

# Verbal short-term memory deficits in Down syndrome: phonological, semantic, or both?

Nancy Raitano Lee · Bruce F. Pennington ·  
Janice M. Keenan

Received: 8 October 2008 / Accepted: 28 July 2009 / Published online: 10 September 2009  
© Springer Science + Business Media, LLC 2009

**Abstract** The current study examined the phonological and semantic contributions to the verbal short-term memory (VSTM) deficit in Down syndrome (DS) by experimentally manipulating the phonological and semantic demands of VSTM tasks. The performance of 18 individuals with DS (ages 11–25) and 18 typically developing children (ages 3–10) matched pairwise on receptive vocabulary and gender was compared on four VSTM tasks, two tapping phonological VSTM (phonological similarity, nonword discrimination) and two tapping semantic VSTM (semantic category, semantic proactive interference). Group by condition interactions were found on the two phonological VSTM tasks (suggesting less sensitivity to the phonological qualities of words in DS), but not on the two semantic VSTM tasks. These findings suggest that a phonological weakness contributes to the VSTM deficit in DS. These results are discussed in relation to the DS neuropsychological and neuroanatomical phenotype.

**Keywords** Down syndrome · Verbal short-term memory · Phonology · Semantics

Numerous studies have documented that the neuropsychological phenotype of Down syndrome (DS) is characterized by significant deficits in verbal short-term memory (VSTM; see Jarrold & Baddeley [1] for a review), with an average effect size, utilizing Cohen's *d* [2] of 1.97 (95% *CI*=1.18 to

2.75) when individuals with DS are compared to mental-age matched typically developing participants [3–10] and 0.68 (95% *CI*=0.17 to 1.20) when individuals with DS are compared to participants with other intellectual disabilities (ID) [3–9, 11–15]. Most studies of VSTM in DS have been conceptualized within the context of Baddeley's [16–18] phonological loop model of VSTM [6–9], which, as the name implies, emphasizes phonological contributions to VSTM. However, recent research has implicated the contributions of semantics to VSTM performance [19, 20]. These studies have led to an alternative model, proposed by R. Martin and colleagues [19, 21, 22], which includes both phonological and semantic-lexical subcomponents. The contribution of semantics to VSTM raises the possibility that the VSTM deficit in DS may be due not just to deficits in phonological processes (which have been implicated in some but not all studies) but also to deficits in semantic processes. Examining this possibility was one of the goals of the current research.

Understanding the nature of the VSTM deficit in DS may shed light on the larger language difficulties that characterize the syndrome. For example, research with typically developing children and children with DS has suggested that VSTM is a strong predictor of vocabulary development longitudinally [23, 24]. In typically developing children, this relationship is strongest early in development (and there is some suggestion that a developmental shift occurs in which vocabulary skills are better predictors of VSTM later in childhood). A recent experimental study has suggested that the VSTM deficit in DS is related to the ability to learn novel phonological forms and to pair these forms with an object [25] providing further evidence for the link between VSTM and vocabulary acquisition in this group. In addition to the well-replicated link between VSTM and vocabulary, VSTM has been found to be related

---

N. Raitano Lee · B. F. Pennington · J. M. Keenan  
Department of Psychology, University of Denver,  
Denver, CO, USA

N. Raitano Lee (✉)  
Unit on Brain Imaging, Child Psychiatry Branch, NIMH,  
National Institutes of Health,  
Bldg. 10, Rm. 4C110 MSC-1367, 10 Center Drive,  
Bethesda, MD 20892-1600, USA  
e-mail: lnancy@mail.nih.gov

longitudinally to receptive syntax in adolescents with DS in a 6-year longitudinal study conducted by Chapman and colleagues [26].

Lastly, research suggests that working memory, a skill that is closely related to short-term memory, but that requires not only briefly holding information online but also manipulating it to complete some task, is strongly related to intelligence [27]. A recent study [28] has shown that a brief intervention (between 8 and 19 daily sessions) aimed at improving working memory implemented with a normal adult population resulted in improvements in fluid intelligence. Thus, by refining knowledge about the nature of the VSTM deficit in DS, it is hoped that more effective interventions could be created to improve VSTM and possibly have a positive impact not only on language development but intelligence more broadly.

### Nature of the VSTM deficit in DS

Many studies have documented that the VSTM deficit in DS is specific to the *verbal* domain (e.g., [7, 14, 15]). Recent research has also suggested that this deficit appears to be due to a capacity limitation of the VSTM system and not atypically rapid decay [29]. In an attempt to isolate causes of this specific verbal deficit, several studies have focused on peripheral hearing and articulatory contributions to VSTM performance, because many individuals with DS experience hearing loss [30] and have articulatory/speech-motor deficits [31]. The approach has been to reduce or eliminate the hearing or articulatory demands of VSTM tasks and determine if DS group performance is then comparable to a matched control group. These studies have failed to find support for either current hearing difficulties [9, 11, 12, 32] or articulatory weaknesses [8, 9] accounting for this deficit.

Research has also examined the contributions of subvocal articulatory rehearsal (as described by Baddeley and colleagues [33]) to VSTM task performance in DS. While training in the use of subvocal articulatory rehearsal improves performance of individuals with DS, such interventions fail to close the gap between the DS and control groups [32, 34, 35]. Moreover, other research has found that typically developing children do not utilize subvocal articulatory rehearsal when completing VSTM tasks before the age of seven (see Gathercole [36] for a review), which is greater than the mental-age of most individuals with DS in VSTM studies and their mental-age matched controls. Somehow these young mental-age matched controls outperform the individuals with DS, despite not utilizing articulatory rehearsal. If concurrent hearing, concurrent articulation, and rehearsal deficits

cannot fully account for the VSTM deficit in DS, then what else could be contributing to this deficit? In the following sections, the contributions of phonology and semantics will be examined.

Several studies assessing the phonological contributions to VSTM performance in DS have manipulated the phonological similarity or confusability of the words included on different lists and assessed the presence of the “acoustic similarity effect” or “phonological similarity effect” (as it will be called in this paper). This effect, first described by Conrad and Hull [37], refers to the finding that words that rhyme or are similar phonologically (e.g., bat, cat, map) are more difficult to recall than phonologically dissimilar words (e.g., bat, sun, food). Theoretically, it follows that individuals who are sensitive to the phonological qualities of words will recall fewer phonologically similar than dissimilar words, while individuals who are less sensitive to these qualities may show a smaller decrement in performance (or no decrement in performance) on the phonologically similar condition.

A reduced phonological similarity effect has been found in groups with DS in some studies but not others. Varnhagen et al. [11] and Hulme and Mackenzie [6] reported a reduced effect in DS relative to controls. In contrast, studies by Broadley et al. [32], Jarrold et al. [8], and Vicari et al. [38] question a reduced phonological similarity effect in DS, either because they found a significant decrement in performance in the DS group for phonologically similar words or because of a possible floor effect. With regard to the latter issue, that of the floor effect, Vicari et al. [38] suggested that the group by condition interaction on the phonological similarity task used in their study could be accounted for by the fact that participants performed at or close to floor-level on the phonologically similar condition. Thus, there appears to be a need to modify either task demands or scoring procedures in span tasks in order to evaluate performance without the potential confound of floor effects. That was one goal of the current study.

Another paradigm utilized to study phonological contributions to VSTM in DS is nonword repetition. Cairns and Jarrold [39] examined nonword repetition of one- and two-syllable nonwords in DS and reported that the DS group was worse at repeating nonsense words than the matched control group, suggesting a deficit in phonological processing. However, given that individuals with DS have severe articulation deficits, it is difficult to separate the phonological and articulatory contributions to performance on nonword repetition tasks. Thus, nonword (and word) discrimination tasks (tasks without significant articulatory demands that involve the presentation of a string of nonwords or words followed by another string of similar or dissimilar nonwords or words) can be used. Using this

technique, Brock and Jarrold [40] reported that individuals with DS were worse at discriminating nonwords than controls overall, suggesting that deficits in phonological processing may underlie the VSTM deficit in DS.

Turning to the contributions of semantics to VSTM more generally [41, 42], such contributions have been found in adults [20] and children [43]. Concrete words are easier to recall than abstract ones in both groups. There are also patients who demonstrate specific deficits on semantic VSTM tasks but who perform similarly to control participants on phonological VSTM tasks [19, 21].

The semantic contributions to VSTM in DS have not been examined closely in child and young adults samples, possibly because receptive vocabulary has been reported to be a relative strength in DS [26], intimating preserved semantic processing. However, research with individuals with semantic STM deficits suggests that having preserved vocabulary knowledge does not preclude the possibility of having a semantic VSTM deficit [44]. To the best of our knowledge, no studies have manipulated directly the semantic qualities of words to examine the possible contributions of deficits in semantic representations to VSTM task performance in children and young adults with DS. However, research with children and young adults with DS has examined lexicality effects during VSTM tasks and reported a greater effect [40], suggesting that individuals with DS relied more on lexical knowledge (and possibly semantic representations) than controls when completing VSTM tasks. Research by Nichols et al. [45] using the California Verbal Learning Test has demonstrated that individuals with DS are more susceptible to intrusions on list recall tasks than typically developing participants. While these results can be suggestive of an over-reliance on semantics during recall tasks, this needs to be explored further utilizing a VSTM task (as opposed to a list-learning task that taps long-term memory more specifically).

