Phonological and Semantic Contributions to Verbal Short-Term Memory in Young Children with Developmental Stuttering

Julie D. Anderson

Indiana University

Stacy A. Wagovich and Bryan T. Brown

University of Missouri

Author Note

Julie D. Anderson, Department of Speech and Hearing Sciences, Indiana University; Stacy A. Wagovich and Bryan T. Brown, Department of Communication Science and Disorders, University of Missouri.

Bryan T. Brown is now at Department of Communication Sciences and Disorders, University of Wisconsin-Eau Claire.

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (R01DC012517).

Correspondence concerning this article should be addressed to Julie D. Anderson, Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405. E-mail: judander@indiana.edu

Abstract

Purpose: The purpose of this study was to examine the verbal short-term memory skills of children who do (CWS) and do not stutter (CWNS) in two experiments, focusing on the influence of phonological and semantic similarity.

Method: Participants were 42 CWS and 42 CWNS between the ages of 3;0 and 5;11 (years; months). In Experiment 1, children completed the phonological similarity task (PST), in which they listened to lists of phonologically similar and dissimilar words and then repeated them when signaled to do so. In Experiment 2, children completed another forward span task, the semantic category task, which is similar to the PST, except that it consisted of lists of semantically homogeneous and heterogeneous words. Main dependent variables were cumulative memory span, proportion of errors by type, and speech reaction time (SRT) for correct responses.

Results: The CWS exhibited significantly shorter memory spans for phonologically dissimilar words and were less affected by the phonological qualities of the words than the CWNS in Experiment 1, based on the findings of both between- and within-group analyses. In Experiment 2, while the groups did not differ in their performance in either condition, within-group analyses revealed that the CWNS benefitted from semantic similarity, whereas the CWS did not. The between-group difference in absolute different scores, however, failed to reach significance. The CWS produced more omissions and false alarms than the CWNS in both experiments, but the two groups of children were otherwise comparable in SRT, although the CWS exhibited overall faster SRT than the CWNS in Experiment 2.

Conclusions: Verbal short-term memory is one domain-general cognitive process in which CWS display weakness relative to typically fluent peers. These weaknesses are likely due, in part, to differences in phonological and, perhaps, semantic processing of words to aid memory.

Keywords: stuttering, fluency disorders, memory, executive functions, phonology, language, children

Phonological and Semantic Contributions to Verbal Short-Term Memory in Young Children with Developmental Stuttering

To date, there is a robust literature focusing on differences between children who stutter (CWS) and children who do not stutter (CWNS) in their language skills, particularly during the preschool years (e.g., Anderson, Pellowski, & Conture, 2005; Silverman & Ratner, 2002; see Ntourou, Conture & Lipsey, 2011 for a meta-analysis). Findings suggest that, early on, CWS may have subtly depressed language skills compared to CWNS (but cf. Watts, Eadie, Block, Mensah, & Reilly, 2017). During the preschool years, children experience dramatic growth, not only in language, but also in motor and cognitive skills. Thus, as many have suggested (e.g., Bloodstein & Ratner, 2008; Guitar, 2013), it is unsurprising that some children develop stuttering during this period of tremendous growth. Within the cognitive realm, the study of executive function skills in young CWS is just beginning. Even so, understanding of this area is critical, in that the development of executive function skills impacts and is impacted by other domains, such as speech and language. Fuller understanding of the onset and development of stuttering requires knowledge of the development of these skills in young, preschool-age children.

Broadly defined, executive function is a collection of individual top-down thought processes that are foundational for problem-solving and self-regulation (Diamond, 2013). Much of the literature has focused on the executive function capabilities of adults, whereas the developmental trajectory of this set of skills in children is not as well understood. Children's executive function skills consist of shifting, inhibition, and updating (Lehto, Juujarvi, Kooistra & Pulkkinen, 2003). According to Lehto et al., shifting refers to cognitive flexibility, and more specifically, the ability to shift attention between mental sets or tasks. Inhibition refers to the ability to suppress (a) distractors that might interfere with the completion of a dominant response, or (b) a dominant response in favor of a non-dominant response. Finally, updating refers to memory skills, specifically the temporary storing and manipulation of information. It is this latter component that is the focus of the current study; in particular, we examined verbal short-term memory (vSTM) in young CWS and their peers with typical fluency. What follows is an overview of the concepts of short-term and working memory, followed by a discussion of the memory skills of CWNS and CWS based on measures of nonword repetition and span. We then consider two phenomena known to influence accuracy on span tasks—the phonological and semantic similarity effects—before addressing the purpose of the study.

Short-Term and Working Memory: Definitions and Theoretical Underpinnings

The process of memory formulation is comprised of three basic steps: acquisition, consolidation, and storage (Baddeley, 2002). Acquisition is the process of acquiring information via sensory organs. Consolidation is the rehearsal and subsequent development of a robust mental representation of that information. Storage is the process of creating a stable, long-term record of that information that is available for future recall. While storage is a component related to long-term memory, acquisition and consolidation are components of both short-term and long-term memory.

Working memory has been widely studied and is commonly defined as a neurocognitive system that facilitates the temporary storage (i.e., short-term memory) and manipulation of incoming visual or acoustic information (Baddeley, 1986, 2002, 2003). Baddeley's (1986, 2003) widely-referenced multi-component model of working memory provides one explanation of the process of working memory. According to Baddeley, there are four components in working memory: two sensory-specific components (*visual-spatial sketchpad* and the *phonological loop*), a secondary temporary store to interface between short-term and long-term memory (*episodic buffer*), and a *central executive* to manage the process of integrating working memory with other cognitive components.

Of particular interest to the present study is the phonological loop component of Baddeley's (1986, 2003) model. The phonological loop is a specialized short-term storage system for verbal information, which consists of two subcomponents: the phonological store and an articulatory rehearsal mechanism (Henry, 2012). The phonological store holds verbal information (referred to as the memory or phonological trace), but only for a short period of time (within two seconds) before it begins to decay. Thus, the second subcomponent of the phonological loop, the articulatory rehearsal mechanism, is needed to prevent this rapid decay. Articulatory rehearsal is used to refresh the information in the phonological store. Fidelity of the rehearsal process, however, is limited by features of the information requiring rehearsal (e.g., phonologically similar vs. distinct; Salamé & Baddeley, 1986), as we discuss in the sections that follow.

Examination of Memory Skills in CWNS and CWS: Nonword Repetition

One popular means of measuring working memory and, in particular, the phonological loop (i.e., vSTM), is through a nonword repetition (NWR) task. In this type of task, a child hears and repeats a series of nonwords (one at a time). The nonwords are consistent with the phonological structure of the native language, but vary in length. The repetition of nonwords has long been viewed as a measure of vSTM (see Gathercole, 2006, for review). In fact, at one time, NWR was considered to be a relatively “pure” measure of vSTM because the production of nonwords, unlike real words, was presumed to be unassisted by long-term lexical knowledge (Baddeley, Gathercole, & Papagno, 1998). However, research has since revealed that long-term lexical and sublexical knowledge does impact NWR performance (Jones & Witherstone, 2011; Leclercq, Maillart, & Majerus, 2013). For example, children tend to repeat nonwords that are higher in wordlikeness more accurately than those lower in wordlikeness (Gathercole, 1995). Likewise, children’s repetition accuracy is improved for nonwords containing high frequency phoneme sequences compared to those containing low frequency phoneme sequences (Munson, Edwards, & Beckman, 2005).

Furthermore, in addition to reliance on long-term lexical and sublexical knowledge, other processes are involved in the ability to successfully repeat a nonword. A child must be able to perceive and encode (i.e., segment) the phonological sequence, hold the sequence in a temporary memory store (i.e., vSTM), assemble the phonemes, and then plan and execute the motor commands necessary for production (Coady & Evans, 2008; Estes, Evans, & Else-Quest, 2007; Gathercole, 2006; Leclercq et al., 2013; Snowling, Chiat, & Hulme, 1991). Thus, reduced NWR ability could be a consequence of difficulty with any of the aforementioned skills involved in the process of perceiving, encoding, storing, and/or producing phonological word forms (Estes et al., 2007).

Research has revealed that children with specific language impairment exhibit marked deficits in their ability to accurately repeat nonwords (see Gathercole, 2006). Likewise, typically-developing children with poor vocabulary skills repeat nonwords significantly less accurately than children with good vocabulary skills (e.g., Gathercole & Baddeley, 1989). This relationship between language learning abilities and nonword repetition has prompted some investigators to examine the NWR performance of CWS, given that CWS tend to perform more poorly on measures of speech and language compared to their typically fluent peers (e.g., Ntourou et al., 2011).

At present, only a small group of studies have documented NWR ability in CWS, yielding conflicting results. Four demonstrate that CWS of various ages produce a higher rate of repetition errors than CWNS (Anderson, Wagovich & Hall, 2006; Hakim & Ratner, 2004, Oyoun, Dessouky, Shohdi & Fawzy, 2010; Pelczarski & Yaruss, 2016), whereas three have failed to identify significant differences between CWS and CWNS in NWR (Bakhtair, Ali, & Sadegh, 2007; Sasisekaran & Byrd, 2013; Smith, Goffman, Sasisekaran, & Weber-Fox, 2012). Some of the variation across these studies is likely due to methodological differences across NWR tasks. In fact, findings of a meta-analysis conducted by Estes et al. (2007), indicated that the magnitude of the effect size difference between children with and without specific language impairment differed significantly depending on the NWR task used; this suggests that different tasks may tax different component skills. Nonetheless, regardless of the reason for the difference in findings, most studies would seem to suggest that CWS of various ages have some degree of difficulty with NWR skills (cf. Ofoe, Anderson &, Ntourou, 2018). Difficulties could be a consequence of poor auditory-perceptual skills or phonetic encoding, reduced vSTM ability, inefficient phoneme selection (i.e., phonological encoding), and/or reduced integrity of motor planning/execution. Nevertheless, vSTM does play a major role in children's performance on NWR tasks (Gathercole, 2006) and, thus, it is a clearly a potential source of difficulty for CWS. Children's performance on NWR tasks tends to be highly correlated with their performance on word span tasks, suggesting that the two measures tap into many of the same processing components (Gathercole, Willis, Baddeley, & Emslie, 1994). Next, we briefly summarize the literature on the use of span tasks with CWS.

Examination of Memory Skills in CWNS and CWS: Span Tasks

vSTM is also routinely assessed through behavioral tasks that require a participant to listen to and then immediately repeat a list of digits, letters, or words presented in a series of increasing lengths. The longest list of digits/letters/words that can be successfully repeated in sequential order is the participant's memory span. Digit, letter, and word span tasks have been used as a means to test hypotheses related to short-term memory (e.g., Baddeley, Thomson, & Buchanan, 1974; Gupta, Lipinski & Aktunc, 2005; Gupta & MacWhinney, 1995), and were instrumental in the development of Baddeley's model of working memory, particularly the phonological loop (Baddeley, 2003).

There are a number of ways in which word span tasks, in particular, differ from NWR tasks. Some of these differences make word span tasks easier (e.g., stimuli consist of single-syllable words, which have fewer motor output demands), and others make NWR tasks easier (e.g., stimuli consist of multisyllabic nonwords, which may provide children with additional coarticulation or prosodic cues across syllables; see Archibald & Gathercole, 2007a, 2007b; Archibald, Gathercole, & Joanisse, 2009; Gathercole, 2006; Henry, 2012). As noted by Archibald and Gathercole (2007a), studies often report that children perform more poorly on NWR tasks than they do span tasks, but direct comparisons are difficult to make because the stimuli in these tasks often differ in length, familiarity, and phonological forms. However, when the stimuli in both tasks are matched by phonological content and, thus, are comparable in vSTM load, children tend to perform more poorly on nonword span tasks than they do

NWR tasks, suggesting that there are other factors outside of vSTM that facilitate performance in NWR tasks (Archibald & Gathercole, 2007a, 2007b).

Like studies of NWR ability, studies that have examined the performance of preschool and/or school-age CWS versus CWNS on simple memory span tasks have yielded somewhat contradictory results, with some studies reporting differences favoring CWNS (e.g., Reilly & Donaher, 2005; Oyoun et al., 2010) and other studies not (e.g., Bakhtiar et al., 2007; Pelczarski & Yaruss, 2016; Sasisekaran & Byrd, 2013). However, findings from a recent meta-analysis, which by combining individual data from many independent studies increases statistical power, revealed that the memory spans of CWS are shorter than those of CWNS (Ofoe, Anderson &, Ntourou, 2018). It should also be noted that most of the studies that have failed to find differences used digit span tasks. Given that memory span tends to be highest for digits and lowest for words, as demonstrated in the classic study by Crannell and Parish (1957), digit span tasks may not be challenging enough to reveal subtle differences between groups, if they exist.

Phonological and Semantic Similarity Effects in Span Tasks

In word span tasks, the selection of words included in lists influences recall accuracy. In particular, the phonological similarity effect (PSE) and the semantic similarity effect (SSE) have been observed to impact list memory. The PSE refers to the finding that phonologically similar words (e.g., mad, map, man) are harder to recall than phonologically different words (e.g., hit, tub, cow; Cowan, Saults, Winterowd, & Sherk, 1991; Gupta et al., 2005; Shulman, 1971). Of interest, the PSE is diminished when participants simultaneously produce irrelevant speech (e.g., counting or repeatedly saying a word; Coltheart & Leahy, 1992). Using Baddeley's (2003) model, this suggests that (a) access to the phonological loop during serial recall of verbal material is facilitative, and that (b) use of the phonological loop when words are phonologically similar is much more difficult than when words are phonologically different.

Only a few investigators have examined the PSE in typically-developing children between the ages of 3 and 5 years, the findings of which have generally revealed a memory advantage for auditorily presented dissimilar words compared to similar words (Brown, 1977; Cowan et al., 1991; Hulme, 1987; Hulme & Tordoff, 1989; Lee, Pennington, & Kennan, 2010; van der Lely & Howard, 1993). However, when items are presented visually (e.g., pictures), preschool children often fail to demonstrate the PSE, as demonstrated by Conrad (1971; cf. Brown, 1977, Henry, Messer, Luger-Klein, & Crane, 2012; Hulme, 1987). As noted by Henry et al. (2012), auditory stimuli have direct access to the phonological store because the phonological forms are already created by the auditory input. In contrast, visual stimuli must first be recoded into a phonological form (via the articulatory rehearsal mechanism) before being accessed by the phonological store (Gathercole, Pickering, Ambridge, & Wearing, 2004; Henry, 2012; Henry et al., 2012). Thus, because of the extra step involved in processing visual stimuli, the PSE is thought to develop later for visually presented items than auditory presented items (Hitch, Halliday, Schaafstal, & Heffernan, 1991). Other methodological factors, such as the type of task used (span vs. fixed length), output modality (spoken vs. probed), and the presentation of conditions (alternating vs. blocked) can also influence children's sensitivity to the PSE (Henry, 1991; Henry et al., 2012).

