for R (R Core Team, 2018). The models linked predictors to the likelihood of no response or incorrect word responses as compared to correct responses. Bayesian models use data to update the prior information known about parameters to the

posterior distribution, the parameters given the data. As we did not have previous data on which to base a prior, we used proper, minimally informative priors for random effects and residuals as recommended for multinomial (or “categorical”) models by Hadfield (2010). The prior distribution for fixed effects had a mean of zero and a large variance ( $10^8$ ) in order to minimally constrain the model estimations.

Monte Carlo Markov chain (MCMC) simulations involve generating large numbers of samples of the distribution of the parameters of interest. The number of samples varied depending on the iterations required for the model to converge. The mean and the 2.5 and 97.5 percentile range indicate the highest posterior density interval (HPDI) for each parameter, similar to 95% confidence intervals for conventional statistical models. The simulations also generate the *pMCMC*, the probability that the parameter estimate includes zero, or no effect.

Model diagnostics for MCMC simulations focus on whether the model converges on a stable set of estimates (Hadfield, 2010). We assessed convergence by assessing graphs of the parameter estimates as they vary with iterations of the model runs: no clear trends indicate convergence. We also assessed convergence using Gelman–Rubin diagnostics (Brooks & Gelman, 1998), which produce a potential scale reduction factor (PSRF) from running two simulations of the same model. The PSRF indicates how closely the two simulations arrive at the same parameter estimates. If a PSRF was less than 1.1, the model was deemed to have converged (von der Malsberg, 2016). Model parsimony was evaluated with deviance information criteria (DIC; Hadfield, 2010). Smaller DICs are preferred and indicate whether the complexity of adding predictors is offset by improving the fit of the model to the data.

As recommended by Hadfield (2010), the intercept was suppressed in models in order to evaluate the effect of the predictor of interest on the likelihood of an incorrect or no response compared to a correct response. We produced separate models for serial position, PA, and word frequency to ease model convergence and to improve the interpretability of models. Before presenting these separate models, we evaluated whether each of these factors systematically varied with the others. The analysis approach for our first question, whether groups differed in correct recall, differed from these models of item-level effects. As it involved participant means without repeated measures, we conducted a between-groups analysis of variance.

## Results

Our first question was whether participants with DLD would recall smaller proportions of target words than their TL peers. The correct recall means (*SD*) for participants with DLD were 0.55 (0.11) (children) and 0.72 (0.11) (adults), as compared to 0.66 (0.15) and 0.88 (0.09) for children and adults with TL. Recall data were analyzed in a 2 (age)  $\times$  2 (language ability) analysis of variance. Adults recalled more target words than children,  $F(1, 96) = 62.3, p < .001$ , partial  $\eta^2 = .393$ . Participants with TL recalled more words than those with DLD,  $F(1, 96) = 30.5, p < .001$ , partial  $\eta^2 = .241$ . Age did not interact with language ability,  $F(1, 96) = 0.84, p = .36$ , partial

$\eta^2 = .009$ . The absence of a significant interaction indicates both adults and children with DLD recalled fewer target words than peers with TL.

Our other research questions concerned how speech production factors affected response types. For children and adults the percentages of response types by set size and language ability group are in Appendix A. For combined language groups, the percentage of incorrect word responses peaked at set size 3 for children, and set size 5 for adults. The percentage of no response increased from set size 1 to 6 for both age groups.

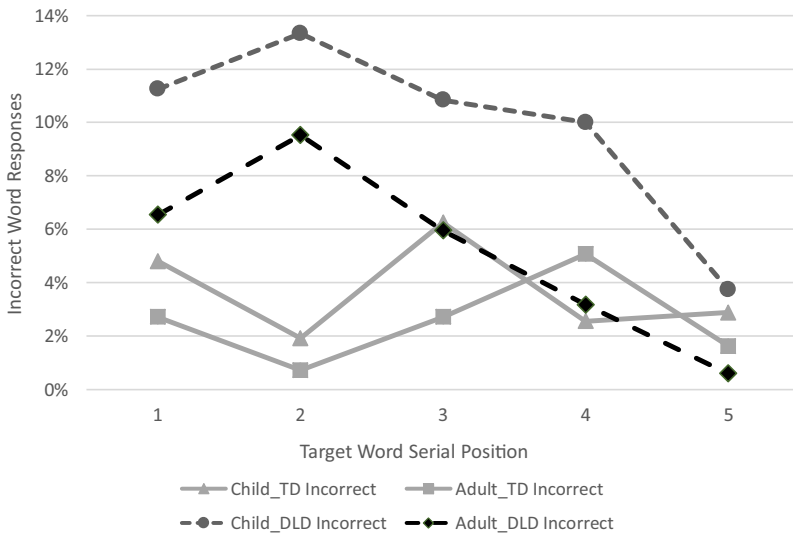
### **Multifactor models**

Before considering models focused separately on serial position, PA, and frequency, we evaluated whether these factors varied systematically with each other. We found little evidence that either PA or target word frequency varied systematically with extreme serial position. Multifactor models indicated that effects of individual factors were not likely to be artifacts of the relation of the factors to one another. Details of the analyses are in Appendix B.