Kittler and colleagues [46] examined semantic contributions to VSTM in a middle-aged sample of adults with DS and a sample of adults with idiopathic ID. They reported that individuals with DS were more sensitive to the semantic similarity of words than adults with idiopathic ID. However, they did not include a typically developing comparison group, so it is difficult to determine if the DS group demonstrated greater sensitivity to the semantic qualities of words, or if alternatively, the idiopathic ID group showed less sensitivity to the semantic qualities of words. Thus, comparing performance of participants with DS to typically developing individuals on semantic VSTM tasks was one goal of the proposed research.

Given that there is a relative dearth of studies examining semantic contributions to VSTM performance directly in young adults with DS, investigating their contributions in concert with the phonological contributions to VSTM

(utilizing scoring procedures that reduce the impact of floor effects) appear to be warranted. Therefore, the current study compared performance of individuals with DS to verbal mental-age matched typically developing children utilizing two phonological and two semantic VSTM tasks in an attempt to seek convergent evidence for the relative contributions of these processes to VSTM performance in DS.

The main hypotheses tested by the current research were as follows.

1. *Phonological hypothesis*: If a phonological deficit underlies weak VSTM skills in DS, performance on tasks tapping the phonological qualities of words will differ from MA matched controls and reflect a less mature pattern of performance, including reduced sensitivity to the phonological qualities of words and greater impairments on tasks that are phonologically demanding, such as nonword tasks.
2. *Semantic hypothesis*: If a semantic deficit underlies weak VSTM skills in DS, performance on tasks tapping the semantic qualities of words will differ from MA matched controls and reflect a less mature pattern of performance, including reduced sensitivity to the semantic qualities of words.

It should be noted that these hypotheses are *not* mutually exclusive. Both hypotheses could be supported by our results, because it is possible that deficits in both phonology and semantics may be contributing to the VSTM deficit in DS.

## Method

### Participants

Participants included 18 children and young adults with DS, ages 11–25, and 18 verbal MA matched typically developing preschool and school-age children, ages 3 to 10. All participants with DS had confirmed chromosomal diagnoses according to parent report (with one participant having mosaic DS). DS and control participants were matched pairwise on gender and receptive vocabulary consistent with other studies of VSTM in DS [8, 9].

Individuals with DS were recruited from two sources. First, individuals who participated in prior research at the University of Denver [10] were sent a letter requesting their participation once again. Twenty-four children were recruited from this source. Second, an advertisement was run in a newsletter of a local Down syndrome family support group, the Mile High Down Syndrome Association. Three individuals were recruited from this source. Of these 27 individuals, only 25 had usable data. For two partic-

ipants with DS (ages 11 and 12), testing was discontinued due to behavioral difficulties or difficulties understanding tasks. Of the 25 remaining participants, 18 individuals with DS were matched to typically developing control participants. Two of these individuals with DS were biological siblings. In order to increase our sample size, we chose to include both participants in analyses. However, all experimental analyses were re-run with one of the siblings removed and results were largely consistent. In order to qualify for the study, participants with DS were required to use at least single words to communicate. Unlike control participants (see below), hearing difficulties and significant birth complications/ medical conditions were not used as exclusionary criteria, given the high rates of these difficulties in the DS population.

Typically developing control participants were recruited through the University of Denver Developmental Psychology Participant Pool. This recruitment source includes children identified through Denver metropolitan area hospitals whose parents expressed an interest in having them participate in future research at the time of their birth. Families of these participants were contacted about the study directly by phone. Thirty-six typically developing children were recruited to be matches to the DS participants, and 18 were deemed appropriate matches. In order to qualify as a match, a participant needed to (a) reside in a monolingual-English home, (b) have no current or past concerns about speech, language, or reading difficulties, (c) pass a hearing screening completed at the University of Denver, and (d) have no history of birth complications, acquired head injury, intellectual disability or autism. S/he was also required to earn a receptive vocabulary test raw score that corresponded to an age-equivalent that was within six to seven months of a participant with DS of the same gender and a standard score between 80 and 120.

Table 1 summarizes the demographic variables for participants in the DS and control groups. Group comparisons on matching and other standardized measures were completed utilizing paired *t*-tests for continuous measures and chi-square for dichotomous measures. As can be seen, groups did not differ on primary matching measures, including gender, ethnicity (percent Caucasian), parental years of education, or the Peabody Picture Vocabulary Test-Third Edition. There was a trend, however, for paternal education to be somewhat higher in the control group than the DS group. To be conservative, all primary analyses were run with father years of education as a covariate following initial analyses and results were largely consistent.

Table 1 also summarizes performance on standardized measures of language and nonverbal intelligence. As expected, the control group significantly outperformed the DS group on the Differential Ability Scales Recall of Digits subtest. The effect size of this difference was 1.34

(Cohen's *d*) which is within the 95% confidence interval of the mean effect size from previous studies of VSTM in DS. Lastly, the DS participants performed nonsignificantly worse on the Pattern Construction subtest than the control participants.

#### Measures and procedures

Testing was completed at the University of Denver for all participants, except for one individual with DS for whom testing was completed in a quiet room of her home. For all but one individual with DS, testing took place during one, two-and-a-half hour testing session. One individual with DS (age 11) could not complete all testing during one session due to difficulties complying with task demands. Thus, testing was completed during two sessions with frequent breaks to maximize performance. For participants with typical development under the age of six, testing took place in two to three testing sessions, depending on the child's attention-level. For typically developing children over the age of six, all testing was completed in one, two-and-a-half hour session.

#### Standardized measures

*Peabody Picture Vocabulary Test — Third Edition [47]:* This is a receptive vocabulary test that requires participants to point to one of four pictures that corresponds to a vocabulary word that is spoken by the examiner.

*Differential Ability Scales Pattern Construction subtest [48]:* This is a measure of visual-spatial construction skills in which participants copy geometric designs utilizing colored blocks.

*Differential Ability Scales Recall of Digits subtest [48]:* This subtest assesses digit recall by having participants repeat increasingly long strings of digits (from two to nine digits in length) that are presented at a rate of two digits per second.

*Hearing screening:* Participants completed a hearing screening at the start of the session. All control participants passed the screening which required identification of pure tones at 25 db HL ISO for 500, 1000, 2000, and 4000 Hz for both ears (procedures outlined in [49]). Of the 18 participants with DS, only 17 were able to understand the demands of the hearing test. Ten of these 17 DS participants (58.82%) failed the hearing screening. Of these 10 participants, eight (80%) had a positive report for past hearing difficulties. Of the seven participants who passed the hearing screening, four (57.1%) had a positive history



**Table 1** Means (SDs) for demographic variables and standardized tests for DS and control groups

	DS	Control	Significance
<i>N</i>	18	18	
Age	19.18 (3.34)	5.85 (2.05)	$t(17)=16.36, p<.001$
Male (%)	27.8	27.8	$\chi^2(1)=0, p=1$
Caucasian (%)	77.8	88.9	$\chi^2(1)<1.0, p>.37$
Mother Years of Education <sup>a</sup>	15.59 (3.15)	16.19 (2.04)	$t(15)=-0.79, p>.43$
Father Years of Education <sup>a</sup>	15.81 (2.40)	16.75 (1.48)	$t(15)=-1.83, p>.08$
PPVT-III <sup>b</sup> Raw Score	89.78 (27.42)	91.00 (27.40)	$t(17)=-1.48, p>.15$
PPVT-III <sup>b</sup> Age Equivalent	6.95 (2.32)	7.03 (2.48)	$t(17)=-.85, p>.40$
DAS PC <sup>c</sup> Ability Score	105.24 (18.83)	114.71 (19.84)	$t(16)=-1.5, p>.14$
DAS PC <sup>c</sup> Age Equivalent	6.27 (2.28)	7.55 (2.63)	$t(16)=-1.63, p>.12$
DAS RD <sup>d</sup> Raw Score	10.39 (3.88)	17.28 (5.14)	$t(17)=-5.87, p<.001$
DAS RD <sup>d</sup> Age Equivalent	3.95 (1.17)	7.48 (3.83)	$t(17)=-4.14, p<.01$

<sup>a</sup> *n* in DS group for father and mother years of education is 16; one participant was missing data for mother education, one participant was missing data for father education, and one of the siblings with DS (see Method) was removed from mother and father years of education analyses

<sup>b</sup> Peabody Picture Vocabulary Test, Third Edition

<sup>c</sup> Differential Ability Scales Pattern Construction Subtest (*n*=17 in control group)

<sup>d</sup> Differential Ability Scales Recall of Digits Subtest

of past hearing difficulties. There were only three participants (17.64%) with DS who passed the hearing screening and had no history of hearing difficulties. For participants with DS, failing the hearing screening was not an exclusionary criterion, as more than 50% of individuals with DS experience some hearing loss [50, 51]. However, contributions of hearing difficulties to VSTM were explored in follow-up analyses to ensure that current hearing difficulties were not accounting for the VSTM deficits in our DS sample. This essentially involved excluding children who failed the hearing screening from analyses to determine if the results remained the same. Unfortunately, because we only utilized a brief hearing screener, hearing acuity thresholds were not obtained for study participants. Thus, we could not examine how quantitative differences in hearing acuity related to performance on the study's tasks.

### Experimental measures

Four experimental tasks were created to test the phonological and semantic hypotheses. To test the phonological hypothesis, a phonological similarity task (a traditional span task) and a nonword discrimination task were used. To test the semantic hypothesis, a semantic category task (a traditional span task) and a semantic proactive interference task were used. Task descriptions are provided below and these are followed by details about task development/administration and the order in which tasks were administered.