While the magnitude of the PSE tends to increase with age for auditorily presented stimuli (Hulme & Tordoff, 1989; but cf. Jarrold & Citroën, 2013), its presence occasionally varies across individuals. For example, some studies have reported that the PSE is not always consistent in healthy adults (Beaman, Neath, & Suprenant, 2008; Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996), neuropsychological patients (Collette, van der Linden, Bechet, & Salmon, 1999), and children with Williams syndrome (Majerus, Barisnikov, Vuillemin, Poncelet, & van der Linden, 2003). The degree to which the PSE is present, at least in adults, seems to depend on whether individuals use subvocal rehearsal and/or their overall level of recall (Logie et al., 1996; cf. Beaman et al., 2008). With respect to the latter, Logie et al. found that adults with shorter word spans in the "easier" (i.e., dissimilar) condition had significantly smaller PSEs than those with longer spans. In the case of young children, the use of memory strategies is not likely to affect the PSE given that these children lack sophisticated rehearsal strategies (e.g., Gathercole, 1998; Gathercole, Adams, & Hitch, 1994; Tam, Jarrold, Baddeley, & Sabatos-DeVito, 2010). Thus, it appears more likely that the size of the PSE in young children would be linked to how they performed in the dissimilar condition and, indeed, there is evidence to suggest that this is the case (Jarrold & Citroën).

The SSE has received substantially less attention than the PSE. In contrast to the PSE, the SSE is described as a *facilitation* of memory recall for lists comprised of words that are semantically related (Monnier & Bonthoux, 2011; Poirier & Saint-Aubin, 1995, Shulman, 1971), compared to lists of semantically unrelated words. Of interest, unlike the PSE, simultaneous production of irrelevant speech does not impact the effect (Poirier & Saint-Aubin), indicating that processing of semantically related or unrelated lists may not be impacted in the same way by the functioning of the phonological loop.

Only three studies, to our knowledge, have examined the SSE in typically-developing preschool children. In the first of these studies, van der Lely and Howard (1993) examined the effect of semantic similarity in six children with specific language impairment and 17 language-matched typically developing children between 3 and 7 years of age. Children completed a span task, consisting of a semantically similar and dissimilar condition, with a set of puppets. The puppets were placed in front of the child, the number of which varied depending on the span length being assessed (e.g., if the list contained two words, then two puppets appeared), and each puppet was given a word by the examiner. The examiner said the word while at the same time pointing to the puppet who “said” it. Children were instructed to repeat the words in the order heard while also pointing to the puppet associated with it. The authors reported that even though the typically-developing children produced longer memory spans in the semantically similar condition than the semantically dissimilar condition, this difference failed to reach statistical significance. However, as noted by Gathercole and Baddeley (1995), the failure to elicit the SSE in this study is a likely consequence of the unorthodox procedures the authors used to measure memory span, for children not only had to remember the sequence of words in each list, but also which puppet produced the word. Consequently, the task posed additional memory demands that may have obfuscated the effect of semantic similarity. Findings from two subsequent studies, in which traditional span tasks were used, substantiate this interpretation. Lee et al. (2010) reported that the memory spans of 18 typically-developing children between 3 and 10 years of age (*M* = 5.8 years) were significantly higher for words that were semantically similar compared to words that were semantically dissimilar. Likewise, Monnier and Bonthoux (2011) successfully elicited the SSE in 74 children between the ages of 4;10 to 6;2 (years; months).

Purpose of the Study

Examination of the PSE and SSE may be particularly relevant in providing a clear picture of the vSTM skills of CWS. We might anticipate differences between groups in the extent to which performance is impacted by the PSE and SSE, given that CWS have been shown to exhibit different phonological and semantic/lexical processing patterns compared to CWNS (e.g., their reactions to priming, the correspondence between vocabulary knowledge and semantic processing, etc.; Anderson, 2008; Byrd, Conture, & Ohde, 2007; Hartfield & Conture, 2006; Pellowski & Conture, 2005).

The purpose of this study, therefore, was to evaluate the vSTM of CWS relative to CWNS using two forward word span tasks validated in preschool children. The burgeoning literature exploring memory processes in CWS is important, because these, as well as other domain-general processes, may link to the development and expression of stuttering in young children. Indeed, as reviewed above, some differences in memory processes have been observed, primarily through NWR tasks. Our focus on vSTM through forward word span tasks is motivated by and serves as a complement to prior NWR studies. Understanding vSTM skills, as one component of the working memory system (Baddeley, 2003), can inform our understanding of the larger working memory system, as well as executive function skills, more broadly, in CWS. This study provides a necessary first step in exploring the linkage between stuttering and vSTM.

In addition, we explored the extent to which language processing impacts vSTM performance for CWS, compared to CWNS. In Experiment 1, we examined the memory spans and recall latencies of CWS and CWNS in response to phonologically similar and dissimilar words, as well as the impact of the PSE. In Experiment 2, we compared the memory spans and recall latencies of both groups of children on a span task designed to elicit the SSE. In both experiments, stimuli were presented auditorily and word lists were presented in increasing length and blocked by condition, as these procedures maximize the likelihood of eliciting the PSE in preschool children (Henry et al., 2012) and perhaps the SSE, as well. Furthermore, to make the tasks more suitable for use with preschool children, each condition was limited to six word lists and each word list was capped at a maximum of four words.

In theory, if CWS differ from peers in phonological and semantic processing skills, this should impact the demonstration of phonological and semantic similarity effects, and thus, the overall pattern of vSTM skills observed in CWS, relative to CWNS. Therefore, this study was designed to explore whether phonologically/semantically similar/dissimilar word lists impacted serial recall for each group of children. We hypothesized that CWS would have greater difficulty with vSTM than CWNS, particularly when phonological and semantic demands are greater.

Experiment 1

Method

Participants. Participants were 42 CWS and 42 CWNS between the ages of 3;0 and 5;11 (years; months). All participants met the following criteria based on parent report and/or examiner observation/testing: (a) spoke American English as their primary language, (b) had no history of neurological or intellectual difficulty, (c) had not been diagnosed with any speech-language disorder(s) besides stuttering, and (d) had hearing within normal limits. Children and their parent(s) were recruited for participation via newspaper/magazine advertisements, posted flyers, and referrals.

Group classification. To be classified as a CWS, children were required to have an average of three or more stuttered disfluencies (part-word repetitions, single-syllable word repetitions, sound prolongations, and/or blocks; Yairi & Seery, 2015; cf. Pellowski & Conture, 2002) per 100 words of conversational speech (described below) and receive a stuttering severity score of 12 or higher on the *Stuttering Severity Instrument-4* (SSI-4; Riley, 2009). To be classified as a CWNS, children were required to have fewer than three stuttered disfluencies, on average, per 100 words of conversational speech.

Children in the CWS group produced a median of 5.15 (*M* rank = 63.50) stuttered disfluencies while children in the CWNS group produced 0.33 (*M* rank = 21.50), a significant difference, $z = -7.92$, $p < .001$. The stuttering severity ratings of the children in the CWS group ranged from “mild” ($n = 25$) to “moderate” ($n = 17$). The CWS had been stuttering, on average, for 16.28 months ($SD = 9.96$) per parent report.

Group matching. At the two data collection sites (see below), both groups of children were matched by chronological age (±4 months) and gender (13 girls and 29 boys per group) and equated by socioeconomic status using Hollingshead’s Four-Factor Index of Social Position (Hollingshead, 1975). The median chronological ages in months for the children in the CWS and CWNS groups were 49.00 and 50.00, respectively. A Mann-Whitney U test revealed no significant difference in chronological age between the two groups of children (CWS: *M* rank = 41.87; CWNS: *M* rank = 43.13), $z = .24$, $p = .81$. The median Hollingshead’s family social position score, which is based on both the mother and father’s education level and occupation, was 52.00 (*M* rank = 46.62) for the children in the CWS group and 51.00 (*M* rank = 38.38) for the children in the CWNS group, a non-significant difference, $z = -1.55$, $p = .12$.

Procedures. Testing was conducted over the course of two sessions at Indiana University and the University of Missouri, with each session lasting approximately 1-1½ hours. At each testing site, children participated in the following tasks: (a) parent-child conversational interaction, (b) standardized speech-language testing and hearing screening, and (c) a vSTM task, the phonological similarity task. Children also participated in several other experimental tasks not related to the present study, with all tasks presented in random order across participants. This study protocol was approved by an institutional review board at each institution.

Parent-child conversational interaction. Children participated in a conversational interaction with their parent(s) for group classification purposes. Children and their parent(s) conversed with one another for approximately 20 minutes while seated at a small table with age-appropriate toys. A 300-word speech sample was collected from each child and analyzed for the presence of stuttered disfluencies and, for CWS, stuttering severity.

Speech-language tests and hearing screening. Children completed the following standardized, norm-referenced speech and language tests for subject inclusion purposes: (a) *Peabody Picture Vocabulary Test-4* (PPVT-4; Dunn & Dunn, 2007), a receptive vocabulary measure; (b) *Expressive Vocabulary Test-2* (EVT-2; Williams, 2007), an expressive vocabulary measure; (c) *Test of Early Language Development-3* (TELD-3; Hresko, Reid, & Hammill, 1999), a receptive/expressive language measure; and (d) "Sounds-in-Words" subtest of the *Goldman-Fristoe Test of Articulation-2* (GFTA-2, Goldman & Fristoe, 2000), a speech sound articulation measure. All children received a standard score of 85 or higher on each test, suggesting the absence of any speech-language disorder(s) other than stuttering for children in the CWS group. Independent sample *t*-tests revealed no significant differences between the two groups of children on all four speech and language tests, with *p*-values ranging from .34 to .98.

Children also completed a hearing screening to ensure that their hearing was within normal limits using bilateral pure tone testing at 20 dB HL for 1000, 2000, and 4000 Hz (American Speech-Language-Hearing Association, 1997). All children passed the hearing screening, suggesting hearing within normal limits.

Phonological similarity task (PST). The PST measures phonological contributions to vSTM in children. It is a traditional forward word span task in which children listen to a series of phonologically similar and dissimilar words and then repeat them when signaled to do so on the computer (e.g., Gupta et al., 2005; Lee et al., 2010).

Materials and Procedure. Children were seated in front of a computer with two microphones and given the following instructions: “We are going to play a listening game. The bird is going to say some words and when he’s done talking, you are going to repeat the words in the order you heard them. So, the first word you hear has to be repeated first and the last word you hear has to be repeated last. We are going to practice first so you can see how the game is played.” Following these instructions, children completed two practice phases. In the first phase, children were introduced to a cartoon picture of a parrot on the computer screen. They then completed two practice trials in which the parrot “said” a word and they repeated the word as soon as the parrot was done “talking,” as indicated visually on the computer screen by a question mark (i.e., the question mark signaled that the parrot was done “talking” and that it was now time for them to respond). This phase was repeated until child was able to correctly respond to both single word practice trials. In the second phase, children were presented with two practice trials with sets of two words to be repeated per trial. This phase was again repeated until the child correctly repeated both words in each trial. Children were instructed to withhold their responses until they saw the question mark to safeguard against premature responses [e.g., to prevent children from saying “*hat*” immediately after the parrot said “*hat*” and “*shoes*” immediately after the parrot said “*shoes*” in a two-word list (*hat, shoes*)]. During both practice phases, children were given cues and/or feedback as needed (e.g., “don’t forget to wait until the parrot gets done talking before you repeat the words,” “don’t forget to say the words in the order you heard them,” etc.), but they had to complete both practice trials in each phase independently to advance to the next phase/experiment.

Following practice, children were told that they were now going to "play the game for real" and, if needed, they were again reminded of the instructions. During the experiment, children were presented, in a fixed random order, with six word lists in two conditions: phonologically similar and phonologically dissimilar (see Table A1 of the Appendix for the list of experimental stimuli). The phonologically similar condition contained 18 nonrhyming three-phoneme consonant-vowel-consonant (CVC) words, similar in that each consistently overlapped in the vowel /æ/ and sometimes in the first consonant. The 18 words in the phonologically dissimilar condition also consisted of three-phoneme CVC sequences, but unlike the phonologically similar condition, they were phonologically unrelated (i.e., none of the phonemes in a list consistently overlapped). Words were drawn from an open set (i.e., no words were repeated across lists) in both conditions so as to minimize the potential influence of proactive interference, a phenomenon in which prior experience with the same or similar words interferes with recall accuracy (Goh & Pisoni, 1998; Nairne & Kelley, 1999). The six word lists in each condition were presented in increasing length, beginning with two word lists and ending with four word lists. Two trials were presented at each word list length (i.e., there were 2 two-word lists, 2 three-word lists, and 2 four-word lists). Children completed all six word lists in each condition.

An open set of nonrhyming words was selected (as opposed to rhyming words) for the phonologically similar condition because of the greater challenge they present; accuracy tends to be consistently lower for phonologically similar nonrhyming words than phonologically dissimilar words (Coady, Mainela-Arnold, & Evans, 2013; Gupta et al., 2005). The nonrhyming words were chosen from among the list of canonically similar words used in Gupta et al. and the Child Corpus Calculator (Storkel & Hoover, 2010) based on mean age of acquisition (AoA), word frequency, positional segment frequency, concreteness, and imageability.[1] Words in the phonologically dissimilar condition were also

---

[1] Mean AoA values were obtained from the database of Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012); values for word frequency and positional segment frequency were obtained from the Child Corpus Calculator (Storkel & Hoover, 2010]; and values for concreteness and imageability were obtained from the MCR Psycholinguistics Database (Wilson, 1988).

selected from the Child Corpus Calculator based on the five aforementioned variables (i.e., AoA, word frequency, etc.). Independent samples *t*-tests revealed no significant differences between the phonologically similar and dissimilar conditions in each of the five variables, with *p*-values ranging from .20 to .99. These variables were also comparable across trial list levels (e.g., 2 words vs. 3 words, etc.) in both the phonologically similar (p = .23 to .90) and dissimilar (p = .08 to .96) conditions.

Words were digitally recorded by a male speaker and edited to 32-bit resolution at a sampling rate of 44 kHz. The intensity of the auditory stimuli was also equated for root mean square amplitude using Adobe Audition CS6. Each word was presented for approximately 500 ms, with an inter-stimulus interval (i.e., the time between words in a list) of 1000 ms. The question mark appeared 175 ms after the last word in the series. The inter-trial interval (the time between each word list) was 2500 ms. The presentation of the conditions was randomized across participants, such that half of the children began with the phonologically similar condition, whereas the other half began with the phonologically dissimilar condition.