### **Serial position**

Our second question was whether target word serial position affected response type. As there were interior serial positions from set sizes 3 to 6, we modeled responses from these set sizes. The first model included the extreme versus interior position contrast, the interactions of extreme position with age (coded 0 for child, 1 for adult) and language group (coded 0 for TL, 1 for DLD), set size (as a control variable), and random effects by participant and by item. More extreme serial position decreased the likelihood of no response (posterior mean;  $M_p = -1.08$ , HPDI [-2.08, -0.07],  $pMCMC = .04$ ) and of incorrect word responses ( $M_p = -3.93$ , HPDI [-5.08, -2.85],  $pMCMC < .001$ ). Extreme position and group did not interact for no response ( $M_p = -0.57$ , HPDI [-1.30, 0.16],  $pMCMC = .13$ ) but did for incorrect word responses. For the group with DLD, the effect of extreme serial position was attenuated ( $M_p = 3.18$ , HPDI [1.92, 4.30],  $pMCMC < .001$ ). For no response extreme serial position did not interact with age ( $M_p = -0.48$ , HPDI [-1.33, 0.46],  $pMCMC = .29$ ) nor was there a three-way interaction of extreme position, age, and group ( $M_p = 1.11$ , HPDI [-0.04, 2.28],  $pMCMC = .06$ ). For incorrect word responses, extreme position interacted with age ( $M_p = 3.18$ , HPDI [1.92, 4.30],  $pMCMC < .001$ ) and there was a three-way interaction of extreme position, group, and age ( $M_p = -2.54$ , HPDI [-4.49, -0.66],  $pMCMC = .005$ ). Target words in extreme list positions were less likely to have incorrect word or no response errors, and for incorrect responses that effect was attenuated for participants with DLD. The three-way interaction suggests that children with DLD differed in their response to extreme position as compared to other groups.

We next evaluated the same model but with the contrast of recency to primacy. Recency reduced the likelihood of both error responses as compared to primacy (no response,  $M_p = -1.17$ , HPDI [-2.19, -0.14],  $pMCMC = .02$ ; incorrect word responses,  $M_p = -4.05$ , HPDI [-5.38, -2.74],  $pMCMC < .001$ ). Recency interacted with group, with a modestly heightened effect for no response ( $M_p = -0.79$ , HPDI



**Figure 1.** Incorrect word responses as a percentage of total responses by target word serial position by participant group. TD, typical development. DLD, developmental language disorder.

$[-1.57, -0.03]$ ,  $pMCMC = .05$ ) but an attenuated effect for incorrect word responses ( $M_p = 2.06$ , HPDI  $[0.74, 3.34]$ ,  $pMCMC = .002$ ). The three-way interaction of recency, age, and group was not significant for no response ( $M_p = -0.06$ ,  $[-1.36, 1.11]$ ,  $pMCMC = .92$ ) but was for incorrect words ( $M_p = -4.72$ , HPDI  $[-6.90, -2.55]$ ,  $pMCMC < .001$ ). This interaction suggests that adults (coded 1) with DLD (coded 1) benefited more from recency than other groups. Figure 1 suggests that both groups with DLD benefited from primacy less than groups with TL, and that adults with DLD benefited more from penultimate position than did other groups.

### Phonological activation

Our next question was whether PA affected response type, and if the effect differed by age and language ability. Because PA varied systematically with set size, we modeled by set size from set size 3, where children's proportion of incorrect word responses peaked, to set size 6 where the proportion of no response was the largest for both groups. Models included PA, the PA  $\times$  Group interaction and random effects for participants and items.

Results for children are in Table 3. At set size 3, the negative coefficients for PA indicate that higher target word PA decreased the likelihood of no response and incorrect word responses. PA interacted with group for incorrect word responses at set sizes 3 and 4. The positive coefficients indicate that the benefit of higher PA was attenuated for children with DLD. At set sizes 4 and 5, higher PA reduced the likelihood of incorrect word responses. PA had no significant effect at set size 6.

Results for adults are in Table 4. Across all set sizes, higher target word PA reduced the likelihood of incorrect word responses. For all but set size 6, higher PA also reduced the likelihood of no response. For incorrect word responses, the

**Table 3.** Models evaluating the effect of phonological activation on response type for children

Factor	Response type	Posterior mean	HPD interval	<i>pMCMC</i>
<i>Children, Set size 3</i>				
PA	No response	-0.46	[-0.68, -0.24]	<.001
PA	Incorrect word	-0.60	[-0.85, -0.36]	<.001
PA × Group	No response	0.17	[-0.20, 0.54]	.34
PA × Group	Incorrect word	0.47	[0.05, 0.88]	.02
<i>Children, Set size 4</i>				
PA	No response	-0.07	[-0.21, 0.08]	.35
PA	Incorrect word	-0.31	[-0.46, -0.15]	.001
PA × Group	No response	-0.01	[-0.20, 0.18]	.92
PA × Group	Incorrect word	0.29	[0.07, 0.52]	.01
<i>Children, Set size 5</i>				
PA	No response	-0.05	[-0.17, 0.07]	.40
PA	Incorrect word	-0.24	[-0.37, -0.12]	<.001
PA × Group	No response	0.00	[-0.17, 0.17]	.98
PA × Group	Incorrect word	0.18	[-0.01, 0.40]	.63
<i>Children, Set size 6</i>				
PA	No response	0.03	[-0.07, 0.13]	.49
PA	Incorrect word	-0.09	[-0.19, 0.02]	.10
PA × Group	No response	0.03	[-0.13, 0.018]	.76
PA × Group	Incorrect word	0.07	[-0.10, 0.24]	.44

Note: PA, phonological activation. HPD, highest posterior density. *pMCMC*, probability that the parameter estimate includes zero

effect of PA interacted with language ability group for set sizes 4 and 5. The positive coefficients indicate that the effect of PA on adults with DLD was attenuated.

Interactions of PA with group by set size are shown in Figure 2. The effect of PA on the likelihood of incorrect word responses interacted with group at set sizes that were at or near the highest proportions of incorrect word responses across set sizes. The interaction term was positive, whereas the PA main effect was negative. Language ability group was coded as “0” for TL and “1” for DLD, so interaction terms reflect how the effect of PA differed for those with DLD. Higher levels of PA generally reduced the likelihood of an incorrect word response, but that effect was attenuated for participants with DLD. Higher PA also reduced the likelihood of no response for both ages at set size 3, and for adults only for set sizes 4 and 5.