### Task descriptions

**Phonological similarity word recall task:** A phonological similarity task was utilized following the methods of Conrad [52] and Hulme [53]. A corpus of seven words were included in the phonologically similar (bag, cat, hat, mat, rat, map, man) and dissimilar (clock, fish, girl, hand, horse, spoon, train) conditions. These words were selected to be concrete nouns with an early age of acquisition. We used seven of the eight words from the similar and dissimilar conditions from Conrad [52], because one of the words from the phonologically similar condition, "tap," did not have a dominant concrete meaning in the American dialect. We chose to drop the word "bus" from Conrad's dissimilar list so that the similar and dissimilar lists would each have one pair of words that were from the same semantic category (i.e., "cat" and "rat" from the phonologically similar list and "fish" and "horse" from the phonologically dissimilar list). This was done to lessen semantic confounds to list recall.

Words included on the similar and dissimilar lists did not differ on ratings of age of acquisition<sup>1</sup> [54] (Similar word  $M=2.64, SD=0.49$ ; Dissimilar word  $M=2.48, SD=0.28$ ), written word frequency [55] (Similar word  $M=193.14, SD=447.47$ ; Dissimilar word  $M=130.14, SD=151.48$ ),

<sup>1</sup> Age of acquisition and imageability ratings were not available from these sources for the word 'horse'. Thus, the means for dissimilar words for these ratings were calculated using 6 of the 7 words in this condition.

concreteness [56] (Similar word  $M=591.29$ ,  $SD=39.58$ ; Dissimilar word  $M=602.57$ ,  $SD=9.47$ ), or imageability<sup>1</sup> [57] ratings (Similar word  $M=640.00$ ,  $SD=30.55$ ; Dissimilar word  $M=648.33$ ,  $SD=24.83$ ); all  $t$ -values  $<1$ , all  $p$ -values  $>.4$ .

Consistent with Conrad [52], each participant and their matched control received a unique test order. Words were selected for each list without replacement (there was replacement across lists, of course). List length increased from two to seven words with two lists of each length (i.e., two lists of two words each, two lists of three words each, and so on). The order of words on the lists was pseudorandomly generated. The three exceptions to completely random order were as follows: (a) no two consecutive lists could have identical word orders, (b) no two consecutive lists could have word orders in which the first two words in a list were identical (e.g., “cat-hat-man” and “cat-hat-map”), and (c) for the phonologically similar condition, the words on the 2- and 3-word lists needed to rhyme. This final constraint was added to increase the phonological confusability of words on shorter lists.

Task-specific procedures were as follows. Participants were instructed to listen to a series of words that were “spoken” by a cartoon character at a rate of one word per second and to say the words after s/he was done “talking,” which was signaled by a question mark on the computer screen. Prior to real test trials, a pre-test vocabulary check was completed, during which participants were required to point to pictures of the seven words from each condition presented on a computer screen. Groups were compared on errors during the vocabulary pre-check, and results revealed a nonsignificant *disadvantage* for the control group (Errors: DS  $M=0.17$ ,  $SD=0.39$ ; Control  $M=0.35$ ,  $SD=0.49$ ). A qualitative examination of errors revealed that the only vocabulary error made by children in both groups was “mat,” a word from the Similar list. (Neither group made errors during the vocabulary pre-check for the dissimilar list). Of the words in each condition, this word is likely to occur at a lower frequency than the other words; thus, it makes sense that the younger control group would tend to make more errors.

Two scoring methods were utilized for this task (and for the semantic category task described below). First, consistent with prior VSTM studies, each participant’s span was recorded (i.e., the highest consecutive list for which a participant recalled all of the words in the correct order). Second, the proportion of words recalled correctly on all lists (including lists that were beyond each participant’s span) irrespective of order was recorded for each condition. Utilizing proportion correct deviates from typical span procedures in which the task is typically discontinued once a participant’s span is reached. This alternative procedure was introduced in order to avoid the floor effects that may arise because many individuals with DS and young mental-

age match controls have memory spans that are between 2 and 4 and thus close to floor levels of performance. These two metrics differ in that the first, span, emphasizes order memory, while the second, proportion correct, emphasizes item memory.

The dependent variables for the phonological similarity task were (a) span for the similar and dissimilar conditions, and (b) the proportion of words recalled correctly (irrespective of order) for each condition, including words on lists that were beyond the participant’s identified span.

*Nonword discrimination task:* A nonword discrimination task was utilized to assess phonological contributions to the VSTM deficit in DS, similar to a task used in a previous study with individuals with DS [40]. The nonword discrimination task developed by Condouris et al. [58] required participants to determine if two nonsense words (which were spoken successively by the same speaker) were the same or different. Thirty-six nonword stimuli, with nine each of two-, three-, four-, and five-syllables in length were administered in pairs with a two-second interstimulus interval (ISI). This ISI was chosen in order to prevent participants from relying entirely on sensory memory as opposed to utilizing phonological memory as desired. Additionally, to avoid confounds associated with recall based on acoustic (as opposed to phonological) qualities of words, different versions of the same nonword stimuli were presented (rather than just presenting the same stimulus twice) when the two nonwords were the same.

Two versions of the nonword task were created so that nonwords that were the same on one list would be different on the other list to prevent possible confounds due to group-specific response biases. List version was counterbalanced across participants and both members of the DS and control pair received the same version of the task.

Nonwords were created following a modified version of Dollaghan and Campbell’s [59] criteria and adhered to English phonotactics and stress patterns. However, the nonwords were unlike any English words. Each nonword had a paired foil that differed from the target nonword by one consonant in manner, place, or voicing. In addition, the nonword foils varied such that the minimal pair change occurred in the initial, medial, or final position. For each syllable length, there were three nonwords with a minimal pair change at one of the three locations. Additionally, for the two-, three-, and five-syllable nonwords, there were either four or five nonwords that were the same (or different), and this was counterbalanced across versions. An error was made when the test versions were being modified at the University of Denver that resulted in either three or six nonwords that were the same (or different) for four-syllable words, unlike the other syllables. Because

versions of the task were counterbalanced across participants (and each matched control participant received the same version of the task as the DS participant), it is presumed that this asymmetry in the number of same and different responses for the 4-syllable nonwords would not present a systematic bias in responses in one group. Nonword stimuli are provided in Appendix 1.

Task-specific procedures were as follows. Participants were told that they would be playing “Copy Cat’s Copying Game.” In this game, a picture of a real cat with a “word bubble” appeared when the first nonword in a pair was played. This was followed by a two-second delay. Then a picture of a cartoon cat, “the copy cat,” appeared with a “word bubble” and the second nonword was played. Participants were instructed that they were going to be playing a “copy cat” game in which the real cat would say a made-up word and the “copy cat” would try to copy her. They were then instructed that “sometimes the copy cat gets it right and sometimes he gets it wrong,” and that it was their job to determine that. Participants completed training to establish understanding of the task. In this training, six words were presented in pairs in which half of the pairs were identical and half of the pairs differed by one phoneme (e.g., car, jar). Then six nonwords were presented in pairs in which half were identical. Participants received feedback if they were correct or incorrect during this training to reinforce task demands.

The dependent variables for this task were the (a) proportion of nonsense words discriminated correctly overall and (b) the proportion of nonsense words discriminated correctly at 2, 3, 4, and 5 syllables in length.

*Semantic category word recall task:* Research with adults has demonstrated that serial recall of words on lists that are from one semantic category (e.g., animals) is greater than recall of words from different semantic categories [60]. Thus, a semantic category recall task was constructed in which 14 lists of words were presented to participants in two conditions. In the homogeneous condition, monosyllabic words belonging to a particular semantic category thought to be familiar to young children (vegetables, gender, things in the sky, senses, furniture, drinks, vehicles, kitchen items, four-legged animals, clothing, colors, and body parts) were presented on lists of increasing length (two lists each from two to seven words in length). In the heterogeneous condition, the same words were utilized; however, they were presented on different lists such that no list had more than one word from a particular semantic category (e.g., “girl-cup-hat-cow-green”). The same words were utilized for the two conditions in order to control for any phonological differences or differences in age of acquisition, frequency and concreteness of words in the two conditions. Given that the same words were utilized in the two conditions, the

homogeneous and heterogeneous conditions of this task were administered at the beginning and the end of the testing session with order counterbalanced across participants. Stimuli for this task are presented in Appendix 2<sup>2</sup>.

Task specific procedures were as follows. Participants were instructed to listen to a series of words that were “spoken” by a cartoon character at a rate of one word per second and to say the words after s/he was done “talking,” which was signaled by a question mark on the computer screen.

Consistent with the Phonological Similarity task, the dependent variables for this task were (a) span for the homogeneous and heterogeneous conditions, and (b) the proportion of words recalled correctly for each condition (irrespective of order), including words on lists that were beyond the participant’s identified span. (For descriptions of how “span” and “proportion correct” were operationalized, refer to the Phonological Similarity task section.)

*Semantic proactive interference task:* A semantic proactive interference task similar to one utilized by Reutener and Fang [61] with preschool children and Reutener and Rubenstein [62] with individuals with intellectual disabilities was administered. Participants were asked to recall three successive word lists (containing three words each) belonging to the same semantic category (e.g., clothing) in order to assess semantic proactive interference. Then, a fourth list was introduced with three words from a different semantic category (e.g., body parts) to assess the release from proactive interference.