The PST was developed and presented using E-Prime v. 2.0 software by Psychology Software Tools, Inc. A Psychology Software Tools Serial Response Box, which has a microphone and voice key, was directly connected to the computer via the serial port. When a child produced a verbal response, the microphone triggered the voice key and the resulting SRT data, expressed in milliseconds, were directly recorded onto the computer via the E-Prime software program. Thus, SRT reflects the amount of time that has elapsed between the offset of the stimulus list to the onset of the child's verbal response, referred to as the "preparatory interval" in the memory literature (Cowan, 1992; Cowan et al., 1994). A second microphone was also connected to the computer. This microphone recorded each child's response to the trial lists through a soundcard directly onto the hard drive.

The SRT measurements collected from reaction time programs can sometimes be unreliable (Protopapas, 2007; Rastle & Davis, 2002), especially with young children. The voice key may be triggered prematurely (e.g., the child audibly inhales, clicks their tongue, bumps the table, etc.), not at all (e.g., the child does not speak loud enough or does so too late), or not until the second or third phoneme of a word (the voice key fails to trigger on the initial phoneme). Technical errors such as these can be mitigated by presenting participants with a large number of trials and removing extreme values from the data corpus (e.g., removing SRT values above and below a certain cut-off point). However, increasing the number of trials is not always a viable option, especially in studies of young children with limited attention spans.

In the present study, each condition of the PST was limited to six trials due to the challenging nature of the task. As a result, there were fewer opportunities for obtaining useable SRT data and those that were obtained were more subject to technical errors—for example, some children noticeably spoke more softly when challenged, presumably because they were uncertain or fearful of making a mistake. In fact, approximately 40% of the SRT data for correct PST responses recorded in E-Prime were not useable due to definitive or suspected technical errors. Thus, to maximize the amount of SRT data available for analysis, all of the children's correct responses to the trial lists, which had been recorded as a waveform directly onto the computer's hard drive (see above), were re-calculated semi-automatically using CheckVocal software (Protopapas, 2007). The preparatory intervals (i.e., SRT) in CheckVocal were measured in milliseconds from the acoustic offset of the auditory stimuli to the onset of the child's correct verbal response. Any correct response containing a speech disfluency (e.g., "b-b-boy" or "uh… car") was excluded from the SRT analyses.

Scoring. Children's responses to the word lists were scored as correct (child repeats all words in the correct order) or incorrect. Incorrect responses were further categorized by type of error: omissions (child fails to repeat one or more words correctly), intrusions (child repeats a word that had not been presented), order errors (child does not repeat the words in the correct order), and false alarms (child repeats one or more words as they are being presented or waits until all words have been presented but responds before the question mark appears). Main dependent variables in each condition include: (a) cumulative memory span (Cowan et al., 1994; Hulme, Maughan, & Brown, 1991), which is the longest word length for which all words, across the two trials of that word length, were correctly repeated in the right order plus .25 for every subsequent individual trial recalled correctly (e.g., if a child correctly repeats both 2-word trials correctly and one 3-word trial, he/she would get a score of 2.25); (b) proportion of errors by type (based on the total number of errors across all word lists); and (c) SRT for correct responses. While responses containing speech disfluencies were excluded from the SRT analyses, they were retained for the analyses of cumulative memory span and error types, as they occurred with insufficient frequency (only 4.96% of CWS's responses and 6.34% of CWNS's responses contained disfluencies) and of those produced, most were interjections.

As previously indicated, children completed all six trials in each condition, including those that exceeded their memory spans. This procedure differs from those of most other span studies, in which the task is typically discontinued once children reach their memory span. This procedure was implemented to calculate the proportion of words recalled correctly in each condition regardless of order, a procedure that is designed to avoid potential floor effects with young children (Lee et al., 2010; cf. Gathercole & Baddeley, 1995). Our original intention was to measure both cumulative memory span, which emphasizes order memory, and proportion of words recalled correctly, which emphasizes item memory. However, the findings of the proportion correct data were identical to those of cumulative memory span, presumably because few order errors were produced (see subsequent Results and Discussion). Thus, the proportion correct data were not reported.

Statistical analyses. All statistical analyses were conducted using IBM SPSS Statistics for Windows, Version 23 (Armonk, NY: IBM Corp). The main dependent variables were assessed for their suitability for parametric testing using various data screening procedures (e.g., outliers, normality, homogeneity of variance, etc.). Two variables failed to meet the normality assumption—cumulative memory span and proportion of error types—and transformations proved to be unsuccessful in correcting the underlying distributions. Thus, these variables were analyzed using nonparametric tests (i.e., Mann Whitney U tests, Wilcoxon Signed Rank tests).[2] The alpha levels for minimum statistical significance were adjusted to reduce the possibility of type I errors (i.e., familywise error rate) using Bonferroni corrections (based on an alpha level of .05), with separate (familywise) groupings of between- and within-group tests for the analyses of cumulative memory span and error types. Pearson's correlation coefficient *r* was used as the effect size measure, with an *r* of .50 representing a "large" effect, .30 a "medium" effect, and .10 a "small" effect (Cohen, 1988, 1992).

Because SRT for correct responses met parametric criteria, this dependent variable was assessed using a mixed model analysis of covariance (ANCOVA) with condition (similar or dissimilar) as the within-subjects variable and group (CWS or CWNS) as the between-subjects variable. Chronological age was added to the model as a covariate, as it was significantly correlated with the main dependent variables for both groups of children. As a measure of the strength of the association, the effect size indicator partial eta square (η^2_p) is reported for each statistical comparison, with .14 representing a "large" effect, .06 a "medium" effect, and .01 a "small" effect (Cohen, 1988).

Results

Cumulative memory span. Children's cumulative memory span performance varied across groups and conditions. For the phonologically similar condition, 22 (52.4%) CWS and 24 (57.1%) CWNS obtained spans of two or less, 11 (26.2%) CWS and 6 (14.3%) CWNS obtained spans between two and three, and 9 (21.4%) CWS and 12 (28.6%) CWNS obtained spans of three or more. A chi-square analysis

---

[2] Most, if not all, studies in the memory literature use parametric statistics to analyze memory span and error data, even with young children whose data are less likely to be normally-distributed (e.g., Cowan, 1992; Cowan et al., 1994; Monnier & Bonthoux, 2011), perhaps because some statisticians have argued that parametric statistics are robust against violations of the normality assumption (e.g., Rasch & Guiard, 2004). However, other statisticians argue that parametric statistics are not appropriate under these circumstances (e.g., Erceg-Hurn & Mirosevich, 2008). In this study, we opted to follow the latter advice, but note that we also analyzed the data using analysis of variance techniques and the results were identical to the nonparametric results.

revealed no significant difference in the distribution of memory spans (two or less, between two and three, or three or more) between the CWS and CWNS, $\chi^2(1, N = 84) = 1.99$, $p = .37$. For the phonologically dissimilar condition, 25 (59.5%) CWS and 22 (52.4%) CWNS obtained spans of two or less, 13 (31.0%) CWS and 10 (23.8%) CWNS obtained spans between two and three, and 4 (9.5%) CWS and 10 (23.8%) CWNS obtained spans of three or more. These differences in memory span distribution between the two groups of children were not statistically significant, $\chi^2(1, N = 84) = 3.15$, $p = .21$. Thus, the CWS and CWNS were similar in that most (just over half) had spans of two or less in both the phonologically similar and dissimilar conditions.

Mann-Whitney U tests were used to examine between-group differences in the phonologically similar and dissimilar conditions (see Figure 1). Findings revealed no significant difference in cumulative memory span between the CWS (*Mdn* = 2.00, *M rank* = 42.70) and CWNS (*Mdn* = 2.00, *M* rank = 42.30) in the phonologically similar condition, $U = 873.50$, $z = -.08$, $p = .94$, $r = -.009$. However, in the phonologically dissimilar condition, the children in the CWS group (*Mdn* = 2.00, *M* rank = 36.44) exhibited significantly shorter cumulative memory spans than the children in the CWNS group (*Mdn* = 2.00, *M* rank = 48.56), $U = 1136.50$, $z = 2.37$, $p = .02$, $r = .26$.

Within-group differences across conditions for each group of children were examined using Wilcoxon Signed Rank tests. Findings revealed that the cumulative memory spans of the CWNS were significantly worse in the phonologically similar condition, $T = 567.00$, $p < .001$, $r = .40$, compared to the phonologically dissimilar condition (*Mdn* difference = .88), but this difference was not significant for the CWS, $T = 229.00$, $p = .80$, $r = .03$ (*Mdn* difference = .01). The absolute difference between the two group of children (CWS: *M* rank = 36.06; CWNS: *M* rank = 48.94) in difference scores (i.e., phonologically dissimilar condition – phonologically similar condition = difference) was also statistically significant, $U = 1152.50$, $z = 2.45$, $p = .01$, $r = .27$.

Individual differences in the absence and presence of the PSE were also explored. Table 1 shows the number of CWS and CWNS who demonstrated or failed to demonstrate (zero or reversal effect) the PSE. It is clear from the table that, despite the significant effects of phonological similarity for the CWNS group, some children in this group failed to exhibit the PSE. Likewise, although the CWS, as a group, did not demonstrate the PSE, some individual CWS exhibited the PSE. A chi-square analysis revealed that the association between group and the PSE distribution was statistically significant, $\chi^2(1, N = 84) = 6.87$, $p = .009$. Thus, despite the individual exceptions, the CWS were still significantly less likely to exhibit the PSE, whereas the CWNS were more likely to exhibit the PSE, consistent with the other findings.

To examine the possibility that the absence of the PSE in our participants is associated with lower performance in the easier dissimilar condition, additional analyses were conducted, comparing the within-group differences in cumulative memory span in the dissimilar condition for children who exhibited the PSE versus children who either failed to exhibit the PSE or exhibited a reversal effect. Findings revealed that the 28 CWS who failed to exhibit the PSE had significantly lower cumulative memory spans in the dissimilar condition (*Mdn* = 2.00, *M* rank = 18.73) than the 14 CWS who exhibited the effect (*Mdn* = 2.25, *M* rank = 27.04), $U = 118.50$, $z = -2.11$, $p = .03$, $r = -.23$. The same trend occurred in the CWNS, with the 16 CWNS who did not exhibit the PSE exhibiting lower memory spans in the dissimilar condition (*Mdn* = 2.00, *M* rank= 19.53) than the 26 CWNS who exhibited the effect (*Mdn* = 2.13, *M* rank = 22.71); however, this difference was not statistically significant, $U = 176.50$, $z = -.89$, $p = .38$, $r = -.10$.

To determine whether chronological age could have affected the previous findings (e.g., perhaps the children who failed to exhibit the PSE had lower memory spans because they were younger in age), between group differences in chronological age between the children who did not exhibit the PSE and those who did were examined. Findings revealed no significant difference in chronological age (in months) between the CWS who did (*Mdn* = 47.50, *M* rank = 20.54) and did not (*Mdn* = 54.50, *M* rank =

21.98) exhibit the PSE, *U* = 209.50, *z* = .36, *p* = .72, *r* = .04. Similar non-significant findings appeared for the CWNS who did (*Mdn* = 49.50, *M* rank = 20.92) and did not (*Mdn* = 51.00, *M rank* = 22.44) exhibit the PSE, *U* = 223.00, *z* = .39, *p* = .70, *r* = .04. Thus, these findings suggest that children who failed to exhibit the PSE were comparable in age to those who exhibited the PSE.

Error types. Fewer than five children in each group exhibited order errors in the phonologically similar and dissimilar conditions and those who did exhibited only one of these errors. Thus, order errors were not included in the present analyses. Responses containing more than one type of error (e.g., both an intrusion and omission) were also excluded. Between-group differences in the remaining error types (omissions, intrusions, and false alarms) were analyzed using Mann-Whitney U tests and within-group differences were analyzed using Wilcoxon Signed Rank tests.

For the phonologically similar condition, the CWS produced more omissions (*Mdn* = 33.33, *M* rank = 45.33) and fewer intrusions (*Mdn* = 0.00, *M* rank = 38.31) than the CWNS (*Mdn* = 25.00, *M* rank = 39.67 and 46.69, respectively), but these differences were not statistically significant (Omissions: *U* = 763.00, *z* = -1.10, *p* = .27, *r* = -.12; Intrusions: *U* = 1058.00, *z* = 1.67, *p* = .09, *r* = .18). However, the CWS (*Mdn* = 0.00, *M* rank = 47.39) exhibited significantly more false alarms than the CWNS (*Mdn* = 0.00, *M* rank = 37.61), *U* = 696.50, *z* = -2.11, *p* = .02, *r* = -.23.

Within-group analyses of item errors, which includes both omissions and intrusions, in the phonologically similar condition revealed that the CWS produced significantly more omissions than intrusions, *T* = 71.50, *p* = .005, *r* = -.31. The CWNS, on the other hand, produced comparable proportions of both omissions and intrusions, *T* = 218.00, *p* = .77, *r* = -.03.[3]

For the phonologically dissimilar condition, the CWS produced more omissions (*Mdn* = 33.33) and fewer intrusions (*Mdn* = 8.34) than the CWNS (*Mdn* = 20.00 and 25.00, respectively). After

---

[3] Within-group analyses of false alarms in the phonologically similar (and dissimilar) condition were not conducted because there are no a priori assumptions concerning the effect of phonological similarity on these types of errors.

Bonferroni correction, there no longer was a significant difference between the CWS (*M* rank = 47.69) and CWNS (*M* rank = 37.31) in the proportion of omissions produced in the phonologically dissimilar condition, *U* = 664.00, *z* = -2.01, *p* = .04, *r* = -.22, but the between-group difference in the proportion of intrusions remained significant (CWS: *M* rank = 36.89; CWNS: *M* rank = 48.11), *U* = 1117.50, *z* = 2.20, *p* = .02, *r* = .24, with CWNS producing more intrusions than the CWS. The proportion of false alarms was also significantly higher for the CWS group (*Mdn* = 25.00, *M* rank = 48.81) than the CWNS group (*Mdn* = 0.00, *M* rank = 36.19), *U* = 617.00, *z* = -2.58, *p* = .01, *r* = -.28.

The within-group analyses of item errors in the phonologically dissimilar condition indicated that the CWS produced significantly more omissions than intrusions, *T* = 119.50, *p* = .01, *r* = -.28. However, like the phonologically similar condition, there was no significant difference in the proportion of omissions and intrusions for the CWNS, *T* = 346.00, *p* = .12, *r* = .17.

Across conditions, within-group analyses indicated that the CWS produced comparable proportions of omissions, *T* = 283.00, *p* = .80, *r* = -.03, intrusions, *T* = 192.50, *p* = .93, *r* = .01, and false alarms, *T* = 291.00, *p* = .11, *r* = .17, in the phonologically similar and dissimilar conditions. There were also no significant differences across conditions in the proportion of omissions, *T* = 132.00, *p* = .10, *r* = -.18, intrusions, *T* = 316.00, *p* = .18, *r* = .15, and false alarms, *T* = 113.50, *p* = .45, *r* = .08, for the CWNS.