### **Word frequency**

Our final question was whether the target word frequency affected response type. The mean (*SD*) log SUBTLex frequency for target words in the CLPT was 3.26 (0.65)



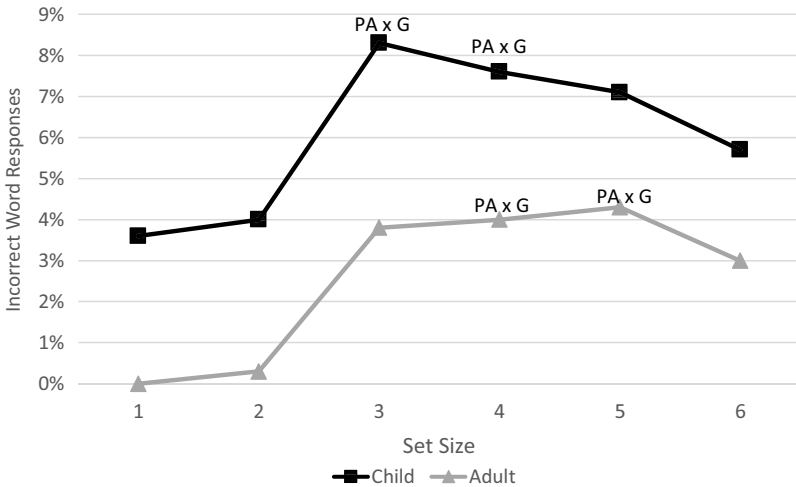
**Table 4.** Models evaluating the effect of phonological activation on response type for adults

Factor	Response type	Posterior mean	HPD interval	<i>pMCMC</i>
<i>Adults, Set size 3</i>				
PA	No response	-1.23	[-1.95, -0.59]	.001
PA	Incorrect word	-1.33	[-2.10, -0.70]	<.001
PA × Group	No response	0.87	[-0.07, 1.85]	.06
PA × Group	Incorrect word	0.35	[-0.59, 1.32]	.46
<i>Adults, Set size 4</i>				
PA	No response	-0.44	[-0.63, -0.25]	<.001
PA	Incorrect word	-0.58	[-0.78, -0.37]	<.001
PA × Group	No response	0.14	[-0.15, 0.43]	.33
PA × Group	Incorrect word	0.38	[0.05, 0.72]	.03
<i>Adults, Set size 5</i>				
PA	No response	-0.25	[-0.40, -0.11]	.001
PA	Incorrect word	-0.39	[-0.55, -0.22]	<.001
PA × Group	No response	0.20	[-0.01, 0.42]	.05
PA × Group	Incorrect word	0.28	[0.04, 0.53]	.02
<i>Adults, Set size 6</i>				
PA	No response	-0.10	[-0.22, 0.02]	.09
PA	Incorrect word	-0.21	[-0.34, -0.09]	.001
PA × Group	No response	0.03	[-0.15, 0.20]	.72
PA × Group	Incorrect word	0.12	[-0.07, 0.33]	.23

Note: PA, phonological activation. HPD, highest posterior density. *pMCMC*, probability that the parameter estimate includes zero.

as compared to a corpus mean (*SD*) of 1.66 (0.86) (Balota et al., 2007). We modeled the effect of frequency on response type, after including set size as a control variable. We evaluated the interactions of frequency with group and age. To parallel the model for serial position, we used data from set sizes 3 to 6. Because frequency did not systematically change with set size, we analyzed the data for the combined sets.

Higher word frequency generally decreased the likelihood of no response ( $M_p = -0.64$ , HPDI [-1.07, -0.23], *pMCMC* = .004) and incorrect word responses ( $M_p = -1.44$ , HPDI [-1.87, -1.00], *pMCMC* < .001). Frequency did not interact with group for no response ( $M_p = 0.07$ , HPDI [-0.37, 0.52], *pMCMC* = .78), but did for incorrect word responses ( $M_p = 0.95$ , HPDI [0.39, 1.49], *pMCMC* < .001). The positive coefficient indicates that the facilitating effect of higher word frequencies was attenuated for the group with DLD. Frequency did not interact with age for no response ( $M_p = 0.51$ , HPDI [-0.04, 1.05], *pMCMC* < .07) nor for incorrect word responses ( $M_p = 0.87$ , HPDI [-0.04, 1.73], *pMCMC* = .053). There were



**Figure 2.** Incorrect word responses as a percentage of total responses by set size and age. Significant phonological activation by language ability group (PA  $\times$  G) interactions indicated for set sizes 3 and 4 for children, set sizes 4 and 5 for adults.

no significant three-way interactions of frequency, group, and age for no response ( $M_p = -0.66$ , HPDI [-1.40, 0.07],  $pMCMC = .08$ ) and incorrect word responses ( $M_p = -0.98$ , HPDI [-2.14, -0.20],  $pMCMC = .10$ ).

Higher frequency target words were less likely to elicit incorrect words and no responses than lower frequency targets. The benefit of frequency was reduced for participants with DLD, but did not interact with age.

## Discussion

Individuals with DLD produced fewer correct recall responses in the listening span task than their peers with TL, consistent with prior reports of more limited VWM capacity in children with DLD (Ellis Weismer et al., 1999; Leonard et al., 2007). Adults with DLD also demonstrated more limited VWM capacity than their peers. Both groups with DLD had lower PIQ scores than their TL peers, consistent with the DLD phenotype (Bishop et al., 2017). The finding for adults with DLD is consistent with evidence for the persistence of the disorder into adulthood (Clegg et al., 2005; Johnson et al., 1999). What has been less clear is why VWM capacity is limited, leading us to analyze the pattern of response types for a VWM task.