Categories and words were selected from Battig and Montague’s [63] category norms and were chosen to be familiar to preschool children: items of clothing, kitchen utensils, four-legged animals, furniture, body parts, and modes of transportation. Stimuli are provided in Appendix 3. Theoretically, if participants are encoding items based on their semantic characteristics, performance over the first three trials of words from the same semantic category should decline due to proactive interference (consistent with the results of [64]) and performance on the fourth trial with words from a different semantic category should improve and look similar to performance on the

<sup>2</sup> An examination of errors on this task revealed that on the heterogeneous condition, 13 participants (5 with DS and 8 controls) said “frown” when the word “brown” was presented on a list. The proportion of children who made this error did not differ significantly by group ( $\chi^2=1.08, p=.30$ ). Given that the task was administered via computer and all words were digitized, this error appeared to be due to a misperception of the stimuli. All participants were given credit if they said “frown” or “brown” on this list during the heterogeneous condition so as not to over-penalize for an error in the perception of the digitized stimuli.

first trial (due to the release from proactive interference). Additionally, it is expected that if participants have encoded the semantic category for a particular list, they will tend to make more errors that are from within a particular semantic category (e.g., incorrectly saying you heard “fork” when a list of other kitchen utensils is presented) than from another semantic category when recalling words after a delay. Thus, errors were also analyzed for this task and categorized into five types. *Within category errors* were defined as errors in which a child stated a word that was not a part of the particular list that they had just heard, but was from the same semantic category (e.g., saying “dog” when a list of other four-legged animals was presented). *Across category errors* were defined as errors in which the child stated a word that was from a previously presented semantic category (e.g., saying “dog” when you were completing a list containing words that were kitchen utensils). *Phonological errors* were defined as errors in which the child stated a word that differed from a target word in a list by one phoneme (e.g., saying “felt” for “belt”). *Practice errors* were defined as errors in which the child stated a word that was presented in the six practice trials during the actual trials. *Anomalous errors* included all other errors that did not fit into one of the four categories described above. Lastly, if an error was ambiguous, meaning that it was difficult to determine if it belonged to one error coding category or another (e.g., saying “cat” for “hat” — this could be coded as a phonological error or it could be coded as an across category error given that four-legged animals is a category), this error was *not* coded but instead was included in the total number of errors for that participant.

Task-specific procedures were as follows. This task was presented to the participants as the “Magic Memory Game.” Words on each list were “spoken” by a cartoon magician. Participants were asked to listen to each list and repeat the list immediately after the magician “said it.” Following immediate recall, an eight-second delay was introduced in which a distractor task was completed. During the distractor task, a screen appeared with 25 pink dots displayed at varying locations across trials. Participants were required to count out loud as many of these 25 dots as they could during the delay. Then participants were asked to recall the words the magician “said” right before they counted the dots.

Prior to completing test trials, all participants completed training, which consisted of six practice trials of lists of words from unrelated categories to familiarize participants with task demands. If participants could not recall any words after the delay or if they did not understand that they were expected to remember the words after this delay, testing was discontinued.

Dependent variables for this task included (a) the proportion of words recalled correctly following the delay

for Lists 1–3 (to assess the effects of proactive interference), (b) the proportion of words recalled during the fourth or release list (to assess the release from interference), (c) total errors made during recall, and (d) proportion of within category, across category, and anomalous errors (described above) made by participants.

*Task development and administration* Tasks were programmed utilizing Superlab Experimental Software Version 2.0.4 [65] and administered via computer to participants. Real word stimuli were recorded by the same female utilizing Goldwave [66] digital recording software, and word clarity was evaluated by a Speech-Language pathologist. For words that were identified as unclear (i.e., the digital recording did not sound like the targeted word), three versions of the targeted word were re-recorded and two adult listeners judged which digital recording sounded the most like the intended word. All words for the span tasks were administered to participants at a rate of one word per second, and participants were administered all lists for each task (i.e., including lists that were beyond their span).

*Task order* The order of semantic and phonological tasks was counterbalanced (i.e., half the subjects started with semantic tasks, and half started with phonological tasks). Additionally, condition order for each task was counterbalanced across participants to control for any possible order effects on performance. Finally, each control participant received the same test order as their matched participant with DS.

## Results

Prior to conducting primary analyses, all data were inspected for deviations from normality. The data for each group for all tasks were normally distributed, with the exception of errors on the semantic proactive interference task. Thus, data for only this task were transformed. In order to examine group differences on tasks, a series of mixed-model ANOVAs with one within-subject factor (condition) and one between-subject factor (group) were completed, followed by tests of simple effects (with *Bonferroni* correction for the number of comparisons being performed for a particular task).

### Group comparisons testing the phonological hypothesis

If the phonological hypothesis is supported, a group by condition interaction is anticipated on the phonological similarity task, such that the DS group is less affected by



the phonologically similar or confusable words than controls. Similarly, a group by syllable interaction is expected on the nonword discrimination task, such that the DS group is more impacted by the longer nonwords than controls, due to their increased phonological complexity.

*Phonological similarity word recall task*

Two, 2×2 mixed-model ANOVAs with one within-subject factor (Similar vs. Dissimilar condition) and one between-subject factor (DS vs. Control group) were completed for span and proportion of total words recalled correctly. Means (and *SDs*) for each group are summarized in Table 2. For span, there was a main effect of condition,  $F(1, 34)=19.47, p<.001$ , a main effect of group,  $F(1, 34)=15.68, p<.001$ , and a group by condition interaction,  $F(1, 34)=5.34, p<.05$ . Tests of simple effects for span (with *Bonferroni* adjustment,  $.05/4=.0125$ ) revealed that the DS group performed worse than the control group on both conditions ( $p$ 's<.01). Within-group tests of simple effects revealed that the DS group's performance did not differ significantly on the two conditions (but there was a trend for a difference in the expected direction such that similar words were harder to recall than dissimilar words;  $p<.1$ ), while the control group had a significantly smaller span for the similar than dissimilar condition ( $p<.01$ ).

For proportion of words recalled, there was a main effect of condition,  $F(1, 34)=18.03, p<.001$ , a trend for a main effect of group,  $F(1, 34)=2.79, p<.11$ , and a group by condition interaction,  $F(1, 34)=8.37, p<.01$ . Tests of simple effects (*Bonferroni* adjusted,  $.05/4=.0125$ ) revealed that for proportion of words recalled, the DS and Control groups did not differ on the similar condition ( $p>.1$ ), but the DS group performed worse than the control group on the dissimilar condition ( $p<.01$ ). Additionally, within-group tests of simple effects revealed that the DS group's performance did not differ on the proportion of words recalled between conditions ( $p>.3$ ), but the control group recalled a significantly smaller proportion of words in the similar than dissimilar condition ( $p<.001$ ). These results support the phonological hypothesis since individuals with DS were less affected by the phonological qualities of words than control participants.

Lastly, in a study utilizing a phonological similarity task analogous to the one used here, Vicari and colleagues [38] interpreted their finding of a group by condition interaction as being due to a floor-effect. We believe that such a floor effect does not account for our findings because participants did not perform near floor level when proportion of words recalled across all lists (including lists beyond a participant's span) was analyzed. As can be seen in Table 2, both the DS and control groups recalled about half of the words presented in the similar

**Table 2** Means (*SDs*), marginal mean (*SEs*) and effect sizes ( $\eta^{2p}$ ) on experimental tasks testing the phonological hypothesis

	DS Mean ( <i>SD</i> )	Control Mean ( <i>SD</i> )	Condition Marginal Mean ( <i>SE</i> )	Effect size ( $\eta^{2p}$ ) for Group, Condition & Interaction
<b>Phonological Similarity Task</b>				
Span				
Similar Condition	2.08 (0.65)	2.61 (0.58)	2.35 (0.10)	Group $\eta^{2p}=0.32$ ; Condition $\eta^{2p}=0.36$ ; Interaction $\eta^{2p}=0.14$
Dissimilar Condition	2.36 (0.68)	3.50 (1.00)	2.93 (0.14)	
Group Marginal Mean ( <i>SE</i> )	2.22 (0.15)	3.06 (0.15)		
Proportion Correct				
Similar Condition	0.51 (0.17)	0.57 (0.18)	0.54 (0.03)	Group $\eta^{2p}=0.08$ ; Condition $\eta^{2p}=0.35$ ; Interaction $\eta^{2p}=0.20$
Dissimilar Condition	0.54 (0.20)	0.69 (0.21)	0.61 (0.03)	
Group Marginal Mean ( <i>SE</i> )	0.53 (0.04)	0.63 (0.04)		
<b>Nonword Discrimination Task <sup>a, b</sup></b>				
2-syllable Proportion Correct	0.69 (0.11)	0.88 (0.13)	0.78 (0.03)	Group $\eta^{2p}=0.66$ ; Syllable $\eta^{2p}=0.13$ ; Interaction $\eta^{2p}=0.17$
3-syllable Proportion Correct	0.65 (0.14)	0.76 (0.18)	0.70 (0.04)	
4-syllable Proportion Correct	0.51 (0.15)	0.86 (0.11)	0.69 (0.03)	
5-syllable Proportion Correct	0.59 (0.13)	0.63 (0.17)		
Group Marginal Mean ( <i>SE</i> )	0.62 (0.02)	0.83 (0.02)		

<sup>a</sup>  $n=11$  in each group

<sup>b</sup> Group marginal mean and syllable effect size reported are for 2–4 syllable nonwords only, given that neither group performed above chance on the 5-syllable nonwords

condition of this task, which corresponds to approximately 27 words.