Speech reaction time. Prior to analyzing the SRT data, SRT values for correct responses that were more than two standard deviations above or below the mean were treated as outliers and, thus, eliminated from the final data corpus for each group of children in each task (see Ratcliff, 1993). This resulted in the removal of 3.5% of the data in the phonologically similar condition and .92% of the data in the phonologically dissimilar condition. Following the removal of outliers, children who had fewer than three (< 50%) useable SRT responses in a condition were excluded from the analyses, along with their matched pairs. This resulted in the removal of 22 pairs of children across both conditions, leaving 20 children in each group (*N* = 40).

On a descriptive basis, the CWS were slower to initiate speech than the CWNS in both the phonologically similar (CWS: adjusted *M* = 975.26 ms, *n* = 20; CWNS: adjusted *M* = 887.41 ms, *n* = 20) and dissimilar (CWS: adjusted *M* = 968.29 ms, *n* = 20; CWNS: adjusted *M* = 920.41 ms, *n* = 20) condition (see Figure 2). However, the mixed-model ANCOVA failed to reveal a significant main effect of group, F (1, 37) = 2.93, p = .09, η^2_p = .07, condition, F (1, 37) = 1.19, p = .28, η^2_p = .03, or group by condition interaction, F (1, 37) = .33, p = .57, η^2_p = .009. Although the covariate main effect was significant for chronological age, F (1, 37) = 8.34, p = .006, η^2_p = .18, the interaction effect between chronological age and condition was not significant, F (1, 37) = 1.36, p = .25, η^2_p = .04.

Discussion

This experiment explored the vSTM of CWS and CWNS using a forward span task consisting of two conditions: phonologically similar and dissimilar. The stimuli in each condition were sampled from an open set of CVC words with and without overlapping phonemes, and auditorily presented to the children in lists ranging from 2 to 4 words. The children verbally recalled the words in each list when prompted to do so. Thus, both the mode of presentation and recall were auditory/verbal, which tends to be easier for children than visual presentation, but more difficult than visual recall (Jarrold & Citroën, 2013). By presenting the words auditorily, recoding demands are also eliminated or reduced, providing direct access to the phonological code (Henry et al., 2012; Jarrold & Citroën).

As previously indicated, phonologically similar words tend to be more difficult for children to recall than phonologically different words, a phenomenon known as the PSE (e.g., Cowan et al., 1991). According to Baddeley (1986; cf. Baddeley, Eysenck, & Anderson, 2009), this effect occurs because phonologically similar words have fewer distinguishing features, making their memory traces more likely to be confused in the phonological store or during their retrieval from the store. Typically-developing preschool children are sensitive to phonological similarity, particularly when the stimuli are presented auditorily (Cowan et al., 1991; Hulme, 1987; Hulme & Tordoff, 1989). However, some studies have demonstrated absent or reduced PSEs in children with developmental disorders, such as specific language impairment and reading disorders (e.g., Coady et al., 2013; Gathercole & Baddeley, 1990; Kibby, 2009; Kibby, Marks, Morgan, & Long, 2004; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977).

The memory span findings of Experiment 1 demonstrated that phonological similarity affected the performance of the CWNS, but not the CWS. That is, while memory span for phonologically similar words was comparable across groups, the CWS were less affected by the phonological qualities of the words than the CWNS due to their poor memory span for phonologically dissimilar words. Analyses of individual differences further revealed that the PSE was absent in the majority of the CWS, but not all—some CWS exhibited the effect and, for that matter, some CWNS failed to exhibit the effect. The CWS who did not exhibit the PSE had lower memory spans in the easier dissimilar condition, which could not be accounted for by chronological age. The same was not true of the CWNS who failed to exhibit the PSE—their memory spans in the dissimilar condition were comparable to those who exhibited the effect. That there were no significant differences in the preparatory interval (i.e., SRT) between the two groups of children suggests that that the CWS were not sacrificing accuracy for speed (i.e., not choosing to respond quickly while sacrificing memory span) and that the failure of CWS to exhibit the PSE was not due to difficulties with motor planning processes (see Cowan et al., 1994, and Gathercole & Baddeley, 1990).

Only one study, to our knowledge, has examined the effect of phonological similarity in CWS. In addition to NWR and digit/letter span tasks, Oyoun et al. (2010) administered a word span task consisting of rhyming (similar) and nonrhyming (dissimilar) words to 60 Persian children (30 CWS, 30 CWNS), ranging from 5 to 13 years of age. The authors provided scant detail about the task other than that the children had to repeat the words in the order they heard them for at least one of two list lengths before the task was terminated. It would appear that while both groups of children recalled more dissimilar words than similar ones, there were no significant differences between the CWS and CWNS for both sets of words. It is not clear if the authors examined within-group differences across word sets for each group of children separately, but on a descriptive basis, the CWNS had a greater mean difference between the similar and dissimilar word sets (*M* = 1.87) than the CWS (*M* = 1.57). Thus, consistent with the findings of the present experiment, it appears that the CWS in the Oyoun et al. study may have exhibited a reduced PSE compared to the CWNS.

The finding that, when compared to the CWNS, the memory span of the children in the CWS group was lower in the dissimilar condition suggests that these children, as a group, have reduced vSTM capacity and that this reduction in capacity is associated with absent, reduced, or reversed effects of phonological similarity. Of course, not all CWS failed to demonstrate the PSE, but there were many more CWS who did not demonstrate the effect than did, and these children had lower memory spans in the easier dissimilar condition. Of importance, some investigators have suggested that an absent or reduced PSE is representative of difficulty with phonological encoding (e.g., Kibby, 2009; Liberman et al., 1977; Nittrouer, Caldwell-Tarr, & Lowenstein, 2003). For, as noted by Coady et al. (2013), if children have difficulty extracting segmental and metrical information from words, they will be less susceptible to interference from other words that are phonologically similar. Further, words that are phonologically dissimilar would be no more phonologically distinct than phonologically similar words (i.e., all word lists are processed as though they are equally challenging and, thus, all compete at retrieval; see Nittrouer et al., 2003). In sum, these memory span findings suggest that CWS have reduced vSTM capacity due to difficulty representing information in the phonological store, and these differences may be due to the phonological encoding skills of CWS.

The results of the error analyses revealed three main findings: (a) the CWS and CWNS produced few order errors and significantly more item errors (omissions and intrusions) in both conditions, (b) the

CWS and CWNS groups produced different proportions of omissions and intrusions in both conditions, and (c) the CWS produced more false alarms than the CWNS in both conditions.

With respect to the first main finding, that children produced few order errors and more item errors in Experiment 1 was not entirely unexpected, given that the stimuli were drawn from an open set of words without replacement. As noted by Archibald and Gathercole (2007a; cf. Campoy & Baddeley, 2008), order errors tend to predominate over item errors when stimuli are selected from among a small, closed set of familiar words. However, when stimuli are selected from an open set of words, the opposite pattern tends to emerge: item errors occur more frequently than order errors. It is clearly easier to remember a small number of frequently appearing words than it is to remember many words that appear infrequently in a task. On the other hand, when stimuli are sampled from a small set of words, the same words will necessarily be repeated across lists in different orders, thereby increasing the likelihood of proactive interference for order information. Since proactive interference is absent when stimuli are sampled from an open set of words without replacement, fewer order errors should occur.

Relative to the second main finding, while not all comparisons were statistically significant, there was a strong tendency for the CWS to produce more omissions and fewer intrusions than the CWNS in the similar, as well as dissimilar condition. All between-group differences in omissions and intrusions at least approached significance, except for the proportion of omissions in the similar condition, and the CWS produced significantly more omissions than intrusions in this condition, as well as the dissimilar condition. To make sense of these findings, it is helpful to consider the amount of information that is retained in different types of errors.

If a child produces a word out of order (i.e., an order error), then it is reasonable to suggest that the target or item information had been retained, but not the order information (Archibald & Gathercole, 2007b). A child who fails to repeat a word that was presented (i.e., omission) or gives a word that had not been presented (i.e., intrusion), according to Archibald and Gathercole, is failing to retain not only the item information, but also the order information. This supposition appears entirely reasonable; however, it is likely the case that intrusions represent the retention of more information than most omissions because a child who gives a non-presented word at least had knowledge that *some* word needed to be produced there. The following example illustrates this point: Child A and Child B are presented with the following three-word list: *cat, spoon, bed*. Child A responds with "*cat, spoon*" and Child B responds with "*cat, spoon, boy*." Thus, Child A produced an omission and Child B produced an intrusion. The third word in Child B's response is incorrect, but it demonstrates that the child remembered that the list contained three words, whereas there is no evidence of this with Child A. Of course, it is possible that Child A was aware that the list contained three words but was uncertain of the word and chose only to produce only the words of which he was certain. Children differ in their willingness to take risks and guess in these tasks. Level of certainty in a particular instance may also play a role in the child's decision to guess the word. Therefore, when a child does take the risk and thereby produces an intrusion, it is possible to credit the child with remembering that *some* word belonged there. When a child simply omits a word, we cannot infer why but only that the word was not recalled. Thus, production of intrusions may reflect the retention of more information than many omissions, but clearly they do not reflect as much information as order errors. In this way, omissions can be construed as a less sensitive indicator of difficulty with vSTM (cf. Monnier & Bonthoux, 2011).

In applying this to the present findings, the fact that the CWS produced more omissions than intrusions, whereas the CWNS tended to produce more intrusions or comparable proportions of intrusions and omissions suggests, through their error patterns, that the CWS recalled less information or demonstrated less recall of information than did CWNS. These findings, along with those of memory span, provide further evidence to suggest that CWS may have weaknesses in vSTM.

The third main finding—that the CWS produced significantly more false alarms than the CWNS[4]—may suggest an increased eagerness to respond as a memory aid. That is, children with weaker vSTM may attempt to enhance their performance by responding as quickly as possible following the prompt. The result of doing so may be an increase in the number of false alarms produced. A second, perhaps more empirically supported explanation is that CWS may have weaker inhibitory processes, the result of which is impulsivity. Impulsive behaviors are a consequence of weaknesses in inhibition coupled with a desire to respond (Bari & Robbins, 2013). In the present study, an impulse to respond was created by instructing the children to repeat what the parrot "said"; however, children had to withhold their response until they were prompted to do so visually with the appearance of the question mark, an act that requires functional inhibitory processes. Thus, the CWS in the present study had more difficulty than the CWNS in inhibiting their desire to respond, consistent with the findings of Eggers, De Nil, and Van den Bergh (2013). In the Eggers et al. (2013) study, CWS and CWNS between the ages of 4;10 and 10;0 completed a simple response inhibition task, the go/no-go task, in which they had to press a button when they saw a symbol of a green man running, but not when they saw a symbol of a red man standing. On this task, the CWS produced significantly more false alarms (pressing the button when they saw the red man standing) and premature responses (pressing the button between 0 and 200 ms after the onset of the stimulus) than the CWNS. These findings were interpreted by the authors to suggest that CWS have a more impulsive, less controlled response style. The present findings are also consistent with the findings of a recent study by Anderson and Wagovich (2017), in which preschool CWS (some of

[4] In theory, it could be that the production of a false alarm impacts a child's subsequent ability to recall a list of items or increases the likelihood of an omission in the list (because the child might not have heard/processed the entire list due to the false alarm). We acknowledge this possibility; however, in the present study, it is less likely to have impacted findings substantially. The CWS exhibited significantly more false alarms in both conditions compared to the CWNS, but cumulative memory span and omissions only differed between the two groups of children in the phonologically dissimilar condition. Thus, if false alarms were impacting word encoding, then the CWS should have exhibited reduced memory spans and increased omissions in both conditions, not just one of them.

whom were the same as in this study) were found to have more difficulty than CWNS with a complex verbal response inhibition task.

Experiment 2

Method

Participants. Participants were identical to those in Experiment 1.

Procedures. In addition to the testing procedures described in Experiment 1, children completed the semantic category task (SCT), which is similar to the PST, except that semantically homogeneous and heterogeneous words were used to measure semantic contributions to vSTM in children. The semantically homogeneous condition contained 18 single-syllable words from taxonomic categories familiar to young children (gender, furniture, vehicles, kitchen items, four-legged animals, and clothing; see Table A2 of the Appendix for the list of experimental stimuli). The homogeneous words were selected from among those used by Lee et al. (2010) in their study of vSTM deficits in children and young adults with Down syndrome. However, with one exception, the categories used in the present study were rearranged into different trial list levels to ensure that all lists contained words that were comparably low in AoA and high in word frequency. Through this process, some of the words in the category lists of Lee et al. were removed (e.g., the 5-word vehicle list from Lee et al. [*bus, train, car, boat, truck*] was presented as a 4-word list [*car, boat, truck, bus*] in the present study). In addition, there was one instance in which a later acquired, lower frequency word was substituted for an earlier acquired, higher frequency word (i.e., the 4-word furniture list from Lee et al. [*chair, lamp, desk, couch*] was presented as a 3-word list in the present study and the word, *bed*, was substituted for lamp/desk [*chair, couch, bed*]).

The heterogeneous condition contained the same 18 words used in the homogeneous condition, but the words were rearranged to create lists of words from different semantic categories. The same list of words was used across conditions to ensure that the words in both conditions were comparable in their psycholinguistic characteristics (e.g., AoA, word frequency, etc.). Independent samples *t*-tests revealed no significant differences in AoA, word frequency, positional segment frequency, concreteness, and imageability across trial list levels (e.g., 2 words vs. 3 words, etc.) for both the semantically homogeneous (p = .15 to .85) and heterogeneous (p = .11 to .58) conditions.

Like the PST, the six word lists in each condition ranged from 2 to 4 words in length, with each list having two trials and no item occurring more than once. Because the same words were used in both conditions, one condition was presented at the beginning of the session and the other at the end of the session, counterbalanced across participants. The materials and procedures (e.g., instructions, practice phases, presentation, etc.), scoring, and statistical analyses were otherwise identical to Experiment 1. While excluded from the SRT analyses, responses containing speech disfluencies were retained for the analyses of cumulative memory span and error types for the same reason indicated in Experiment 1 (only 3.37% of CWS's responses and 3.17% of CWNS's responses contained disfluencies).

Results

Cumulative memory span. Children's cumulative memory span performance was somewhat variable across groups and conditions. For the semantically homogeneous condition, 12 (28.6%) CWS and 9 (21.4%) CWNS obtained spans of two or less, 12 (28.6%) CWS and 20 (47.6%) CWNS obtained spans between two and three, and 18 (42.9%) CWS and 13 (31.0%) CWNS obtained spans of three or more. A chi-square analysis revealed no significant difference in the distribution of these memory spans (two or less, between two and three, or three or more) between the CWS and CWNS, $\chi^2(1, N = 84) = 3.24$, $p = .20$. For the semantically heterogeneous condition, 19 (45.2%) CWS and 14 (33.3%) CWNS obtained spans of two or less, 10 (23.8%) CWS and 14 (33.3%) CWNS obtained spans between two and three, and 13 (31.0%) CWS and 14 (33.3%) CWNS obtained spans of three or more. These differences in memory span distribution between the two groups of children were not statistically significant, χ^2(1, N =

84) = 1.46, *p* = .48. Thus, the CWS and CWNS were similar in that most children had spans greater than two in both the semantically homogeneous and heterogeneous conditions.