No response was the most common error in the VWM task and increased consistently with set size. In contrast, the proportion of incorrect word responses for participants with DLD increased to set size 3 for children and to set size 5 for adults. After these peaks, the proportion of incorrect words declined as set size increased, consistent with findings of Marton et al. (2007) who also reported that the incorrect word responses of children with DLD reached a maximum and then did not increase further with set size. Those authors noted that no response was the most

common error in their data. We sought to understand factors contributing to both incorrect word and omitted responses.

Omitted responses were reduced for all participants by the factors of extreme serial position, higher PA, and higher lexical frequency. Finding fewer omissions for higher frequency words is consistent with prior findings of frequency effects in verbal short-term memory in typical young adults (Allen & Hulme, 2006). The effects of all three factors are consistent with prior work showing that speech production mechanisms and long-term language representations influence VWM performance (Hulme et al., 1997; Roodenrys et al., 2002, Martin & Saffan, 1997), but our findings do not favor a particular VWM model or mechanisms by which language influences VWM. Instead, our findings provide a perspective on how DLD affects the production of correct, omitted, and incorrect responses during the course of a VWM task.

For omitted responses there was an interaction of recency with language group. Interactions with language ability were also found for incorrect word responses for serial position, PA, and frequency. The interactions indicated that the influence of language production factors that facilitated correct responses were attenuated for individuals with DLD, who presumably have less than optimal language production systems. We next explore these differences in our data in light of research on DLD resulting from limited processing capacity and as an emergent phenomenon.

### ***The role of dynamic task demands***

One possible explanation for the response patterns we observed centers on processing limitations and poor inhibitory control (Marton et al., 2007). By this view, children with DLD have more difficulty inhibiting competing stimuli in VWM tasks. The percentage of incorrect word responses increased with set size for participants with DLD to set size 3 for children and 5 for adults. As predicted by the inhibitory control account, as demands increased on the limited VWM capacity of the participants with DLD, they had more difficulty with inhibiting competing stimuli, resulting in more incorrect word responses. The percentage of incorrect word responses beyond these peaks, however, declined whereas no response errors increased, consistent with prior findings (Marton et al., 2007). It is not clear how the inhibitory control account on its own explains the absence of an increase in incorrect words to the largest set sizes. It is also unclear how poor inhibitory control, as a domain general ability, accounts for the pattern of interactions with recency, word frequency, and PA.

The peak of incorrect word responses took place at set sizes likely to be the limits of most participants' VWM span limit. Without a direct measure of VWM span, we must be cautious in interpreting the data as suggesting the size of capacity limits. Prior research, however, suggests a span limit of three to five items for adults with TL (Belleville, Rouleau, & Caza, 1998; Cowan, 2010; Komori, 2016), consistent with the pattern in our data. The capacity for children is more limited (Nicolaou et al., 2018). In any case, increasing set size may have resulted in an initial increase followed by a decline in uncertainty for those with DLD. At small set sizes, they were sure of knowing the target word. At larger sets, they were uncertain, not sure of knowing or not knowing the target, hurting their ability to self-monitor and inhibit

novel responses (Coutinho et al., 2015). At the largest set sizes, they were again certain, but now of not knowing many target words.

The interaction between recency and group for incorrect word responses is consistent with this account of changing levels of uncertainty affecting inhibitory control as the task evolves. In Figure 1, positions coded 2 and 4 elicited more incorrect word responses from children with DLD, consistent with the attenuation of the effect of extreme serial position indicated by the interaction. The penultimate position of a target word may have shifted children with DLD into a more uncertain state resulting in more incorrect word responses. The uncertainty shift related to recency is supported by the information distinctiveness account of serial position effects (Burgess & Hitch, 2006; Oberauer et al., 2016). By this account, early and late items in lists stand out from more central items. The more distinctive targets are more activated than competing alternatives, resulting in better recall. In the case of participants with DLD, increased distinctiveness may affect levels of certainty.

In contrast to the attenuation of recency effects for incorrect responses, we found a heightened recency effect for no response for the group with DLD. This finding is consistent with Mainela-Arnold and Evans's (2005) finding of a heightened recency effect relative to primacy for children with DLD, but differs from Gillam et al. (1998), who found attenuated recency effects for children with DLD. Mainela-Arnold and Evans indicated that the demands of their complex span task were greater than those of the simple span task used in the Gillam et al.'s study, which encouraged their participants to adopt a strategy of focusing on recall of the set-final words. Our study used the same complex span task, but we included adult participants. For both children and adults with DLD, the effect of recency for omission errors varies with task conditions. Our findings add evidence that recency affected no response and incorrect word responses differently. The reduction in no response was accompanied by an increase in incorrect word responses in otherwise facilitating conditions.

The interaction of target word frequency with language ability group for incorrect words followed the same pattern as serial position. Higher frequency generally reduced the likelihood of no response and incorrect word responses, in line with prior findings for typical adults (Allen & Hulme, 2006; Hulme et al., 1997; Luce & Pisoni, 1998). For the group with DLD, however, the effects of frequency was attenuated for incorrect word responses. Compared to lower frequency words at the same level of memory demand, participants may have been shifted toward a threshold state of uncertainty for higher frequency words. In conditions where they otherwise would not have produced an incorrect word, less inhibited participants with DLD now did, resulting in a higher rate of incorrect word responses. A limitation of the study is that we did not attempt to estimate frequencies specific to singular versus plural forms, or to syntactic categories (e.g., *fly* as a verb vs. *fly* as a noun). Frequency effects may vary by word class and form (Rice, Oetting, Marquis, Bode, & Pae, 1994), and these variables should be considered in future research.