#### *Nonword discrimination task*

Prior to completing analyses on the nonword discrimination task, pairs of children (DS and their matched control participant) who had passed the training portion of the task were selected for analyses. The criterion used was that participants discriminated five of six training items correctly so that correct responses could not be accounted for by chance alone ( $p < .05$ ). This eliminated seven pairs of participants for whom either the DS ( $n=7$ ) and/or the control ( $n=3$ ) participant failed to pass the training. Thus, 11 participants from each group were included in the nonword discrimination task analyses.

Given that nonword discrimination was a forced choice task in which participants could be correct half of the time by chance alone, one-sample *t*-tests (with *Bonferroni* correction for multiple tests;  $.05/5=.01$ ) were completed to determine if the proportion correct overall and proportion correct by syllable length (two to five) differed significantly from chance (0.5) for each group. Once these analyses were completed, group comparisons of the overall proportion of words recalled correctly and proportion correct for each syllable length were completed.

When compared to chance-level performance, the proportion of nonwords discriminated correctly overall was significantly greater than chance for both the DS and control groups. Syllable-level analyses revealed that the DS group's performance differed significantly from chance for only the two- and three-syllable nonwords ( $ps < .01$ ), while the control group's performance differed significantly from chance for the two-, three-, and four-syllable nonwords ( $ps < .01$ ). Neither the DS nor the control group's performance differed significantly from chance for the five-syllable nonwords ( $ps > .03$ ), suggesting that these stimuli were too challenging to discriminate for all participants. Thus, syllable-level analyses excluded the five-syllable nonwords.

Means and *SDs* for overall proportion correct and 2–5 syllable nonwords are presented in Table 2. Data were analyzed with a  $3 \times 2$  mixed-model ANOVA with one within-subject factor (syllable: 2-, 3-, and 4- syllable nonwords) and one between-subject factor (group: DS vs. Control). Because sphericity was violated, the Greenhouse-Geisser adjusted *F*-statistic was used to evaluate the results of mixed-model ANOVA. Results revealed a main effect of group,  $F(1, 20)=39.31$ ,  $p < .001$ , such that the control group outperformed the DS group, as expected. Additionally, there was a trend for a main effect of syllable,  $F(1.55, 30.97)=3.03$ ,  $p < .08$ , such that longer syllable nonwords were harder to discriminate. Finally, there was a significant group by syllable interaction,  $F(1.55, 30.97)=4.17$ ,  $p < .05$ .

Tests of simple effects (with *Bonferroni* correction for multiple tests;  $.05/9=.006$ ) revealed that the DS group performed less well than the control group on the two- and four-syllable nonwords ( $ps < .004$ ) but not on the three-syllable nonwords ( $p < .1$ ). Additionally, an examination of performance on nonwords of differing syllable length within each group revealed no significant differences in discrimination of nonwords of differing lengths in the control group (all  $ps > .05$ ). In contrast, the DS group performed significantly worse on the four-syllable nonwords compared to two-syllable nonwords ( $p < .003$ ). Performance on the three-syllable nonwords did not differ from the two- and four-syllable nonwords in the DS group ( $ps > .10$ ). These results also support the phonological hypothesis, as children with DS were worse at discriminating nonwords overall and their performance was disproportionately affected by syllable length.

#### *Hearing difficulties and phonological task performance*

Follow-up analyses were completed to examine the possibility that a current hearing deficit contributed to the DS group's pattern of performance on the phonological VSTM tasks. This was explored by comparing only the performance of individuals with DS who passed the current hearing screening to their paired control participants. For the phonological similarity task, only 7 participants in each group were included in these analyses, as 10 of the DS participants failed the hearing screening and one participant could not be tested due to difficulties understanding the task demands. With these seven participants, the group by condition interaction on the phonological similarity task remained for both span,  $F(1,12)=5.33$ ,  $p < .05$ , and proportion of words recalled for each group,  $F(1,12)=13.95$ ,  $p < .01$ .

Similarly, proportion correct on the nonword discrimination task was analyzed with 6 pairs of participants (one DS participant who passed the hearing screening did not pass the training items for the nonword discrimination task; thus, this participant and matched control were not included). For these 6 pairs, percent correct overall continued to differ significantly between the control and DS groups,  $t(5)=-5.24$ ,  $p < .001$ . In order to be thorough, we also completed the mixed model ANOVA testing the group by syllable interaction, but the interaction did not reach statistical significance with this reduced sample size ( $F[2, 20]=1.70$ ,  $p > .39$ ). Because of the limited power to detect the interaction with this small sample, we examined mean proportion correct at the 2-, 3-, and 4-syllable levels for members of the DS groups who did ( $n=6$ ) and did not ( $n=5$ ) pass the hearing screening in order to ensure that results were similar across these two groups. Results were largely consistent, with a reduction in performance from the 2- to 4-syllable nonwords for those who did and did not

pass the hearing screening. Means (and *SDs*) for the 2-, 3-, and 4-syllable nonwords, respectively, were as follows: DS Passed Hearing Screening = 0.67 (0.10), 0.66 (0.15), 0.55 (0.13); DS Failed Hearing Screening = 0.71 (0.13), 0.64 (0.14), 0.47 (0.16).

Thus, based on these analyses for the phonological similarity and nonword discrimination tasks, it appears that current hearing difficulties alone cannot account for our phonological task findings. However, we cannot rule out the possibility that differences (including past differences) in hearing acuity between the groups did not contribute to these findings in part. Because we only screened hearing and did not utilize the lengthier procedure needed to identify hearing thresholds at various frequencies, we cannot directly assess how individual differences in hearing acuity relate to performance on these and other tasks included in this study.

#### Group comparisons testing the semantic hypothesis

If the semantic hypothesis is supported, a group by condition interaction is anticipated on the semantic category task, such that the DS group will benefit less when the words to be recalled are from the same semantic category (homogenous condition) than control participants. On the semantic proactive interference task, a group by list interaction would support the semantic hypothesis. Specifically, there would be support for this hypothesis if the DS group was less impacted by semantic proactive interference on the first three lists and showed a less significant release from interference when the semantic category was changed on the fourth list. Additionally, there would be support for the semantic hypothesis if the DS group made fewer within category than across category errors than the control group on the semantic proactive interference task.

#### *Semantic category word recall task test*

Two,  $2 \times 2$  mixed model ANOVAs were completed with one within-subject factor (Homogeneous vs. Heterogeneous condition) and one between-subject factor (DS vs. Control group) for span and the proportion of total words recalled correctly. Means (and *SDs*) are summarized in Table 3. For both span and the proportion of words recalled correctly, there were main effects of condition (Span:  $F[1, 34]=5.83$ ,  $p<.05$ ; Proportion correct:  $F[1, 34]=49.12$ ,  $p<.001$ ) and group (Span:  $F[1, 34]=12.59$ ,  $p<.01$ ; Proportion correct:  $F[1, 34]=6.42$ ,  $p<.05$ ), but there were no interactions (both  $F_s<.05$ ,  $p_s>.8$  for span and proportion correct). These results suggest that individuals with DS rely as much as controls on the semantic qualities of the words to be recalled and thus do not support the semantic hypothesis.

#### *Semantic proactive interference task*

Because the semantic proactive interference task involved the recall of words following a delay in which a verbal distractor was completed, some of the DS participants and younger control participants did not understand the demands of the task. (As stated previously, testing was discontinued for participants who did not recall any words following the delay during the six practice trials.) Thus, prior to completing analyses, pairs of children who understood the demands of the task were selected for analyses. This eliminated three pairs of participants for whom either the DS ( $n=1$ ) or control ( $n=2$ ) participant failed to understand the demands of the task. Thus, 15 participants from each group were included in these analyses.

Analyses were as follows. Performance across the three sets of semantically-related lists (items of clothing, kitchen utensils, and four-legged animals) and their respective release lists (furniture, body parts, and modes of transportation) were examined to determine if there were differences in performance across sets. For the three sets, there were no main effects for list or group and there were no group by list interactions. Thus, for parsimony, the proportion of words recalled for each list was combined across sets so that participants had four scores that represented the proportion of words they recalled for List 1, List 2, List 3, and the Release list.

These scores were submitted to a  $4 \times 2$  mixed-model ANOVA with one within-subject factor (List: 1, 2, 3, and release) and one between-subject factor (Group: DS v. Control). Consistent with analyses for individual sets, results revealed no main effect of condition ( $F<1.1$ ,  $p>.3$ ) or group ( $F<2.4$ ,  $p>.1$ ); there was also no group by condition interaction ( $F<1.6$ ,  $p>.20$ ).

Means (and *SDs*) for each group for the proportion of words recalled correctly are summarized in Table 3. As can be seen, the proportion of words recalled correctly by each group was quite low overall. Thus, it was difficult to assess proactive interference due to the fact that participants' initial recall levels were so low. This was particularly pronounced in the DS group, where recall was equal to one word or less on average.