Mann-Whitney U tests revealed no significant differences between CWS and CWNS in cumulative memory span for either the semantically homogeneous (CWS: *Mdn* = 2.50, *M* rank = 42.23; CWNS: *Mdn* = 2.25, *M* rank = 42.77), *U* = 893.50, *z* = .10, *p* = .92, *r* = .01, or the heterogeneous (CWS: *Mdn* = 2.25, *M* rank = 41.86; CWNS: *Mdn* = 2.25, *M* rank = 43.14) condition, *U* = 909.00, *z* = .25, *p* = .81, *r* = .03 (see Figure 3).

Wilcoxon Signed Rank tests revealed no significant difference between conditions for children in the CWS group (*Mdn* difference = .25), *T* = 271.00, *p* = .32, *r* = -.11, but the same was not true of the CWNS. In particular, the CWNS exhibited significantly shorter cumulative memory spans in the semantically heterogeneous condition compared to the semantically homogeneous condition (*Mdn* difference = .25), *T* = 119.00, *p* = .007, *r* = -.30. However, the absolute difference across conditions (i.e., semantically homogeneous condition – semantically heterogeneous condition = difference) between the two groups of children (CWS: *M* rank = 39.45; CWNS: *M* rank = 45.55) was not statistically significant, *U* = 1010.00, *z* = 1.15, *p* = .25, *r* = .13.

Individual differences in the number of CWS and CWNS who demonstrated or failed to demonstrate the SSE (i.e., the tendency for memory recall to be higher for semantically homogeneous than heterogeneous words) were also examined. As revealed in Table 2, the distribution of children who did and did not exhibit the SSE was identical in both groups of children; a chi-square analysis revealed that the association between group and the SSE distribution was not statistically significant, χ^2(1, *N* = 84) = 0.00, *p* = .99.

There is no precedence in the literature to suggest that children who fail to exhibit the SSE have shorter word spans in the easier, semantically homogeneous condition than those who exhibit the SSE. However, given the precedence in this regard related to the PSE and for ease of comparison across experiments, we adopted a similar approach for the analysis of the SSE; namely, within-group differences in cumulative memory span were examined for the children who did and did not exhibit the SSE. Findings revealed that the 19 CWS who failed to exhibit the SSE (zero or reversal effect) had significantly lower cumulative memory spans in the semantically homogeneous condition (*Mdn* = 0.50, *M* rank = 15.05) than the 23 CWS who exhibited the SSE (*Mdn* = 3.00, *M* rank = 26.83), *U* = 96.00, *z* = -3.14, *p* = .002, *r* = -.34. While the CWNS exhibited the same pattern, with the 19 CWNS who failed to exhibit the SSE having lower cumulative memory spans (*Mdn* = 2.25, *M* rank = 21.30) than the 23 CWNS who exhibited the effect (*Mdn* = 2.25, *M* rank = 21.74), this difference was not statistically significant, *U* = 223.00, *z* = .12, *p* = .91, *r* = .01.

Further analyses revealed no significant difference in chronological age between the CWS who did (*Mdn* = 55.00, *M* rank = 24.80) and did not exhibit the SSE (*Mdn* = 44.00, *M* rank = 17.88), *U* = 147.50, *z* = -1.83, *p* = .07, *r* = -.28. The same was true for the CWNS who exhibited the SSE (*Mdn* = 49.00, *M* rank = 20.57) and those who did not (*Mdn* = 55.00, *M* rank = 22.63), *U* = 240.00, *z* = .54, *p* = .59, *r* = .06. These findings suggest that the lower memory spans in the easier homogeneous condition for the children who did not exhibit the SSE were not due to their being younger in age.

Error Types. Like Experiment 1, order errors were excluded from the analysis because fewer than six children in each group exhibited order errors in the semantically homogeneous and heterogeneous conditions, and those who did exhibited only one of these types of errors. Responses containing more than one type of error were also excluded from the analyses. The remaining error types (omissions, intrusions, and false alarms) were analyzed between-groups using Mann-Whitney U tests and within-groups using Wilcoxon Signed Rank tests.

For the semantically homogeneous condition, the proportion of omissions were comparable for both groups of children (CWS: *Mdn* = 0.00, *M* rank = 43.24; CWNS: *Mdn* = 8.34, *M* rank = 41.76), *U* = 851.00, *z* = -.30, *p* = .77, *r* = -.03. While the CWS (*Mdn* = 0.00, *M* rank = 39.05) produced fewer intrusions than the CWNS (*Mdn* = 0.00, *M* rank = 45.95) in the semantically homogeneous condition, this difference was not statistically significant, *U* = 1027.00, *z* = 1.71, *p* = .09, *r* = .19. The CWS also produced more false alarms (*Mdn* = 29.17, *M* rank = 45.63) than the CWNS (*Mdn* = 0.00, *M* rank = 39.37) in the semantically homogeneous condition, but this too failed to reach significance, *U* = 750.50, *z* = -1.27, *p* = .21, *r* = -.14.

Within-group analyses of item errors indicated that the CWS produced significantly more omissions than intrusions in the semantically homogeneous condition, *T* = 39.00, *p* = .01, *r* = -.27. However, the proportion of omissions and intrusions in the semantically homogeneous condition were comparable for the CWNS, *T* = 128.00, *p* = .35, *r* = -.10.

For the semantically heterogeneous condition, the CWS (*Mdn* = 16.67, *M* rank = 42.51) and CWNS (*Mdn* = 8.34, *M* rank = 42.49) again produced a similar proportion of omissions, *U* = 881.50, *z* = -.005, *p* = .99, *r* < .005. The CWS (*Mdn* = 0.00, *M* rank = 38.14) produced fewer intrusions than the CWNS (*Mdn* = 0.00, *M* rank = 46.86) in the semantically heterogeneous condition, a difference which only approached significance, *U* = 1065.00, *z* = 1.86, *p* = .06, *r* = .20. However, the proportion of false alarms was significantly higher for the CWS (*Mdn* = 25.00, *M* rank = 49.39) than the CWNS (*Mdn* = 0.00, *M* rank = 35.61), *U* = 592.50, *z* = -2.90, *p* = .004, *r* = -.32, in the semantically heterogeneous condition.

The within-group analyses indicated that despite the higher proportion of omissions than intrusions in the semantically heterogeneous condition for the children in the CWS group, this difference was not significant, *T* = 99.50, *p* = .15, *r* = -.16. The proportion of omissions and intrusions in the semantically heterogeneous condition was also similar for the CWNS, *T* = 222.50, *p* = .65, *r* = .05.

Across conditions, within-group analyses indicated that the CWS produced comparable proportions of omissions, *T* = 163.50, *p* = .37, *r* = -.10, intrusions, *T* = 80.00, *p* = .25, *r* = .12, and false alarms, *T* = 205.50, *p* = .96, *r* = .006, in the semantically homogeneous and heterogeneous conditions. There were also no significant differences across conditions in the proportion of omissions, *T* = 211.50, *p*

= .85, *r* = .02, intrusions, *T* = 185.00, *p* = .06, *r* = .21, and, after Bonferroni correction, false alarms, *T* = 79.00, *p* = .04, *r* = -.22, for the CWNS.

Speech reaction time. Like the PST, the SCT values for correct responses, which were determined using CheckVocal (Protopapas, 2007), that were more than two standard deviations above or below the mean were removed from the final data corpus for each group of children as outliers. As a result, 1.68% of the data were removed from the semantically homogeneous condition and 2.08% of the data were removed from the semantically heterogeneous condition. Following the removal of outliers, children who had fewer than three (< 50%) useable SRT responses in a condition were excluded from the analyses, along with their matched pairs. This resulted in the removal of 12 pairs of children across both conditions, leaving 30 children in each group (*N* = 60).

Findings from the mixed-model ANCOVA revealed a significant main effect of group, *F* (1, 57) = 8.82, *p* = .004, η^2_p = .13, with the CWS (adjusted *M* = 852.77 ms, *n* = 30) exhibiting faster SRT across both conditions than the CWNS (adjusted *M* = 989.59 ms, *n* = 30; see Figure 4). The main effect of condition (Homogeneous: adjusted *M* = 912.31, *n* = 60; Heterogeneous: adjusted *M* = 930.06, *n* = 60) approached significance, *F* (1, 57) = 2.95, *p* = .09, η^2_p = .05, but the group by condition interaction effect did not, *F* (1, 57) = .25, *p* = .62, η^2_p = .004. Although the covariate main effect was significant for chronological age, *F* (1, 57) = 4.70, *p* = .03, η^2_p = .08, there was no significant interaction effect between chronological age and condition, *F* (1, 57) = 3.35, *p* = .07, η^2_p = .06.

Given the significant main effect of group, the relationship between SRT and cumulative memory span was evaluated for each group of children to determine the potential for speed-accuracy tradeoffs using Spearman’s rank partial correlation coefficients, with chronological age serving as the covariate. For the semantically homogeneous condition, analyses revealed no significant correlations between SRT and cumulative memory span for children in both the CWS (*r* = .28, *p* = .14) and CWNS (*r* = -.14, *p* = .48) groups. Similarly, neither the CWS (*r* = -.05, *p* = .81) nor the CWNS (*r* = .13, *p* = .49) exhibited a significant correlation between SRT and cumulative memory span in the semantically heterogeneous condition.

Discussion

In Experiment 2, semantic contributions to vSTM were examined in CWS and CWNS using the SCT. The SCT is a traditional span task in which children verbally recall auditorily presented words in two conditions: semantically homogeneous and heterogeneous. The stimuli in each condition were sampled from an open set of words from the same or different taxonomic semantic categories, with lists ranging from 2 to 4 words. Thus, like the PST, both the mode of presentation and recall were auditory/verbal.

The SSE, as previously indicated, refers to the tendency for words that are semantically similar to be easier for adults and children to recall than words that are semantically different, which generally suggests that long-term lexical/semantic knowledge influences verbal STM (Monnier & Bonthoux, 2011; Poirier, Saint-Aubin, Mair, Tehan, & Tolan, 2015). Two theoretical accounts have been proposed to explain the SSE. First, according to the *redintegrated hypothesis* (Hulme et al., 1991; Schweickert, 1993), verbal information is initially encoded into phonological forms in vSTM as per Baddeley's multicomponent model (1986, 2003). These phonological traces, however, rapidly degrade due to decay or interference, unless they are rehearsed. If the phonological trace is degraded at the time of recall (and, thus, cannot be produced), then long-term lexical/semantic information is accessed to reconstruct or "fill in" the degraded memory trace. Although semantically similar and dissimilar words are presumed to degrade to the same extent, words that are semantically similar have an advantage because shared semantic information provides an additional retrieval cue, thereby limiting the number of potential lexical candidates to be recalled from long-term memory and, thus, increasing the likelihood that their phonological traces will be correctly reconstructed (Poirier et al., 2015; Saint-Aubin, Ouellette, & Poirier, 2005; Saint-Aubin & Poirier, 1999). Second, according to the *interactive* account, long-term lexical/semantic information is activated immediately upon hearing a word, keeping the memory trace in the phonological store active and, thus, preventing the trace from degrading (Campoy & Baddeley, 2008; Thorn, Gathercole, & Frankish, 2005). With this account, the advantage that semantically similar words have over semantically dissimilar words stems from the fact that their long-term representations are more activated: representations that are more highly activated will increase the level of activation of information in the phonological store, thereby increasing the likelihood that the word will be correctly recalled (Monnier & Bonthoux, 2011).

The findings of Experiment 2 revealed that while the two groups of children exhibited similar memory spans in both conditions of the SCT, the CWS produced significantly more omissions than intrusions in the semantically homogeneous condition which could, as further discussed below, indicate subtle weaknesses in vSTM capacity. The memory span findings also demonstrated that the CWNS benefitted from semantic similarity, whereas the CWS did not. That is, memory span was significantly higher for the CWNS, as a group, when the words to be recalled were from the same semantic category than when they were from a different semantic category. The CWS, on the other hand, were less sensitive to the effect of similarity and, thus, performed similarly in both conditions. Nevertheless, there was no significant difference between the two groups of children in the absolute difference score, and identical proportions of CWS and CWNS either exhibited (55%) or failed to exhibit (45%) the SSE. These latter findings suggest that it is not just a matter of the proportion of children in each group who demonstrated the SSE—rather, when children showed the SSE, it was the *extent* to which performance in the homogeneous condition exceeded that of the heterogeneous condition. The CWS who did not exhibit the SSE also exhibited lower memory spans in the easier homogeneous condition compared to CWS who exhibited the effect, but the same was not true of CWNS.

The finding that the CWS were less sensitive to the SSE than the CWNS should be interpreted with caution, given the lack of convergence in results. Nevertheless, the findings provide *preliminary* evidence to suggest that the CWS, as a group, *may* be less able to use long-term semantic knowledge to support memory performance. According to the redintegration hypothesis, if a degraded trace is not successfully redintegrated, an error will be produced in which the missing information is either incorrectly replaced or not replaced at all (Thorn et al., 2005). While there is some evidence to suggest that the vSTM capacity of the CWS may have been subtly reduced in the homogeneous condition, since they produced significantly more omissions than intrusions, there were otherwise no statistically significant differences between the CWS and CWNS in memory span in either condition. Thus, if redintegration is the force behind the SSE, then it seems reasonable to suggest that we should have seen clearer evidence of reduced memory span for the children in the CWS group.

The interactive account, which has received more support in recent years (Poirier et al., 2015), would seem to provide a more parsimonious explanation for the present findings. As described above, the interactive account implies that long-term knowledge influences activity in the phonological store, such that the memory traces of semantically similar words receive additional activation from the lexical/semantic level over those of semantically dissimilar words. Accordingly, if a child's lexical/semantic representations are not as well established or if the interactive connection between the phonological store and long-term memory is less than optimal, then there will be no activation advantage for semantically similar words. As a result, memory span for semantically similar and dissimilar words will not differ. Our results, considered in this context, tentatively suggest that some CWS, particularly those with lower memory spans, may have weaker semantic associations and/or associative links to the phonological store.

The results of the error analyses suggest that, while the groups did not differ in their errors in the homogeneous condition, they differed in the production of false alarms in the heterogeneous condition, with the CWS producing more false alarms. It is possible that this eagerness to respond on the part of the CWS group, resulting in false alarms, was an attempt to keep the memory trace from decaying. It is also possible that the additional false alarms produced by the CWS reflects difficulty inhibiting a response, as discussed in relation to the previous experiment. On average, the CWS also produced more false alarms than the CWNS in the homogeneous condition, although this difference did not reach significance. In addition, the error analysis revealed that the CWS produced more omissions than intrusions in the homogeneous condition; this pattern was not observed in the CWNS. As discussed in Experiment 1, intrusions may reflect the retention of more information than omissions, because when a child produces an intrusion, the child inserts an incorrect word as a placeholder, demonstrating knowledge that the list contained another word of some sort. The production of more omissions by the CWS may be a subtle indicator of weakness in vSTM among CWS. Neither group produced many order errors, which is likely a result of the fact that the stimuli were from a large, open set of words.