The interactions of PA with language ability group add another dimension: variation by set size. PA generally reduced the likelihood of error responses, but this effect was attenuated for the groups with DLD at set sizes near peaks in the proportion of their incorrect word responses (Figure 2). These interactions may have

occurred at these set sizes simply because variability in response type was not attenuated by floor or ceiling effects. However, this was also true at set size 5 for children, yet there was no interaction. Furthermore, there was little evidence of interactions of PA and group for no response. The differential effect of PA for participants with DLD is detectable only for incorrect word responses when the task is neither too easy nor too difficult. The interaction effects appear at points of greater uncertainty associated with the limits of VWM, and the incorrect word responses may be a result of poorer control of inhibition under these conditions for participants with DLD (Coutinho et al., 2015; Marton et al., 2007).

Greater PA may also have supported the compressibility of target words, accounting for the reduction in both error types. Greater PA may have enhanced the strength of representations encoded in short-term memory, facilitating the recognition of phonological patterns in the memoranda that aided recall. The role of compression is dynamic (Chekaf et al., 2018), playing a greater role supporting recall as task demands increase. For typical adults, compression is not demanding of cognitive processing resources (Mathy et al., 2018), but may be for individuals with DLD (Montgomery et al., 2018). At peak task demands, participants with DLD may have been more constrained by efforts to divide their attention between storage and compression on the one hand, and processing on the other hand, resulting in the attenuated benefit of PA on compression. As a result, they produced more overt errors and fewer accurate responses at the limits of their capacities.

Changing demands during the course of the VWM task resulted in more rising then falling proportions of incorrect word responses whereas no response errors increased as set size increased. Incorrect word responses, but not no response errors, were more likely for participants with DLD in the presence of generally facilitating factors of frequency and PA. This pattern of errors may be explained by changing levels of uncertainty coupled with more limited abilities of participants with DLD to inhibit competing stimuli or divide their attention between storage and compression and processing the sentences. This account is consistent with a view of the language performance of individuals with DLD as emergent, resulting from interactions of the individual's capacities and the changing demands on the language production system (Evans, 2001). This emergent account of DLD provides an alternate explanation for the interactions of facilitating speech production factors with language ability group, the effect of language experience.

### ***A role for effective language experience***

Interactions of speech production factors and language ability group consistently indicated that generally facilitating factors had a reduced effect for the group with DLD. These interactions may reflect the fact that individuals with DLD have had less effective language experience. Compared to peers with TL, children with DLD require more exposure to incorporate statistical information on language into their long-term memories (Evans et al., 2009). The implication is that individuals with DLD are less sensitive to the regularities of language that support language learning. As a result, higher frequency for target words, for example, is less effective in suppressing incorrect word responses for the group with DLD.

Phonological activation effects reflected both short- and long-term experience with phonemes. The experience of phonemes within the task builds as the task progresses. Following speech production models (Foygel & Dell, 2000), hearing onsets and vowel nuclei in the task activates those sounds, and by linkage activates candidate words containing those sounds in the task. However, this within-task activation takes place on a backdrop of the participant's long-term experience of the phonemes (Hsiao & Nation, 2018). Individuals come to the task with a base level of connection strength between phonemes and words that are target or alternative responses in the task. Individuals with DLD may come to the task with lower base levels of connection strength, so that when the within-task additional activation is added, there is a smaller effect compared to peers with TL.

Within the study, the mean levels of the model coefficient estimates are consistent with an effect of long-term language experience. In separate models for children and adults estimating the effects of PA on response type, coefficients for the effect of PA were consistently further from zero for the adults. This suggests that the same within-task experience of hearing target word phonemes resulted in larger effects for adults. These are descriptive differences, not statistically tested differences, but the child-to-adult differences are consistent with a role for long-term language experience in how individuals respond to the within-task activation of target-word phonemes. The consistently attenuated effects of lexical and memory factors on the group with DLD, together with a differing magnitude of effects by age group imply that different degrees of effective language experience affect response patterns in VWM tasks.

### ***Observations on incorrect responses***

The characteristics of the words produced incorrectly are of considerable interest. Because these responses were so sparse relative to the overall body of data, statistical analysis is not feasible. However, descriptive observations can be made regarding the errors produced by children; adults did not produce enough errors to consider systematically. For all children, by far the most common error was to repeat a word encountered earlier in the task, including targets and nontargets. Children with DLD appeared to differ from peers in producing relatively more errors related to words not in the sentence set they were attempting to recall. Furthermore, if we identify which sentence set contains the word produced in error, children with TL usually produced a word that appeared in the previous set. In contrast, children with DLD often produced words that could be traced back, two, three, four, or more sets. We speculate that the children with DLD had more difficulty inhibiting words that had been activated earlier in the task. This is an intriguing entry point for further research into recall errors.

### ***Conclusions and implications***

In order to understand how the development of VWM in individuals with DLD is related to language deficits, it is crucial to consider properties of the linguistic system as well as mechanisms of WM. Performance on a complex listening span task was influenced for adults and children with and without DLD by the

memory mechanism represented by serial order, the lexical-level variable of word frequency, and by the phonological properties of the words encountered in the task. The influence of these variables was attenuated for adults and children with DLD, with the exception of a heightened effect of recent serial position on no responses in individuals with DLD. Our findings add to existing research by providing evidence for the complex and dynamic effects of PA, as well as serial position and frequency, in states of uncertainty or instability for individuals with DLD. We also show that these effects are found in adults as well as in children.

We suggest that future research on VWM in individuals with DLD should take the word production system into account. For example, complex span tasks (listening span and others; Jarrold, 2017), could be manipulated to systematically vary serial position, lexical frequency, and PA, as well as the position of phones within words (onset, vowel, and coda), phonological and semantic similarity among items, and lexical variables such as phonological neighborhood density (Acheson & McDonald, 2009). Our study did not consider syllable codas for PA, as in the Foygel and Dell (2000) model. The VWM task in the current study included multisyllabic target words, which raised the question of whether all syllable codas, or only codas of stressed syllables, have an impact on PA and ultimately VWM. These variables are not new to WM research, but the challenge is to manipulate and/or control multiple variables within a single experiment. Mixed-effects models offer tools to help analyze such complex designs, although more complex models become challenging to interpret and may not converge. A systematic set of experimental investigations, each with a manageable number of variables, would enhance our understanding of VWM in DLD.