Because proactive interference (and the release from interference) could not be examined as planned, the contributions of semantics were examined by analyzing errors made by participants during recall. The method used to classify five types of errors is summarized in the Method section; error types included within semantic category errors, across semantic category errors, anomalous errors, phonological errors, and practice errors. An analysis of error frequency revealed that phonological and practice errors occurred at a low frequency; thus, only the within

**Table 3** Means (SDs), marginal mean (SEs) and effect sizes ( $\eta^{2p}$ ) on experimental tasks testing the semantic hypothesis

	DS Mean (SD)	Control Mean (SD)	Condition Marginal Mean (SE)	Effect size ( $\eta^{2p}$ ) for Group, Condition & Interaction
<b>Semantic Category Task</b>				
Span				
Homogenous Condition	2.44 (0.70)	3.33 (0.92)	2.89 (0.14)	Group $\eta^{2p}=0.27$ ; Condition $\eta^{2p} = 0.15$ ; Interaction $\eta^{2p}=0.00$
Heterogeneous Condition	2.11 (0.78)	3.06 (1.01)	2.58 (0.15)	
Group Marginal Mean (SE)	2.28 (0.18)	3.19 (0.18)		
Proportion Correct				
Homogenous Condition	0.56 (0.16)	0.68 (0.20)	0.62 (0.03)	Group $\eta^{2p}=0.16$ ; Condition $\eta^{2p} = 0.59$ ; Interaction $\eta^{2p}=0.00$
Heterogeneous Condition	0.41 (0.12)	0.54 (0.18)	0.48 (0.03)	
Group Marginal Mean (SE)	0.48 (0.04)	0.61 (0.04)		
<b>Semantic Proactive Interference Task</b>				
Recall <sup>a</sup>				
List 1 Proportion Correct	0.27 (0.21)	0.53 (0.39)	0.40 (0.06)	Group $\eta^{2p}=0.08$ ; List $\eta^{2p} = 0.04$ ; Interaction $\eta^{2p}=0.05$
List 2 Proportion Correct	0.30 (0.19)	0.43 (0.33)	0.36 (0.05)	
List 3 Proportion Correct	0.30 (0.23)	0.41 (0.37)	0.36 (0.06)	
Release Proportion Correct	0.35 (0.30)	0.48 (0.40)	0.42 (0.07)	
Group Marginal Mean (SE)	0.31 (0.07)	0.47 (0.07)		
Errors				
Total Errors (raw number)	11.15 (7.25)	9.31 (7.89)		
Proportion Within Category <sup>b</sup>	0.47 (0.20)	0.46 (0.29)	0.46 (0.05)	Group $\eta^{2p}=0.05$ ; Error type $\eta^{2p} = 0.30$ ; Interaction $\eta^{2p}=0.00$
Proportion Across Category <sup>b</sup>	0.18 (0.17)	0.16 (0.20)	0.17 (0.04)	
Proportion Anomalous <sup>b</sup>	0.22 (0.19)	0.18 (0.21)	0.20 (0.04)	
Group Marginal Mean (SE)	0.29 (0.02)	0.27 (0.02)		

<sup>a</sup>  $n=15$  in each group

<sup>b</sup>  $n=13$  in each group. Note that this  $n$  is smaller than the total number of participants who passed the training and completed the task, because some participants did not make scoreable errors in their recall; i.e., they omitted words rather than recalling an incorrect word that could be scored as within/across category or anomalous

category, across category, and anomalous errors were included in these analyses.

Preliminary data inspection for errors revealed that the distribution of total errors for the control group was significantly kurtotic, such that the majority of children made between zero and five errors (but a few participants made as many as 20 to 30 errors). Thus, to correct for this deviation from normality, total error scores were submitted to square root transformation and normality was re-evaluated. Kurtosis decreased substantially with this transformation. *T*-tests comparing errors across group were run with both the transformed variable and the raw variable, and the results were largely consistent. Thus, non-transformed raw scores for total errors are presented in Table 3 for ease of interpretation.

For error analyses (for immediate recall and delay trials), 13 pairs of children were included. Two additional pairs of children were excluded because the control participant in that pair made no scoreable errors. With regard to the raw number of errors, the DS group made a nonsignificantly greater number of errors overall ( $t < 1$ ,  $p > .5$ ). The propor-

tion of errors by type was submitted to a  $3 \times 2$  mixed model ANOVA with one within-subject factor (error type: within vs. across vs. anomalous errors) and one between-subject factor (group: DS vs. control). Results revealed a main effect of error type,  $F(2,48)=10.42$ ,  $p < .001$ , but no effect of group,  $F < 1.14$ ,  $p > .2$ . There was also no interaction,  $F < 0.5$ ,  $p < .9$ . Tests of simple effects revealed that both groups made a greater number of within semantic category errors (errors that were from the same semantic category) than across semantic category errors (errors involving a word from a previously presented list of another semantic category), DS  $t(12)=4.83$ ,  $p < .001$ ; Control  $t(12)=2.37$ ,  $p < .04$ . These results, like those from the semantic category task, do not support the semantic hypothesis.

#### *Alternative explanations for failure to support the semantic hypothesis*

Given our small sample size, the question of insufficient power to detect interactions on the semantic tasks arises. To evaluate



this possibility, we calculated effect sizes (using Cohen's  $d$ ) for the magnitude of the semantic manipulation for the Semantic Category task (contrasting group performance on the homogenous vs. heterogeneous conditions). We then contrasted this with the effect size for the phonological manipulation on the Phonological Similarity task (contrasting group performance on the similar vs. dissimilar conditions).

While the effect size for the semantic manipulation on the Semantic Category Task for controls was small for span ( $d=0.28$ ), it was large for proportion correct ( $d = 0.74$ ). Thus, while power to detect a group by condition interaction may have been limited by a small condition effect for the span measure, this was not the case for proportion correct. Further, an evaluation of the magnitude of the semantic manipulation for the DS group for both span and proportion correct suggested that the DS group actually showed a somewhat greater effect of the semantic manipulation than controls (even though the group by condition interaction did not reach statistical significance in the ANOVA analyses summarized above). Specifically, for span, the DS group showed an effect size of 0.45, and for proportion correct, the DS group showed an effect size of 1.07. These effect sizes run counter to the semantic hypothesis, since it predicts less sensitivity to the semantic qualities of words in DS and consequently a smaller difference in performance on the homogenous and heterogeneous conditions than controls. This is the opposite of what we found. Rather, it appears that the DS group was impacted by the semantic relatedness of words *more* than controls (albeit, non-significantly).

Lastly, contrasting the effect sizes from the Semantic Category task with those obtained from the Phonological Similarity Task, we found that the effect size for the semantic manipulation for the control group was smaller than the phonological manipulation for span (phonological:  $d=1.13$ ). While this may suggest insufficient power to detect an interaction on the Semantic Category task, again the effect size for proportion correct scores runs counter to this. Specifically, the size of the phonological effect for controls for proportion correct in the similar vs. dissimilar conditions was 0.62. This is actually smaller than the size of the effect for the Semantic Category task which was 0.73. Thus, if we were able to detect an interaction on the Phonological Similarity task with a smaller effect size for condition, it seems unlikely that low power could account for the lack of an interaction on the Semantic Category task (at least for proportion correct).

## Discussion

The current study examined the phonological and semantic contributions to the VSTM deficit in DS by experimentally

manipulating phonology and semantics in VSTM tasks. Overall, the results consistently supported the phonological hypothesis and consistently did not support the semantic hypothesis. Support for the phonological hypothesis was evident from several findings. First, on a phonological similarity task, the DS group was less affected by phonologically similar words than the control group. The DS group's recall accuracy did not differ for phonologically similar and dissimilar words, while the control group's accuracy did. Moreover, for proportion of words recalled, the DS group's performance did not differ significantly from the control group's performance when the words were phonologically similar, showing (counterintuitively) that their weakness in phonological processing permitted them to perform similarly to the control group (who was penalized presumably for their greater sensitivity to the phonological qualities of words).

These results are consistent with those of Hulme and Mackenzie [6] and Varnhagen et al. [11], but inconsistent with Jarrold et al. [8] who reported similar degrees of phonological sensitivity during a phonological similarity task in which a probed memory recall procedure was used (as opposed to a serial recall procedure like the one used in the current study). This discrepancy in method may explain the inconsistency in our findings. Vicari et al. [38] found an interaction on a similar task to the one used in this study, but interpreted their results as being due to a floor effect and not to a specific phonological deficit in the DS group. As described in the results section, we do not believe that a similar floor-effect can account for our findings.

Additional support for a phonological deficit underlying the VSTM weakness in DS comes from the results of the nonword discrimination task in which individuals with DS discriminated fewer nonwords than controls overall. These results were consistent with the results of Brock and Jarrold [40] in which individuals with DS were found to discriminate fewer one-syllable nonwords than controls. Our study adds to Brock and Jarrold's findings by demonstrating that individuals with DS were not only worse at discriminating nonwords overall, but that they were also more impacted by the length of the nonwords than controls. While the control group's performance did not differ significantly when discriminating nonwords of increasing length, the DS group's performance decreased with greater syllables. Finally, the control group was able to discriminate nonwords of two- three-, and four- syllables at a rate greater than chance while the DS group was only able to discriminate two- and three- syllable nonwords at a rate greater than chance. In spite of the floor effect (chance-level performance) in the DS group in the discrimination of four-syllable nonwords, a group by syllable interaction was detected, indicating greater difficulty with nonword discrimination in the DS group when the task became more

phonologically challenging (as reflected in the greater number of syllables).