The findings from SRT analyses were largely unremarkable, with one exception: the SRT (preparatory interval) was significantly shorter for the CWS than the CWNS across both conditions of the SCT. However, consistent with Cowan (1992), there were no significant correlations between SRT and memory span for both groups of children. This suggests that children with higher cumulative memory spans take the same amount of time to prepare their responses as do children with lower cumulative memory spans (i.e., there was no evidence of speed-accuracy tradeoffs).

The fact that in this study, the CWS exhibited faster SRTs than the CWNS is a rather unusual finding, as most studies have revealed that young CWS tend to have significantly slower reaction times when completing various tasks (e.g., Anderson & Conture, 2004; Hartfield & Conture, 2006; Pellowski & Conture, 2005). However, in the aforementioned study by Eggers et al. (2013), the CWS also had significantly faster reaction times for false alarms than CWNS. The authors interpreted these findings to suggest that the CWS do not change their speed in response to errors; rather, they persist with a faster response style even though they produced more errors as a result (i.e., a speed-accuracy tradeoff). Another possibility, however, is that the CWS may be attempting to respond quickly following the completion of the stimulus, as a strategy to aid memory. This would explain the increased occurrence of false alarms among the CWS group, as well as the shorter SRTs within this group.

General Discussion

This two-experiment study was undertaken to evaluate vSTM in CWS. Two forward word span tasks were developed, the PST and the SCT, requiring children to recall lists of similar and dissimilar items in the order presented. Of importance, in addition to evaluating performance separately on each of the four list types (phonologically similar/dissimilar; semantically homogeneous/heterogeneous), we also explored the presence of the phonological and semantic similarity effects. To review, together, the PSE and SSE describe the fact that items that are phonologically dissimilar and items that are semantically related (homogeneous) are recalled more easily than their counterparts (phonologically similar and semantically heterogeneous lists). SRT and the nature of children's errors were also examined for each of the two experiments. We discuss the overarching findings, across experiments, and their implications below.

Memory Span

The memory span findings were somewhat more revealing for the PST than for the SCT. The groups differed only in their spans of phonologically dissimilar lists; the CWS produced spans that were shorter than those of the CWNS. No differences were observed for the phonologically similar lists or either of the semantic lists. Performance in the easier condition can be construed as an index of vSTM capacity. Thus, the fact that the CWS exhibited poorer performance in the phonologically dissimilar condition suggests that they may have reduced vSTM capacity.

Of note, the children in our study displayed memory spans that, while clearly shorter than the findings of adult memory spans within the literature, were consistent with reports of the spans of preschool-age children. For example, Cowan et al. (1994) found that their children with a mean age of 60 months had an average cumulative span for short words of 2.80. In comparison, the CWNS in the present study were, on average, 10 months younger than those in the study by Cowan et al., and displayed a mean span of 2.35 in the dissimilar condition (i.e., the condition most comparable to Cowan and colleagues). Thus, the spans obtained for the typically developing children in the present study are consistent with the extant developmental literature in this area.

Implications of Phonological and Semantic Similarity Effects

With regard to the PST, it is noteworthy that, in computing absolute difference values between the phonologically similar and dissimilar items, the CWNS demonstrated a significantly greater absolute difference than the CWS, consistent with our finding that the CWNS demonstrated the expected PSE, whereas the CWS, as a group, did not. Of course, it is not the case that every CWS failed to show the expected PSE or that every CWNS showed the effect. Therefore, we further explored these findings by examining subgroup performance of those who demonstrated the PSE in each group, in comparison to those of each group who did not demonstrate the expected pattern. Importantly, significantly more children in the CWNS group displayed the PSE than children in the CWS group. Age did not appear to be a relevant factor for the CWS or CWNS, in explaining which children demonstrated the PSE.

Perhaps the most parsimonious explanation for why the CWS, as a group, failed to demonstrate the PSE is that it is a consequence of weaknesses in vSTM, for if the size of the effect is proportional to vSTM performance, then one would expect smaller PSEs when vSTM is weak (see Jarrold & Citroën, 2013; Wang, Logie, & Jarrold, 2016 for discussion of the proportional scaling account of the PSE). This interpretation is supported by the fact that the absence of the PSE in CWS was associated with *lower* memory spans in the phonologically dissimilar condition. Overall, this suggests that for the CWS group, the phonologically dissimilar condition was less facilitating or more challenging than expected. In fact, consistent with this idea, the phonologically dissimilar condition was the single condition of the four in which the CWS, as a group, performed significantly less well than the CWNS. However, the converse—that the CWNS exhibited the PSE because they have stronger vSTM—does not entirely follow from the present results because even though there was a trend for the CWNS who failed to exhibit the PSE to have reduced memory spans in the phonologically dissimilar condition than the CWNS who exhibited the effect, this difference was not statistically significant.

While we cannot rule-out the possibility that deficient vSTM is responsible for the reduced PSE found in CWS, absent or reduced PSEs are often interpreted in the literature as being a product of inefficient phonological encoding (e.g., Gathercole & Baddeley, 1990; Kibby, 2009; Liberman et al., 1977; Nittrouer et al., 2003), because word span tasks not only require vSTM, but also phonological encoding. During a word span task, efficient phonological encoding enables children to break words down into their constituent phonemes, making them more likely to experience interference from other phonologically similar words in the phonological store or during retrieval, the consequence of which is the PSE (Coady et al., 2013). In contrast, if children are less able to exploit phonological differences among words in a list, then these words will be less vulnerable to interference from other words and, hence, the PSE will be absent or reduced. Thus, according to this conceptualization, the fact that the CWS, as a group, failed to demonstrate the PSE may indicate weaknesses in phonological encoding. Of course, it is also possible that both interpretations are correct: weaknesses in vSTM coupled with subtle to not-so-subtle difficulties with phonological encoding resulted in reduced or absent PSE in the CWS.

With regard to the SCT, though the groups did not differ in their performance in either condition, within-group analyses revealed that the CWNS demonstrated the expected SSE, with performance in the homogeneous condition significantly exceeding that of the heterogeneous condition. This pattern was not observed for the CWS. That is, word lists that were semantically similar (e.g., *chair*, *couch*, *bed*) did not facilitate vSTM performance for CWS, in contrast to the performance for CWNS. While these within-group findings are intriguing, they are tempered by the fact that there was no significant between-group difference in absolute difference scores. There was also no significant difference between groups in the number of children who displayed the SSE. The CWS who did not display the SSE, however, had lower spans in the easier, homogeneous condition, than the CWS who displayed the SSE (with the CWNS displaying a similar trend on a descriptive basis).

Children with strong lexical/semantic knowledge should most effectively exploit their mental integration of words and concepts, resulting in the strongest recall performance on word lists that are semantically related (Monnier & Bonthoux, 2011; Poirier et al., 2015). Both the redintegration (Hulme et al., 1991; Schweickert, 1993) and interaction accounts (Campoy & Baddeley, 2008; Thorn, Gathercole, & Frankish, 2005) predict that memory performance for semantically similar lists will be aided by long-term lexical knowledge, either by providing a retrieval cue (redintegration) or through the activation of related semantic information (interaction) to aid retrieval.

As previously indicated, the interactive account (Poirier et al., 2015) may provide a more compelling explanation for the SSE observed in CWNS and the subtle lack thereof in CWS. The interactive account (that the SSE occurs as a result of stronger activation of words that are semantically related) hinges on strong lexical/semantic representations, allowing for interaction between the phonological store and long-term memory, thereby facilitating memory of semantically similar words. Thus, if CWS display somewhat weaker lexical/semantic representations in long-term memory, theoretically, they may be less able to exploit semantic similarity (through strengthened activations of similar words) to enhance short-term memory of similar words, compared to dissimilar words. This would also account for the finding that the CWS who did not display the SSE had lower spans in the homogeneous condition compared to those who displayed the effect.

As with the PSE, another potential explanation is that absent or reduced SSEs are simply a consequence of weak vSTM. In this way, if CWS have less robust vSTM skills, then smaller SSEs ought to be the result—that is, the CWS ought to perform more similarity in both the semantically homogeneous and heterogeneous conditions. Support for this explanation is admittedly tenuous because there were otherwise no significant memory span differences between the CWS and CWNS in both conditions of the

SCT. Evidence from the present study for the lack of the SSE in the group of CWS is also not as strong as it was for the PSE. While the CWS did not experience a reduction in memory span performance in the heterogeneous condition compared to the homogeneous condition, as their normally-fluent peers did, there was no significant difference between the CWS and CWNS in the absolute difference score.

In sum, the PSE and SSE may provide information about children's vSTM skills and their ability to phonologically encode words and semantically integrate lexical items, respectively. Current findings of reduced phonological and, to some extent, semantic similarity effects provide evidence to suggest that CWS, as a group, may have difficulty with phonological and/or semantic processing skills, at least as they relate to supporting vSTM performance. These findings replicate those of previous studies in which CWS were found to have less robust or less mature lexical/semantic and phonological processing skills, which may manifest themselves as a different phonological or lexical/semantic processing style (Anderson, 2008; Byrd et al., 2007; Hartfield & Conture, 2006), slower speech-language planning processes (Hartfield & Conture, 2006; Pellowski & Conture, 2005), or difficulties with lexical encoding and/or retrieval (Pellowski & Conture, 2005). Such weaknesses could result in degraded phonological and/or lexical representations in memory, which may make them more susceptible to stuttering (see Anderson, 2007, and Anderson & Byrd, 2008, for further discussion). Thus, current results are informative to this body of work in that both phonological and semantic processing skills and their role in verbal short-term memory were considered together, within the same groups of children.

Errors in Memory Span Tasks

Examination of error types across phonological and semantic conditions revealed most notably that the CWS tended to produce more omissions, fewer intrusions, and more false alarms than the CWNS. The presence of more omissions and fewer intrusions among the CWS group in both conditions of the PST and the homogeneous condition of the SCT may indicate subtle difficulties with vSTM. Thus, these findings provide additional support to suggest that CWS may have difficulties with vSTM, particularly for phonological information. A reasonable inference for the finding that the CWS produced significantly more false alarms than the CWNS in both phonological conditions and the more challenging semantically heterogeneous condition is that their response style is more impulsive (cf. Eggers et al., 2013). This is an interesting supplement to the memory span findings, in that it highlights, perhaps, the way children respond to these types of cognitively challenging tasks.

Speech Reaction Time

Like the results pertaining to false alarms discussed above, the SRT findings of the present study, in part, support the notion that the CWS displayed a more impulsive response style. Although SRT did not differ between groups for the phonological conditions, the CWS displayed faster SRTs than the CWNS in the semantic conditions. Results do not point to differences in speed-accuracy tradeoffs between groups, but they do suggest that, as a group, at least for the semantic conditions, the CWS tended to resort to responding quicker than their peers who do not stutter, although this speed difference did not impact their span length.

The SRT findings for the PST and SCT may, at first glance, appear inconsistent. Indeed, compared to the PST, cumulative spans were higher in the SCT, with fewer CWS and CWNS obtaining spans of two or less. This suggests that the SCT was generally easier for both groups of children than the PST. Thus, if CWS have a tendency to be more impulsive, an easier task would have allowed them to respond faster. In fact, on a descriptive basis, the CWS were, on average, 119.01 ms faster in the SCT than the PST, whereas the CWNS showed the opposite effect with an even smaller margin of difference (M = 85.68 ms.). That said, the SRT results across tasks are in some ways fairly consistent; while the between-group difference in overall SRT was significant for the SCT, the same between-group difference approached significance for the PST with a p-value of .09. Moreover, the main effects of condition for both tasks and the interaction between group and condition for both tasks failed to reach significance. Thus, the SRT findings for the two experiments were, for the most part, consistent.

Limitations and Conclusions

This study is the first to explore phonological and semantic vSTM in the same groups of CWS and peers, enabling us to explore vSTM with greater specificity and depth with this population. We acknowledge several limitations of the work. As described in the Method sections, the word stimuli were carefully selected with regard to mean age of acquisition, word frequency, positional segment frequency, concreteness, and imageability. However, it was not possible to control for these variables across experiments, because our primary concern was the phonological or semantic content of the words, consistent with the aims of the study. Post-hoc analyses revealed that the words used in the SCT were significantly higher in word frequency, concreteness, and imageability (meaning that the SCT stimuli were easier in these respects) than the words used in the PST dissimilar condition. On the other hand, the words in the SCT had a significantly higher age of acquisition than the words in the PST dissimilar condition, suggesting that the SCT stimuli were more advanced developmentally. Ideally, to make strong comparisons between phonological and semantic memory findings, the word lists would not have varied on any psycholinguistic factors (e.g., age of acquisition, number of phonemes, word frequency, segment frequency, imageability, etc.). However, the fact that (a) these four factors did not all vary in the same direction and (b) some psycholinguistic factors did not differ across tasks would seem to suggest that neither task was more difficult, at least in terms of the difficulty of the words themselves.

Between the phonological and semantic tasks, the stimuli did differ in that, while the SCT stimuli were the same words, manipulated to create word lists that were homogeneous and heterogeneous, the PST stimuli were, by necessity, different words across the similar and dissimilar conditions. That is, it would have been impossible to use phonologically similar word lists to create phonologically dissimilar word lists. Taken together, these two limitations suggest that, though the findings of each experiment provide insights into phonological and semantic verbal short-term memory, caution is warranted in directly comparing the two experiments. Though every effort was made to keep the tasks comparable, we acknowledge that there were some subtle differences between them.

It is also important to acknowledge that many of the CWS who participated in this study will spontaneously recover from stuttering. Consequently, we do not know how these children might differ in vSTM compared to those who continue to stutter. That said, it seems reasonable to suggest that one must first determine whether a difference exists between CWS and CWNS in a given skill, like vSTM, prior to assessing the impact such a difference has, if any, on stuttering persistence. Of importance, factors that contribute to the onset of stuttering need not be the same as those that are associated with its persistence. Thus, while the issue of persistence versus recovery in early stuttering is clearly an important one, so too is the study of factors associated with onset, regardless of whether they are also germane to persistence.

In summary, findings of the present study suggest a difference in memory span, such that CWS, as a group, demonstrate shorter spans for phonologically dissimilar items. Thus, our findings implicate vSTM as one cognitive difference between CWS and CWNS, adding to the scant literature on short-term/working memory skills in CWS. The differences observed in vSTM and the similarity effects for phonologically similar vs. dissimilar and semantically homogeneous vs. heterogeneous words enable a deeper understanding, not only of short-term memory processes, but also of phonological and semantic processing skills in general and the use of these skills to aid memory.