From a clinical perspective, valid assessment and effective intervention for individuals with DLD depend on understanding the locus of VWM deficits. Our findings suggest that VWM assessment consider different levels of task demand, as factors affecting VWM performance change as the task transitions from easy to challenging to impossible. Intervention to remediate VWM limitations is an active area of research. The evidence supporting intervention in VWM is controversial as researchers debate whether immediate gains in VWM result in downstream gains in functional language ability (Gillam, Holbrook, Mecham, & Weller, 2018). An alternative is to intervene on the language system with downstream gains in VWM. Interventions focused on improving phonological abilities have been shown to benefit VWM capacity (Gillam et al., 2018), a finding consistent with the significant role of PA in our study. Research is needed to determine which intervention approaches result in well-maintained, functional language gains.

We set out to explore why children and adults with DLD have poorer performance on VWM tasks. We found that traditional influences on list recall as well as lexical characteristics of target memoranda affected how participants responded. Factors in common with models of speech production affected VWM response. The conditions under which these factors affected response varied with the level of task challenge. For both children and adults with DLD, these external factors differentially affected their performance when they were in an uncertain state.

**Acknowledgements.** This research was supported by the National Institute on Deafness and Other Communication Disorders Grant 5 R03 DC007312, funding from the Pennsylvania State University Social Science Research Institute (second author), and Ruth L. Kirschstein National Research Service Award 1F31DC010960 (first author). The views expressed are those of the authors and do not necessarily reflect any official position of the National Institutes of Health. Preliminary versions of these analyses were presented at the Symposium for Research in Child Language Disorders, Madison, Wisconsin, in June 2019. Thanks to Michael Dickey for an inspiring conversation about the project, and to Maura Jaeger and Patrick Schoeppner for coding assistance.

## References

- Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*, 50–68.
- Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*, 64–88.
- Archibald, L. M. D. (2017). Working memory and language learning: A review. *Child Language Teaching and Therapy*, *33*, 5–17.
- Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189–208.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.
- Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource sharing model account. *Developmental Psychology*, *45*, 477–490.
- Belleville, S., Rouleau, N., & Caza, N. (1998). Effect of normal aging on the manipulation of information in working memory. *Memory & Cognition*, *26*, 572–583.
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*, 1068–1080. doi: [10.1111/jcpp.12721](https://doi.org/10.1111/jcpp.12721)
- Bracken, B., & McCallum, S. (1998). *Universal nonverbal intelligence test*. Itasca, IL: Riverside.
- Brooks, P. J., Seiger-Gardner, L., Obeid, R., & MacWhinney, B. (2015). Phonological priming with non-words in children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, *58*, 1210–1223.
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.
- Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*, 627–652. doi: [10.1016/j.jml.2006.08.005](https://doi.org/10.1016/j.jml.2006.08.005)
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.
- Chekaf, M., Gauvrit, N., Guida, A., & Mathy, F. (2018). Compression in working memory and its relationship with fluid intelligence. *Cognitive Science*, *42*, 904–922.
- Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders—A follow-up in later adult life. Cognitive, language, and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, *46*, 128–149.
- Coutinho, M., Redford, J., Church, B., Zakrzewski, A., Couchman, J., & Smith, J. D. (2015). The interplay between uncertainty monitoring and working memory: Can metacognition become automatic? *Memory & Cognition*, *43*, 990–1006.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, *104*, 163–191.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford Psychological Series, No. 26. New York: Oxford University Press.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*, 51–57.



- Cowan, N., Li, Y., Glass, B. A., & Saults, J. S. (2018). Development of the ability to combine visual and acoustic information in working memory. *Developmental Science*, *21*, e12635.
- Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, *15*, 331–343.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1136–1146.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.
- Ellis Weismer, S., Evans, J., & Hesketh, L. J. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*, 1249–1260.
- Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *43*, 865–878.
- Evans, J. L. (2001). An emergent account of language impairments in children with SLI: Implications for assessment and intervention. *Journal of Communication Disorders*, *34*, 39–54.
- 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*, 321–335.
- Fidler, L. J., Plante, E., & Vance, R. (2011). Identification of adults with developmental language impairments. *American Journal of Speech-Language Pathology*, *20*, 2–13.
- Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*, 182–216.
- Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (1997). The origins of formal paraphasias in aphasic's picture naming. *Brain and Language*, *59*, 450–472.
- Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *57*, 1363–1382.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Erlbaum.
- Gaulin, C. A., & Campbell, T. F. (1994). Procedure for assessing verbal working memory in normal school-age children: Some preliminary data. *Perceptual and Motor Skills*, *79*, 55–64.
- Gillam, R. B., Cowan, N., & Marler, J. A. (1998). Information processing by school-age children with specific language impairment: Evidence from a modality effect paradigm. *Journal of Speech, Language & Hearing Research*, *41*, 913–926.
- Gillam, S., Holbrook, S., Mecham, J., & Weller, D. (2018). Pull the Andon rope on working memory capacity interventions until we know more. *Language, Speech and Hearing Services in Schools*, *49*, 434–448.
- Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., . . . Turpin, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 355–369.
- Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin*, *99*, 221–228.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*, 1–22.
- Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126.
- Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal of Psychology*, *86*, 527–536.
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology*, *23*, 1217–1232.
- Jarrold, C. (2017). The mid-career award. *Quarterly Journal of Experimental Psychology*, *70*, 1747–1767.
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.

- Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., . . . Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research*, *42*, 744–760.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.
- Komori, M. (2016). Effects of working memory capacity on metacognitive monitoring: A study of group differences using a listening span task. *Frontiers in Psychology*, *7*, 285. doi: [10.3389/fpsyg.2016.00285](https://doi.org/10.3389/fpsyg.2016.00285)
- Kowialiewski, B., & Majerus, S. (2018). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language*, *101*, 64–83.
- Lee, J. C., & Tomblin, J. B. (2015). Procedural learning and individual differences in language. *Language Learning and Development*, *11*, 215–236.
- Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.
- Leonard, L. B., Ellis Weismer, S., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*, *50*, 408–428.
- Leonard, L. B., Ellis Weismer, S., Weber-Fox, C., & Miller, C. A. (2014). The role of processing in children and adolescents with language impairment. In J. B. Tomblin & M. Nippold (Eds.), *Understanding individual differences in language development across the school years* (pp. 117–143). Hove, UK: Psychology Press.
- Leonard, L. B., Nippold, M. A., Kail, R., & Hale, C. A. (1983). Picture naming in language-impaired children. *Journal of Speech and Hearing Research*, *26*, 609–615.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- Mainela-Arnold, E., & Evans, J. (2005). Beyond capacity limitations: Determinants of word recall performance on verbal working memory span tasks in children with SLI. *Journal of Speech, Language, and Hearing Research*, *48*, 897–909.
- Mainela-Arnold, E., Evans, J., & Coady, J. A. (2008). Lexical representations in children with SLI: Evidence from a frequency-manipulated gating task. *Journal of Speech, Language, and Hearing Research*, *51*, 381–393.
- Mainela-Arnold, E., Evans, J. L., & Coady, J. A. (2010). Explaining lexical-semantic deficits in specific language impairment: The role of phonological similarity, working memory, and lexical competition. *Journal of Speech, Language, and Hearing Research*, *53*, 1742–1756.
- Mainela-Arnold, E., Misra, M., Miller, C. A., Poll, G. H., & Park, J. (2012). Investigating sentence processing and language segmentation in explaining children's performance on a sentence-span task. *International Journal of Language and Communication Disorders*, *47*, 166–175.
- Majerus, S., Leclercq, A.-L., Grossmann, A., Billard, C., Touzin, M., van der Linden, M., & Poncelet, M. (2009). Serial order short-term memory capacities and specific language impairment: No evidence for a causal association. *Cortex*, *45*, 708–720.
- Martin, N., & Saffran, E. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*, 641–682.
- Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, *41*, 3–29.
- Marton, K., & Eichorn, N. (2014). Interaction between working memory and long-term memory: A study in children with and without language impairment. *Zeitschrift für Psychologie*, *222*, 90–99.
- Marton, K., Eichorn, N., Campanelli, L., & Zakarias, L. (2016). Working memory and interference control in children with specific language impairment. *Language and Linguistics Compass*, *10*, 211–224.
- Marton, K., Kelmenson, L., & Pinkhasov, M. (2007). Inhibition control and working memory capacity in children with SLI. *Psychologia (Ramat-Gan)*, *50*, 110–121.
- Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and complex working memory tasks allow similar benefits of information compression. *Journal of Cognition*, *1*, 31. doi: [10.5334/joc.31](https://doi.org/10.5334/joc.31)

- Miller, C. A., & Wagstaff, D. A. (2011). Behavioral profiles associated with auditory processing disorder and specific language impairment. *Journal of Communication Disorders*, *44*, 745–763.
- Montgomery, J. W. (2003). Working memory and comprehension in children with specific language impairment: What we know so far. *Journal of Communication Disorders*, *36*, 221–231.
- Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*, 269–288.
- Montgomery, J. W., Evans, J., Fargo, J., Schwartz, S., & Gillam, R. (2018). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *61*, 2950–2976.
- Morice, R., & McNicol, D. (1985). The comprehension and production of complex syntax in schizophrenia. *Cortex*, *21*, 567–580.
- Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Burkner, P. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*, 1225–1242.
- Nicolaou, E., Quach, J., Lum, J., Roberts, G., Spencer-Smith, M., Gathercole, S., . . . Wake, M. (2018). Changes in verbal and visuospatial working memory from Grade 1 to Grade 3 of primary school: Population longitudinal study. *Child: Care, Health and Development*, *44*, 392–400.
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., . . . Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*, 1247–1257.
- Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*, 758–799. doi: [10.1037/bul0000046](https://doi.org/10.1037/bul0000046)
- Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.
- Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language and Communication Disorders*, *48*, 329–342.
- Poll, G. H., Miller, C. A., & van Hell, J. G. (2015). Evidence of compensatory processing in adults with specific language impairment: Testing the predictions of the procedural deficit hypothesis. *Journal of Communication Disorders*, *53*, 84–102.
- Poll, G. H., Miller, C. A., & van Hell, J. G. (2016). Sentence repetition accuracy in adults with developmental language impairment: Interactions of participant capacities and sentence structures. *Journal of Speech, Language, and Hearing Research*, *59*, 302–316.
- Poll, G. H., Watkins, H. S., & Miller, C. A. (2014). Lexical decay during online sentence processing in adults with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *57*, 2253–2260.
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research*, *37*, 106–122.
- 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*, 1239–1257.
- Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1019–1034.
- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168–175.
- Schwering, S. C., & MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience*, *14*, 68. doi: [10.3389/fnhum.2020.00068](https://doi.org/10.3389/fnhum.2020.00068)
- Seiger-Gardner, L., & Brooks, P. J. (2008). Effects of onset- and rhyme-related distractors on phonological processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *51*, 1263–1281.