In contrast to their performance on the phonological tasks, the DS group appeared to utilize semantic processes during VSTM tasks in a manner very similar to the controls. While they performed less well than controls overall, they benefited as greatly as the control group when recalling words from the same semantic category. In fact, a comparison of the effect size for the semantic manipulation for the DS and control groups on the Semantic Category task suggested that the DS group relied (non-significantly) more on semantics than control participants, as evidenced by a larger effect of condition in the DS than control group. This clearly runs counter to the semantic hypothesis. Also counter to this hypothesis, error analyses on the semantic proactive interference task revealed a similar pattern of errors made by the DS and the control groups, with both groups showing the use of semantic processing in remembering the words because they made more errors within a semantic category than across semantic categories. The results of our semantic VSTM tasks are consistent with the greater lexicality effect reported by Brock and Jarrod [40], suggesting similar or somewhat greater reliance on semantics during VSTM tasks for participants with DS relative to control participants.

To the best of our knowledge, this is the first study to investigate systematically the semantic contributions to the VSTM deficit in DS using a child and young adult sample. Our literature review identified only one other study that examined directly the semantic contributions to VSTM task performance in DS; however, this study was completed with middle-age adults [46]. Results for this older sample revealed greater interference during recall on lists including semantically confusable words (synonyms) for a DS group than an idiopathic ID group. These findings suggested that the DS group relied on the semantic qualities of words *more* than other middle-age adults with idiopathic ID. Inconsistent with the current study's phonological findings, Kittler et al. found that adults with DS were impacted by the phonological similarity of words to a similar degree as individuals with idiopathic ID. While these findings could suggest a possible developmental shift in VSTM strategies in DS over time, such that there is increased reliance on semantics to complete VSTM tasks and possibly increased awareness of the phonological qualities of words, this interpretation must be made with caution because of the differences in the comparison groups used in the two studies (an idiopathic ID group in Kittler et al. and a typically developing younger control group in the current study). Additionally, an examination of scores on the Semantic Category task in the current study suggests a non-significant over-reliance on semantics in the DS group. While this difference did not reach statistical significance, it

does appear that individuals with DS may rely on the semantic qualities of words more than non-DS controls both in young and middle adulthood.

Thus, future research on the nature of the VSTM deficit in DS should explore the phonological and semantic contributions to VSTM tasks longitudinally, employing both a typically developing control group and a group with idiopathic ID. This would not only provide valuable information about the developmental progression of VSTM in DS, but it would also help to identify etiology-specific effects on phonological and semantic VSTM tasks in individuals with DS overtime. Because we did not include an idiopathic ID group in our study, we cannot say for certain that our pattern of findings is specific to DS and that it does not apply to individuals with ID in general.

Given that VSTM skills have been shown to contribute to vocabulary development in young children [23] and to syntactic development in individuals with DS longitudinally [26], understanding the relations between phonological development, VSTM, and language skills may be important for understanding the DS neuropsychological phenotype. In particular, it may be informative to study the development of phonological processing skills longitudinally in DS in order to identify when difficulties with phonological processing are first evident, how phonological processing changes over time, and how changes in phonological processing relate to VSTM and language functioning. In addition, future research should investigate the contributions of individual differences in hearing acuity to phonological processing and VSTM in DS. While the current research demonstrated that even participants with DS who passed the hearing screening demonstrated phonological processing deficits on VSTM tasks, this does not preclude the possibility that differences in hearing acuity in the DS group could be accounting for these phonological deficits. Thus, research should investigate the contributions of hearing acuity to phonological processing and VSTM development longitudinally in order to examine the role that commonly reported peripheral hearing difficulties play in the development of VSTM difficulties in DS.

Lastly, the neurobiological underpinnings of the VSTM and language deficits in DS should be investigated in future studies. The extant neuroimaging literature in DS is scant and includes only structural magnetic resonance imaging (MRI) studies. The majority of these studies have focused on lobar-level volumetric differences in adults with DS (see [67] for a review). The few pediatric studies have utilized structural MRI as well and have identified reductions in overall brain volume [68–70] and specific reductions in cerebellar [69, 71], frontal [71], and hippocampal [71, 72] volumes. Additionally, reductions in temporal lobar regions (including the superior temporal gyrus) have been reported

[69, 70]. Conflicting findings exist for the parietal lobes, with one study suggesting preserved volumes [69] and another suggesting reduced volumes [70] in DS.

Studies examining the neural correlates of VSTM tasks in typical populations using functional neuroimaging have highlighted the involvement of the left inferior frontal gyrus, the inferior parietal lobe, and the temporal lobes (see [73] for a review). With regard to the neural correlates of phonological and semantic VSTM tasks in particular, one study [74] highlighted involvement of different components of the frontal and temporal lobes for both phonological and semantic processes, while another study pointed to involvement of the parietal lobe (specifically the supramarginal gyrus) for phonological VSTM tasks in particular [75].

It is difficult to integrate the existing neuroimaging literature for DS with what is known about the neural correlates of VSTM, given that many of the studies of DS have utilized whole brain or lobar-level measurements to characterize the neuroanatomical phenotype of the syndrome, while functional neuroimaging studies have tended to focus on more discrete brain regions. Clearly, additional research is needed in order to identify the neural underpinnings of the VSTM deficit in DS. In particular, functional MRI studies of VSTM task performance could be informative, as they may permit a direct comparison of activation patterns during phonological and semantic VSTM tasks in children with DS and matched controls. Such research may refine our understanding of the DS neuropsychological and neuroanatomical phenotype and help to advance studies aimed at developing biomedical and educational interventions to ameliorate the cognitive deficits associated with DS.

**Acknowledgements** The present study was supported by National Institutes of Health grants, HD 17449 and HD 49027. We would like to express our sincere appreciation to the participants and families who contributed to this research. We would also like to thank Jennifer Branback, Christine Brennan, Irina Kaminer, Christa Hutaff-Lee, and Katherine Stedron for their contributions to task development, data collection, and data verification.

**Appendix 1**

Nonword Discrimination Task Stimuli [58]

Syllable Length	Nonword	Foil
2 syllable	<b>t</b> Sigav	<b>t</b> Sidav
	<b>t</b> akoub	<b>t</b> Sakoub
	<b>ma</b> ufup	<b>ma</b> ufub
	<b>kou</b> naitS	<b>kou</b> naip
	<b>d</b> zau <b>da</b> ek	<b>ga</b> u <b>da</b> ek
	<b>doi</b> baef	<b>doi</b> vaef

3 syllable	<b>no</b> ipa <b>ed</b> z	<b>mo</b> ipa <b>ed</b> z
	<b>da</b> u <b>ke</b> i <b>g</b>	<b>da</b> u <b>ke</b> i <b>b</b>
	<b>ke</b> it <b>Sou</b> n	<b>ke</b> ipou <b>n</b>
4 syllable	<b>vo</b> ut <b>Se</b> ima <b>ef</b>	<b>vo</b> u <b>ke</b> ima <b>ef</b>
	<b>na</b> ib <b>ado</b> it	<b>na</b> ib <b>ado</b> ip
	<b>di</b> ma <b>no</b> iv	<b>d</b> z <b>ai</b> ma <b>no</b> iv
	<b>na</b> et <b>Sida</b> g	<b>ma</b> et <b>Sida</b> g
	<b>mo</b> itau <b>vup</b>	<b>mo</b> idau <b>vup</b>
	<b>fa</b> teip <b>oid</b>	<b>fa</b> teip <b>oig</b>
	<b>yu</b> noit <b>Sig</b>	<b>fu</b> noit <b>Sig</b>
	<b>ba</b> give <b>ip</b>	<b>ba</b> dive <b>ip</b>
	<b>d</b> z <b>ai</b> ma <b>bo</b> it	<b>d</b> z <b>ai</b> ma <b>bo</b> if
	<b>ka</b> po <b>it</b> ae <b>vou</b> n	<b>ka</b> po <b>it</b> ae <b>vou</b> m
5 syllable	<b>ga</b> ud <b>ite</b> ka <b>dz</b>	<b>ga</b> ud <b>it</b> Se <b>ik</b> adz
	<b>d</b> z <b>ae</b> ta <b>upou</b> fo <b>ig</b>	<b>d</b> z <b>ae</b> ta <b>upou</b> fo <b>ig</b>
	<b>t</b> Se <b>imo</b> ig <b>afau</b> b	<b>t</b> Se <b>imoid</b> z <b>afau</b> b
	<b>pa</b> ed <b>zau</b> vit <b>So</b> im	<b>pa</b> ed <b>zau</b> vit <b>So</b> in
	<b>fu</b> ka <b>na</b> iteit <b>S</b>	<b>yu</b> ka <b>na</b> iteit <b>S</b>
	<b>po</b> uf <b>ag</b> ae <b>ma</b> ut	<b>po</b> uf <b>ada</b> e <b>ma</b> ut
	<b>fa</b> ud <b>oi</b> kou <b>nae</b> dz	<b>fa</b> ud <b>oi</b> kou <b>nae</b> t <b>S</b>
	<b>t</b> Sout <b>anoid</b> ib	<b>k</b> ou <b>ta</b> noid <b>ib</b>
	<b>po</b> iv <b>ifa</b> baed <b>zau</b> t	<b>po</b> iv <b>ifa</b> baed <b>zau</b> k
	<b>fo</b> inae <b>git</b> at <b>Se</b> ik	<b>fo</b> inae <b>gid</b> at <b>Se</b> ik
	<b>ba</b> ed <b>zau</b> vug <b>ifam</b>	<b>pa</b> ed <b>zau</b> vug <b>ifam</b>
	<b>vo</b> ig <b>au</b> baet <b>Sidap</b>	<b>fo</b> ig <b>au</b> baet <b>Sidap</b>
	<b>ta</b> u <b>fa</b> mo <b>ig</b> ak <b>ain</b>	<b>ta</b> u <b>fa</b> mo <b>ib</b> ak <b>ain</b>
	<b>ga</b> po <b>u</b> ke <b>id</b> z <b>ae</b> vid	<b>ga</b> po <b>u</b> ke <b>id</b> z <b>ae</b> vit
	<b>gi</b> d <b>z</b> ae <b>mau</b> vut <b>ab</b>	<b>gi</b> d <b>ae</b> mau <b>vut</b> ab
<b>ma</b> e <b>ve</b> ino <b>id</b> z <b>ai</b> pou <b>d</b>	<b>ma</b> e <b>ve</b> ino <b>id</b> z <b>ai</b> pou <b>b</b>	
<b>te</b> ivou <b>fub</b> amo <b>ig</b>	<b>t</b> Se <b>ivou</b> f <b>ub</b> amo <b>ig</b>	

Syllables in bold are stressed.