References

American Speech-Language-Hearing Association. (1997). Guidelines for audiologic screening. Rockville, MD: Author.

Anderson, J. D. (2007). Phonological neighborhood and word frequency effects in the stuttered disfluencies of children who stutter. *Journal of Speech, Language, and Hearing Research, 50,* 229-247.

Anderson, J. D. (2008). Age of acquisition and repetition priming effects on picture naming of children who do and do not stutter. *Journal of Fluency Disorders, 33,* 135-155.

Anderson, J. D., & Byrd, C. T. (2008). Phonotactic probability effects in children who stutter. *Journal of Speech, Language, and Hearing Research, 51,* 851-866.

Anderson, J. D., & Conture, E. G. (2004). Sentence-structure priming in young children who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 47,* 552-571.

Anderson, J. D., Pellowski, M. W., & Conture, E. G. (2005). Childhood stuttering and dissociations across linguistic domains. *Journal of Fluency Disorders, 30,* 219-253.

Anderson, J. D., Wagovich, S. A., & Hall, N. (2006). Nonword repetition skills in young children who do and do not stutter. *Journal of Fluency Disorders, 31,* 177-199.

Anderson, J. D., & Wagovich, S. A. (2017). *Explicit and implicit verbal response inhibition in preschool-age children who stutter. Journal of Speech, Language, and Hearing Research*, *60,* 836-852.

Archibald, L. M. D., & Gathercole, S. E. (2007a). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics, 28,* 587-606.

Archibald, L. M. D., & Gathercole, S. E. (2007b). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic Bulletin & Review, 14,* 919-924.

Archibald, L. M. D., Gathercole, S. E., & Joanisse, M. F. (2009). Multisyllabic nonwords: More than a string of syllables. *Journal of the Acoustical Society of America, 125,* 1712-1722.

Baddeley, A. D. (1986). *Working Memory.* Oxford, UK: Oxford University Press.

Baddeley, A. D. (2002). Fractionating the central executive. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 246–260). New York, NY: Oxford University Press.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of communication disorders*, *36*, 189-208.

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory.* New York: Psychology Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1974). Word length effect and the structure of short-term-memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Bakhtiar, M., Ali, D. A. A., & Sadegh, S. P. M. (2007). Nonword repetition ability of children who do and do not stutter and covert repair hypothesis. *Indian Journal of Medical Sciences, 61,* 462-470.

Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology, 44-79*.

Beaman, C. P., Neath, I., & Suprenant, A. M. (2008). Modeling distributions of immediate memory effects: No strategies needed? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 219-229.

Bloodstein, O., & Bernstein Ratner, N. (2008). *A handbook on stuttering,* 6th ed. San Diego, CA: Singular Publishing Group.

Brown, R. M. (1977). Visual and phonetic memory in preschool children: Evidence for coding flexibility. *Perceptual and Motor Skills*, *45*, 1043-1050.

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological priming in young children who stutter: Holistic versus incremental processing. American Journal of Speech-Language Pathology, 16, 43-53.

Campoy, G., & Baddeley, A. (2008). Phonological and semantic strategies in immediate serial recall. *Memory, 16,* 329-340.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of nonword repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 43,* 1-40.

Coady, J. A., Mainela-Arnold, E., & Evans, J. L. (2013). Phonological and lexical effects in verbal recall by children with specific language impairments. *International Journal of Communication Disorders, 48,* 144-159.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin,* 112, 155-159.

Collette, G., van der Linden, M., Bechet, S., & Salmon, E. (1999). Phonological loop and central executive functioning in Alzheimer's disease. *Neuropsychologia, 37,* 905-918.

Coltheart, V., & Leahy, J. (1992). Children's and adults' reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 718.

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology, 5,* 398.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31,* 668-864.

Cowan, N., Keller, T., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language, 33,* 234-250.

Cowan, N., Saults, J., Winterowd, C., & Sherk, M. (1991). Enhancement of 4-year-old children's memory span for phonologically similar and dissimilar word lists. *Journal of Experimental Child Psychology, 51,* 30-52.

Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *The Journal of Psychology, 44,* 319-327.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.

Dunn, L. & Dunn, L. (2007). *Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4).* Circle Pines, MN: American Guidance Service, Inc.

Eggers, K., De Nil, L. F., & Van den Bergh, B. R. (2013). Inhibitory control in childhood stuttering. *Journal of Fluency Disorders, 38,* 1–13.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63,* 591-601.

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *50*, 177-195.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83-94.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39,* 3–27.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513-543.

Gathercole, S. E., Adams, A. M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory and Cognition, 22,* 201-207.

Gathercole, S. E. & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Journal of Memory & Language, 28, 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29,* 336-360.

Gathercole, S. E., & Baddeley, A. D. (1995). Short-term memory may yet be deficient in children with language impairments: A comment on van der Lely & Howard (1993). *Journal of Speech, Language, and Hearing Research, 38,* 463-466.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40,* 177-190.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory, 2,* 103-127.

Goh, W. D., & Pisoni, D. B. (1998). *Effects of neighborhoods on immediate memory span for spoken words: A first report* (Research Report No. 22). Retrieved from Indiana University, Speech Research Laboratory website: http://www.indiana.edu/~srlweb/pr/22/WG-DP1.pdf.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation – Second Edition (GFTA-2).* Circle Pines, MN: American Guidance Service, Inc.

Guitar, B. (2013). *Stuttering: An integrated approach to its nature and treatment*. Lippincott Williams & Wilkins.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition, 33,* 1001-1016.

Gupta, P., & MacWhinney, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language*, *34*, 63-88.

Hakim, H., & Ratner, N. B. (2004). Nonword repetition abilities of children who stutter: an exploratory study. *Journal of Fluency Disorders, 29,* 179-199.

Hartfield, K. N., & Conture, E. G. (2006). Effects of perceptual and conceptual similarity in lexical priming of young children who stutter: Preliminary findings. *Journal of Fluency Disorders, 31,* 303-324.

Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *The Quarterly Journal of Experimental Psychology, 43,* 35-52.

Henry, L. (2012). *The development of working memory in children.* SAGE Publications, Ltd: London.

Henry, L., Messer, D. J., Luger-Klein, S. & Crane, L. (2012). Phonological, visual and semantic coding strategies in short-term picture memory span: How do they develop? *Quarterly Journal of Experimental Psychology, 65*, 2033-2053.

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Heffernan, T. M. (1991). Speech, “inner speech,” and the development of short-term memory: Effects of picture labeling on recall. *Journal of Experimental Child Psychology, 51,* 220–234.

Hollingshead A. (1975). Four-factor index of social status. Yale University, New Haven, Connecticut.

Hresko, W., Reid, D., & Hammill, D. (1999). *Test of Early Language Development – Third Edition (TELD-3).* Austin, TX: PRO-ED.

Hulme (1987). The effects of acoustic similarity on memory in children: A comparison between visual and auditory presentation. *Applied Cognitive Psychology, 1,* 45-51.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30,* 685-701.

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology, 47,* 72-87.

Jarrold, C., & Citroën, R. (2013). Reevaluating key evidence for the development of rehearsal: Phonological similarity effects in children are subject to proportional scaling artifacts. *Developmental Psychology, 49,* 837-847.

Jones, G., & Witherstone, H. L. (2011). Lexical and sub-lexical knowledge influences the encoding, storage, and articulation of nonwords. *Memory & Cognition, 39,* 588-599.

Kibby, M. Y. (2009). There are multiple contributors to the verbal short-term memory deficit in children with developmental reading disabilities. *Child Neuropsychology, 15,* 485-506.

Kibby, M. Y., Marks, W., Morgan, S., & Long, C. J. (2004). Specific impairment in developmental reading disabilities: A working memory approach. *Journal of Learning Disabilities, 37,* 349-363.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods, 44,* 978-990.

Leclercq, A-L., Maillart, C., & Majerus, S. (2013). Nonword repetition problems in children with Specific language impairment: A deficit in accessing long-term linguistic representations? *Topics in Language Disorders, 33,* 238-254.

Lee, N. R., Pennington, B. F., & Keenan, J. M. (2010). Verbal short-term memory deficits in Down syndrome: Phonological, semantic, or both? *Journal of Neurodevelopmental Disorders, 2,* 9-25.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59-80.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In: A. S. Reber, & D. L. Scarborough (Eds.). *Toward a psychology of reading: The proceedings of the CUNY conferences* (pp. 207-225). Hillsdale, NJ: Erlbaum.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory and Cognition, 24,* 305-321.

Majerus, S., Barisnikov, K., Vuillemin, I., Poncelet, M., & van der Linden (2003). An investigation of verbal short-term memory and phonological processing in four children with Williams syndrome. *Neurocase, 9,* 390-401.

Monnier, C., & Bonthoux, F. (2011). The semantic-similarity effect in children: Influence of long-term knowledge on verbal short-term memory. *British Journal of Developmental Psychology, 29,* 929-941.

Munson, B., Edwards, J., & Beckman, M.E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61–78.

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition, 27*, 45-53.

Nittrouer, S., Caldwell-Tarr, A., Lowenstein, J. H. (2003). Working memory in children with cochlear implants: Problems are in storage, not processing. *International Journal of Pediatric Otorhinolaryngology, 77,* 1886-1898.

Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. *American Journal of Speech-Language Pathology*, *20*, 163-179.

Ofoe, L., Anderson, J. D., & Ntourou, K. (2018). Short-Term Memory, Inhibition, and Attention in Developmental Stuttering: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research.*

Oyoun, H. A., El Dessouky, H., Shohdi, S., & Fawzy, A. (2010). Assessment of working memory in normal children and children who stutter. *Journal of American Science, 6,* 562-569.

Pelczarski, K. M., & Yaruss, J. S. (2016). Phonological memory in young children who stutter. *Journal of Communication Disorders, 62,* 54-66.

Pellowski, M. W., & Conture, E. G. (2002). Characteristics of speech disfluency and stuttering behaviors in 3- and 4-year-old children. *Journal of Speech, Language, and Hearing Research, 45*, 20-35.

Pellowski, M. W., & Conture, E. G. (2005). Lexical priming in picture naming of young children who do and do not stutter. *Journal of Speech, Language, and Hearing Research, 48,* 278-294.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology*, *48*, 384-404.

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory and Cognition, 43, 4*89-499.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39,* 859-862.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

Rasch, D., & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science, 46,* 175-208.

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 307–314.

Reilly, J., & Donaher, J. (2005). Verbal working memory skills of children who stutter: A preliminary investigation. *CICSD Journal, 32,* 38-42.

Riley, G. D. (2009). *Stuttering Severity Instrument for Children and Adults – Fourth Edition (SSI-4).* Austin, Tx: Pro-Ed.

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial recall: Is there an effect on all trials? *Psychonomic Bulletin & Review, 12,* 171–177.

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52A,* 367–394.

Salamé, P., & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, *24*, 263-265.

Sasisekaran, J., & Byrd, C. T. (2013). Nonword repetition and phoneme elision skills in school-age children who do and do not stutter. *International Journal of Language and Communication Disorders, 48,* 625-639.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory and Cognition*, *21*, 168–175.

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin, 75,* 399-415.

Silverman, S., & Ratner, N. B. (2002). Measuring lexical diversity in children who stutter: Application of vocd. *Journal of Fluency Disorders*, *27*, 289-304.

Smith, A., Goffman, L., Sasisekaran, J., & Weber-Fox, C. (2012). Language and motor abilities of preschool children who stutter: evidence from behavioral and kinematic indices of nonword repetition performance. *Journal of fluency disorders*, *37*, 344-358.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*, 369-373.

Storkel, H. L. & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods, 42,* 497-506.

Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology, 107,* 306-324.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2005). Redintegration and the benefits of long-term knowledge in verbal short-term memory: An evaluation of Schweickert's (1993) multinomial processing tree model. *Cognitive Psychology, 50,* 133-158.

van der Lely, H. K., & Howard, D. (1993). Children with specific language impairment: Linguistic impairment or short-term memory deficit*? Journal of Speech, Language, and Hearing Research, 36,* 1193-1207.

Wang, X., Logie, R. H., & Jarrold, C. (2016). Interpreting potential markers of storage and rehearsal: Implications for studies of verbal short-term memory and neuropsychological cases. *Memory and Cognition, 44,* 910-921.

Watts, A., Eadie, P., Block, S., Mensah, F., & Reilly, S. (2017). Language skills of children during the first 12 months after stuttering onset. *Journal of Fluency Disorders, 51*, 39-49.

Williams, K. T. (2007). *Expressive Vocabulary Test – Second Edition (EVT-2)*. Circle Pines, MN: American Guidance Service, Inc.

Wilson, M. D. (1988). The MCR Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20,* 6-11.

Yairi, E., & Seery, C. H. (2015). *Stuttering: Foundations and Clinical Applications (2nd ed)*. Columbus, OH: Allyn & Bacon.