- Seiger-Gardner, L., & Schwartz, R. G.** (2008). Lexical access in children with and without specific language impairment: A cross-modal picture–word interference study. *International Journal of Language and Communication Disorders*, **43**, 528–551.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2003). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: PsychCorp.
- 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**, 336–344.
- Tan, L., & Ward, G.** (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**, 1589–1625.
- 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**, 1245–1260.
- van der Lely, H., Rosen, S., & Adlard, A.** (2004). Grammatical language impairment and the specificity of cognitive domains: Relations between auditory and language abilities. *Cognition*, **94**, 167–183.
- von der Malsburg, T.** (2016). *Using MCMCglmm to implement lme4-like Bayesian mixed-effects models (DRAFT)*. Retrieved from <https://github.com/tmalsburg/MCMCglmm-intro#using-mcmcglmm-to-implement-lme4-like-bayesian-mixed-effects-models-draft>
- Wechsler, D.** (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wechsler, D.** (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Harcourt Assessment.
- Williams, K. T.** (1997). *Expressive Vocabulary Test*. Circle Pines, MN: American Guidance Service.
- Woodcock, R. W., McGrew, K. S., & Mather, N.** (2001). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.
- Zhao, Y., Staudenmayer, J., Coull, B. A., & Wand, M. P.** (2006). General design Bayesian generalized linear mixed models. *Statistical Science*, **21**, 35–51.

## Appendix A. Response Distribution

**Table A.1.** Response type distribution by age, set size, and typical language (TL) or developmental language disorder (DLD)

Set size (sentences)	Percentage of response type by set size					
	Incorrect word		No response		Correct word	
	TL	DLD	TL	DLD	TL	DLD
<i>Children</i>						
1	3.8	3.3	0.0	1.7	96.2	95.0
2	2.4	5.4	2.9	11.7	94.7	82.9
3	4.5	11.7	12.2	25.0	83.3	63.3
4	2.9	11.7	37.5	41.3	59.6	47.1
5	3.1	10.7	42.7	46.7	54.2	42.7
6	4.8	6.4	47.1	52.2	48.1	41.4
All sets total	3.6	8.6	30.2	36.3	66.2	55.1
<i>Adults</i>						
1	0.0	0.0	0.0	0.0	100.0	100.0
2	0.0	0.6	0.0	3.0	100.0	96.4
3	1.4	6.3	2.2	7.1	96.4	86.5
4	2.2	6.0	9.2	26.2	88.6	67.9
5	2.6	6.2	13.9	36.2	83.5	57.6
6	3.3	2.8	21.4	38.9	75.4	58.3
All sets total	2.0	4.0	10.5	24.0	87.5	71.9

## Appendix B. Multifactor Models

To evaluate whether any effect of PA was an artifact of its relationship to serial position, we compared mean (*SD*) target word PA by serial position. PA did not vary significantly by serial position for set sizes 3–6,  $F(2) = 2.48$ ,  $p = .10$ .

There were an equal number of items at each serial position for set sizes 4 and 6, so we modeled PA with extreme serial position at those set sizes. Models including extreme serial position with PA had lower DICs than those excluding it, so we continued with both predictors. A model based on child data from set size 4 found no significant effect for extreme serial position, but did find an effect of PA on incorrect word responses ( $M_p = -0.36$ , HPDI  $[-0.59, -0.14]$ ,  $pMCMC = .001$ ). There was also a significant PA  $\times$  Group interaction for incorrect words ( $M_p = 0.29$ , HPDI  $[0.04, 0.53]$ ,  $pMCMC = .02$ ). The model for child data at set size 6 found no significant effects of PA.

For adults at set size 4, a model including extreme serial position found that higher PA reduced the likelihood of both no response ( $M_p = -0.30$ , HPDI  $[-0.53, -0.08]$ ,  $pMCMC = .005$ ) and incorrect word responses ( $M_p = -0.39$ , HPDI  $[-0.66, -0.13]$ ,  $pMCMC = .003$ ). Interactions of group and PA were not significant. At set size 6 for adults, there were effects of extreme serial position on no response ( $M_p = -1.20$ , HPDI  $[-1.76, -0.62]$ ,  $pMCMC < .001$ ) and incorrect words ( $M_p = -0.96$ , HPDI  $[-1.78, -0.12]$ ).

$pMCMC = .02$ ). Higher PA reduced incorrect word responses for adults ( $M_p = -0.18$ , HPDI  $[-0.30, -0.06]$ ,  $pMCMC = .004$ ).

To evaluate whether any effect of target word frequency was an artifact of its relationship with extreme serial position, we assessed whether frequency systematically varied with serial position. Mean (*SD*) target word frequency for set sizes 3–6 did not differ significantly by serial position,  $F(2) = 0.174$ ,  $p = .841$ . A regression model with extreme serial position, set size, and target word frequency had a lower DIC (4850.8) than a model without extreme serial position (DIC = 4863.1), so we retained serial position in the model. This model found significant effects for extreme serial position ( $M_p = -0.43$ , HPDI  $[-0.82, -0.02]$ ,  $pMCMC = .04$ ) and word frequency for no response ( $M_p = -0.50$ , HPDI  $[-0.90, -0.08]$ ,  $pMCMC = .02$ ) and frequency on incorrect word responses ( $M_p = -1.42$ , HPDI  $[-1.82, -0.99]$ ,  $pMCMC < .001$ ). There was no significant frequency by group interaction for no response but there was for incorrect word responses ( $M_p = 0.82$ , HPDI  $[0.38, 1.26]$ ,  $pMCMC < .001$ ).

We evaluated models including both word frequency and PA and found that models including both variables had higher DICs than those with PA alone, so we did not pursue models including both predictors.

These multifactor models indicate that extreme serial position, PA, and target word frequency have effects on response type that are not simply artifacts of the relation of one factor with another.

---

**Cite this article:** Poll, GH. and Miller, CA. (2021). Speech production factors and verbal working memory in children and adults with developmental language disorder. *Applied Psycholinguistics* 42, 673–702. <https://doi.org/10.1017/S0142716421000011>