Table 1

Number (Percent) of Children Who Stutter (CWS) and Children Who Do Not Stutter (CWNS) Who Demonstrated or Failed to Demonstrate the Phonological Similarity Effect (PSE)

Group	PSE Present	PSE Absent
CWS	14 (33.3%)	28 (66.7%)
CWNS	26 (61.9%)	16 (38.1%)
Total	40 (47.6%)	44 (52.3%)

Table 2

Number (Percent) of Children Who Stutter (CWS) and Children Who Do Not Stutter (CWNS) Who Demonstrated or Failed to Demonstrate the Semantic Similarity Effect (SSE)

Group	SSE Present	SSE Absent
CWS	23 (54.8%)	19 (45.2%)
CWNS	23 (54.8%)	19 (45.2%)
Total	46 (56.1%)	38 (45.2%)

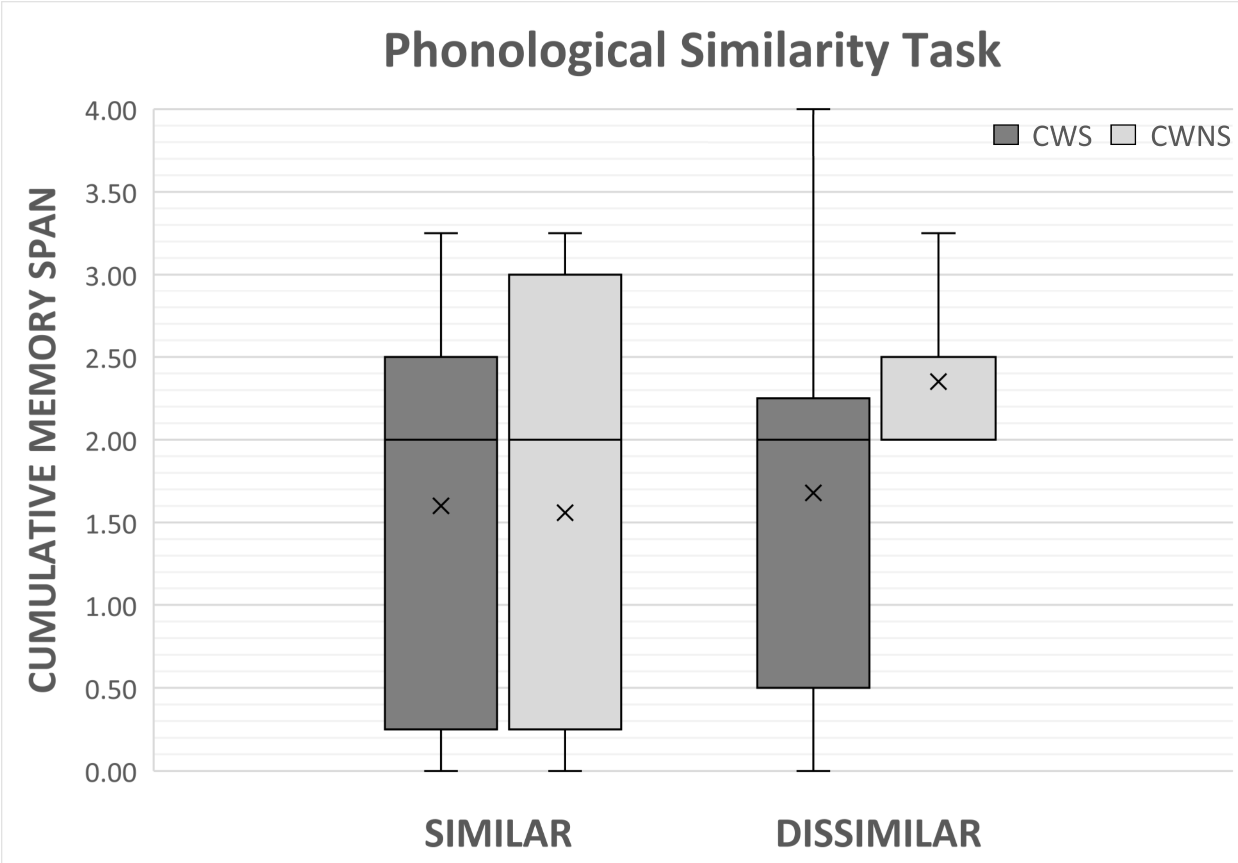

Figure 1. Box-and-whisker plot showing cumulative memory span for the similar and dissimilar conditions of the phonological similarity task for children who do (CWS) and do not stutter (CWNS). The horizontal edges of each box represent the 25th and 75th percentiles (interquartile range) and the line within each box represents the median. The whiskers represent the minimum (10th percentile) and maximum (90th percentile) values, excluding outliers. The “x” in each box represents the mean.

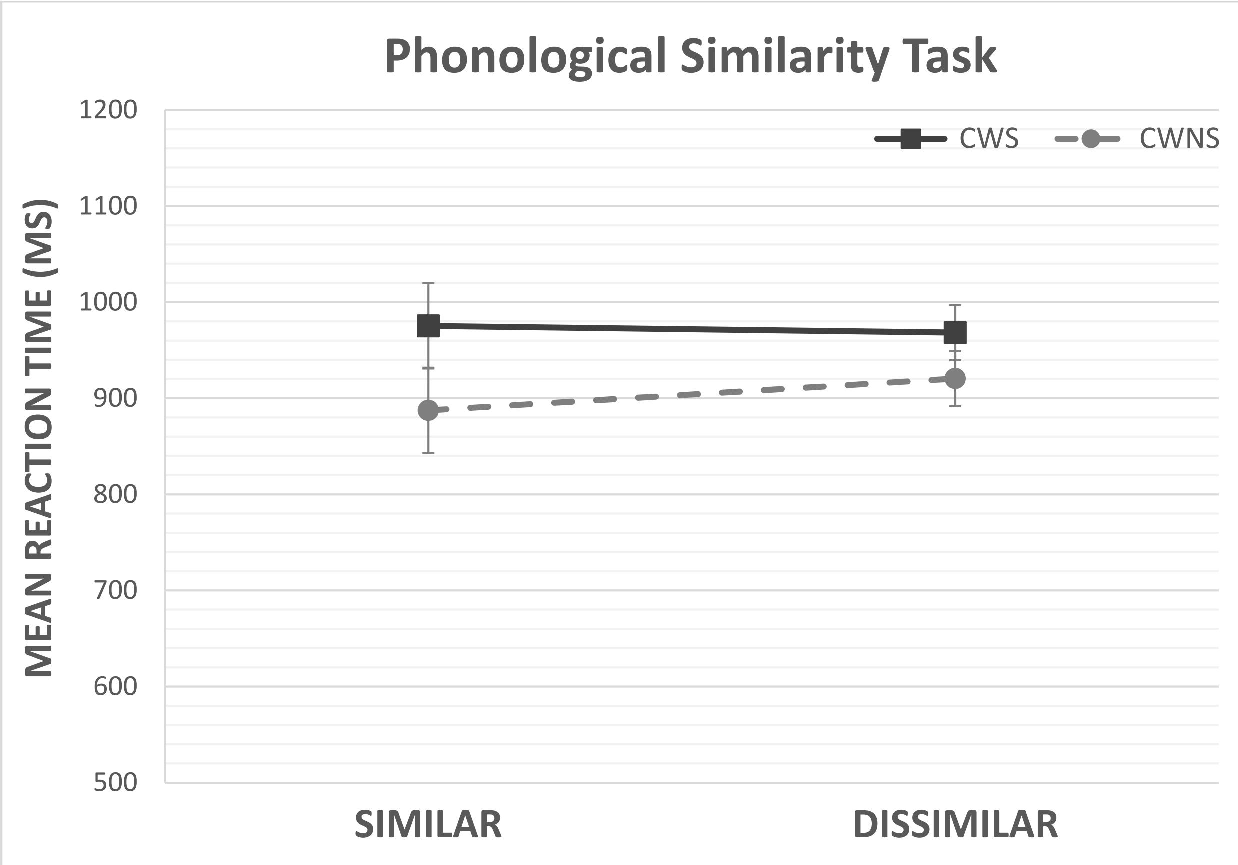

Figure 2. Adjusted mean (and standard error of the mean) reaction time (ms) for children who do (CWS) and do not stutter (CWNS) in the similar and dissimilar conditions of the phonological similarity task.

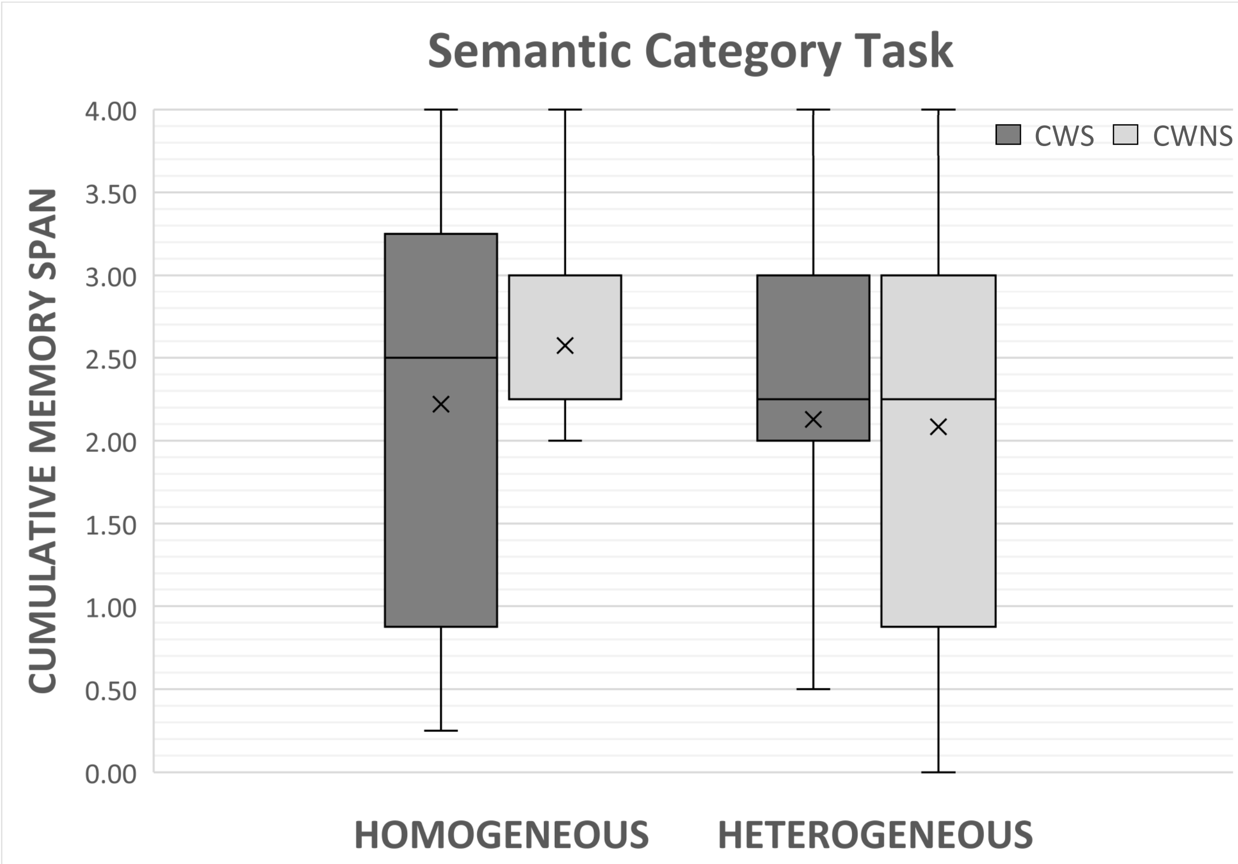

Figure 3. Box-and-whisker plot showing cumulative memory span for the homogeneous and heterogeneous conditions of the semantic category task for children who do (CWS) and do not stutter (CWNS). The horizontal edges of each box represent the 25th and 75th percentiles (interquartile range) and the line within each box represents the median. The whiskers represent the minimum (10th percentile) and maximum (90th percentile) values, excluding outliers. The “x” in each box represents the mean.

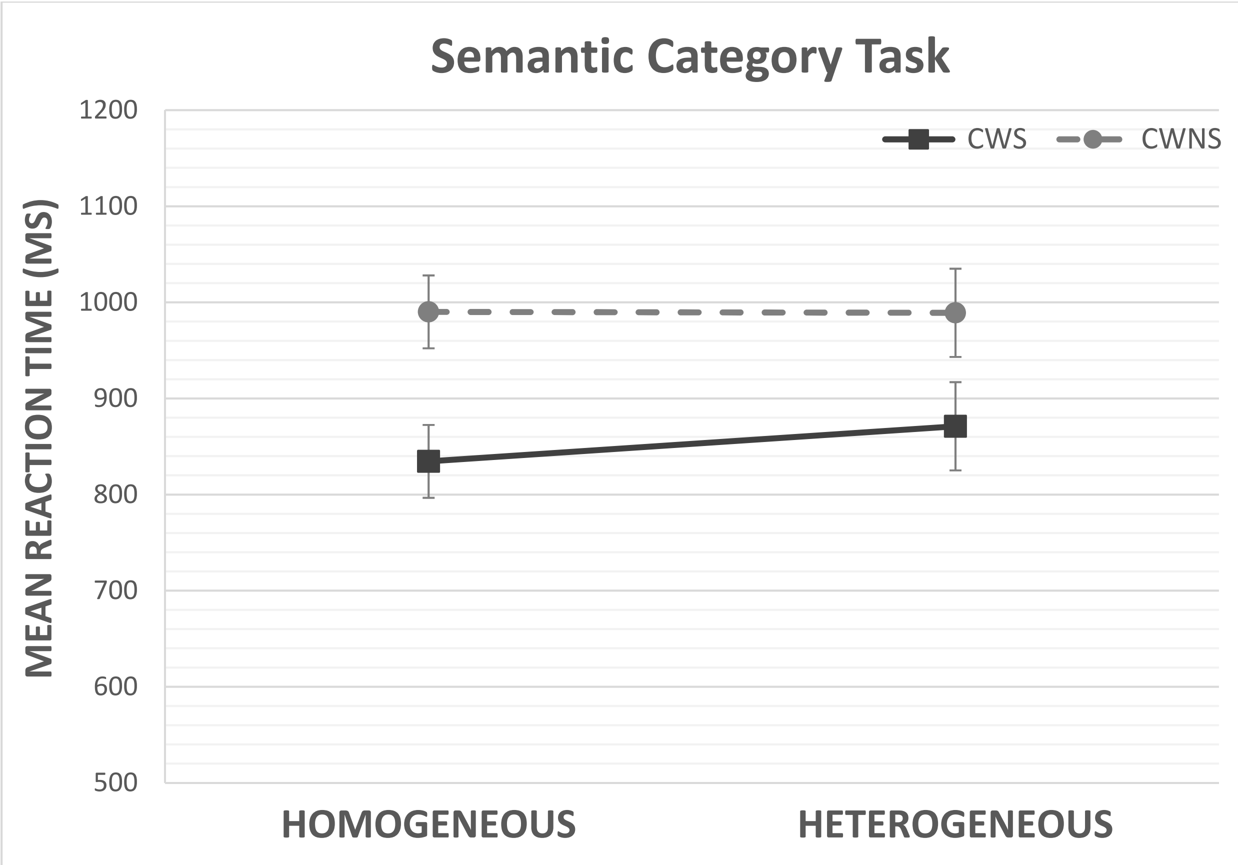

Figure 4. Adjusted mean (and standard error of the mean) reaction time (ms) for children who do (CWS) and do not stutter (CWNS) in the homogeneous and heterogeneous conditions of the semantic category task.

Appendix

Experimental Stimuli

Table A1

Experimental Stimuli for the Phonological Similarity Task in Experiment 1

Trial List	1st word	2nd word	3rd word	4th word
Phonologically Similar Condition				
1	mad	map		
2	bad	bat		
3	rag	rat	ran	
4	fat	fan	dad	
5	bag	bath	hat	can
6	tap	tag	sad	sat
Phonologically Dissimilar Condition				
1	hit	tub		
2	good	sun		
3	bell	gum	van	
4	beg	pen	mud	
5	hair	dot	bug	coat
6	kiss	pin	nut	hop

Table A2

Experimental Stimuli for the Semantic Category Task in Experiment 2

Trial List	1st word	2nd word	3rd word	4th word
Semantically Homogeneous Condition				
1	hat	shoes		
2	boy	girl		
3	spoon	cup	knife	
4	chair	couch	bed	
5	dog	cat	horse	cow
6	car	boat	truck	bus
Semantically Heterogeneous Condition				
1	shoes	couch		
2	cow	spoon		
3	car	girl	hat	
4	horse	knife	boat	
5	boy	truck	cat	bed
6	cup	dog	bus	chair