Pronunciation Key: tS=cheetah; dz=judge; ai=knife; ae=rack; ei=cave; au=cow

**Appendix 2**

Semantic Category Task Stimuli

Homogenous Condition					
corn	peas				
boy	girl				
sun	moon	star			
see	smell	hear			
chair	lamp	desk	couch		
milk	juice	tea	punch		
bus	train	car	boat	truck	
spoon	dish	fork	cup	knife	
dog	cat	horse	cow	wolf	goat
gloves	dress	shorts	hat	shoes	scarf

red	blue	green	white	black	brown	pink
leg	hand	arm	foot	head	eye	ear

#### Heterogeneous Condition

scarf	ear					
shoes	brown					
hear	pink	goat				
truck	head	black				
boat	knife	wolf	eye			
milk	couch	horse	spoon			
girl	cup	hat	cow	green		
star	desk	arm	juice	car		
corn	boy	sun	white	shorts	foot	
leg	blue	cat	gloves	chair	see	
peas	moon	smell	lamp	punch	train	dish
dress	red	hand	dog	fork	bus	tea

### Appendix 3

#### Semantic Proactive Interference Task Stimuli

##### Set 1

List 1	socks	dress	scarf
List 2	pants	gloves	belt
List 3	hat	shorts	coat
Release List	chair	bed	lamp

##### Set 2

List 1	pan	fork	dish
List 2	spoon	glass	stove
List 3	knife	plate	cup
Release List	legs	arms	head

##### Set 3

List 1	dog	bear	goat
List 2	cat	pig	sheep
List 3	horse	cow	wolf
Release List	car	bus	train

### References

- Jarrold C, Baddeley AD. Short-term memory in Down syndrome: applying the working memory model. *Downs Syndr Res Pract.* 2001;7(1):17–23.
- Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
- MacKenzie S, Hulme C. Memory span development in Down's syndrome severely subnormal and normal subjects. *Cogn Neuro-psych.* 1987;4(3):303–19.
- Marcell MM, Weeks SL. Short-term memory difficulties and Down's syndrome. *J Ment Defic Res.* 1988;32(2):153–62.
- Marcell MM, Harvey CF, Cothran LP. An attempt to improve auditory short-term memory in Down's syndrome individuals through reducing distractions. *Res Dev Disabil.* 1988;9(4):405–17.
- Hulme C, MacKenzie S. Working memory and severe learning difficulties. East Sussex, UK: Erlbaum; 1992.
- Jarrold C, Baddeley AD. Short-term memory for verbal and visuospatial information in Down's syndrome. *Cogn Neuropsychiatr.* 1997;2(2):101–22.
- Jarrold C, Baddeley AD, Hewes AK. Verbal short-term memory deficits in Down syndrome: a consequence of problems in rehearsal? *J Child Psychol Psychiatry.* 2000;41(2):233–44.
- Jarrold C, Baddeley AD, Phillips CE. Verbal short-term memory in Down syndrome: a problem of memory, audition, or speech? *J Speech Lang Hear Res.* 2002;45(3):531–44.
- Pennington BF, Moon J, Edgin J, Stedron J, Nadel L. The neuropsychology of Down syndrome: Evidence for hippocampal dysfunction. *Child Dev.* 2003;74(1):75–93.
- Varnhagen CK, Das JP, Varnhagen S. Auditory and visual memory span: Cognitive processing by TMR individuals with Down syndrome or other etiologies. *Am J of Ment Defic.* 1987;91(4):398–405.
- Marcell MM, Armstrong V. Auditory and visual sequential memory of Down syndrome and nonretarded children. *Am J Ment Def.* 1982;87(1):86–95.
- Bower A, Hayes A. Short-term memory deficits and Down's syndrome: A comparative study. *Downs Syndr Res Pract.* 1994;2(2):47–50.
- Wang PP, Bellugi U. Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *J Clin Exp Neuropsych.* 1994;16(2):317–22.
- Edgin J. A neuropsychological model for the development of the cognitive profiles in mental retardation syndromes: Evidence from Down syndrome and Williams syndrome [doctoral dissertation]. Denver, CO: University of Denver; 2003.
- Baddeley A. The concept of working memory. In: Gathercole SE, editor. *Models of short-term memory.* East Sussex, UK: Erlbaum; 1996. p. 1–27.
- Baddeley A. Short-term and working memory. In: Tulving E, Craik FIM, editors. *The Oxford handbook of memory.* London: Oxford University Press; 2000. p. 77–92.
- Baddeley AD. Working memory. In: Bower G, editor. *The psychology of learning and motivation.* New York: Academic; 1974. p. 47–89.
- Martin RC, Shelton JR, Yaffee LS. Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *J Mem Lang.* 1994;33(1):83–111.
- Walker I, Hulme C. Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *J Exp Psychol Learn.* 1999;25(5):1256–71.
- Freedman ML, Martin RC. Dissociable components of short-term memory and their relation to long-term learning. *Cogn Neuropsychol.* 2001;18(3):193–226.
- Martin RC, Freedman ML. Short-term retention of lexical-semantic representations: Implications for speech production. *Memory.* 2001;9(4):261–80.
- Gathercole SE, Baddeley AD. The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *Brit J Psychol.* 1990;81(4):439–54.
- Laws G, Gunn D. Phonological memory as a predictor of language comprehension in Down syndrome: a five-year follow-up study. *J Child Psychol Psych.* 2004;45(2):326–37.
- Jarrold C, Thorn AS, Stephens E. The relationships among verbal short-term memory, phonological awareness, and new word learning: evidence from typical development and Down syndrome. *J Exp Child Psychol.* 2009;102(2):196–218.



26. Chapman RS, Hesketh LJ, Kistler DJ. Predicting longitudinal change in language production and comprehension in individuals with Down syndrome: Hierarchical linear modeling. *J Speech Lang Hear R.* 2002;45(5):902–15.
27. Kyllonen P. Is working memory capacity Spearman's  $g$ . In: Dennis I, Tapsfield P, editors. *Human abilities: Their nature and measurement.* Mahwah, NJ: Lawrence Erlbaum Associates; 1996.
28. Jaeggi SM, Buschkuhl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. *Proc Natl Acad Sci U S A.* 2008;105(19):6829–33.
29. Purser HRM, Jarrold C. Impaired verbal short-term memory in Down syndrome reflects a capacity limitation rather than atypically rapid forgetting. *J Exp Child Psychol.* 2005;91(1):1–23.
30. Marcell MM, Cohen S. Hearing abilities of Down syndrome and other mentally handicapped adolescents. *Res Dev Disabil.* 1992;13(6):533–51.
31. Dodd B. Recognition and reproduction of words by Down's syndrome and non-Down's syndrome retarded children. *Am J of Ment Defic.* 1975;80(3):306–11.
32. Broadley I, MacDonald J, Buckley S. Working memory in children with Down syndrome. *Downs Syndr Res Pract.* 1995;3:3–8.
33. Baddeley AD, Lewis V, Vallar G. Exploring the articulatory loop. *Q J Exp Psychol A.* 1984;2:233–52.
34. Comblain A. Working memory in Down's syndrome: Training the rehearsal strategy. *Down Syndrome: Research & Practice* 1994;2(3):123–6
35. Laws G, MacDonald J, Buckley S. The effects of a short training in the use of a rehearsal strategy on memory for words and pictures in children with Down syndrome. *Downs Syndr Res Pract.* 1996;4(2):70–8.
36. Gathercole SE. The development of memory. *J Child Psychol and Psych.* 1998;39(1):3–27.
37. Conrad R, Hull AJ. Information, acoustic confusion and memory span. *Br J Psychol.* 1964;55(4):429–32.
38. Vicari S, Marotta L, Carlesimo GA. Verbal short-term memory in Down's syndrome: An articulatory loop deficit? *J Intell Disabil Res.* 2004;48(2):80–92.
39. Cairns P, Jarrold C. Exploring the correlates of impaired non-word repetition in Down syndrome. *Br J of Dev Psych.* 2005;23(3):401–16.
40. Brock J, Jarrold C. Language influences on verbal short-term memory performance in Down syndrome: item and order recognition. *J Speech Lang Hear Res.* 2004;47(6):1334–46.
41. Kintsch W, Buschke H. Homophones and synonyms in short-term memory. *J Exp Psychol.* 1969;80(3):403–7.
42. Levy BA, Baddeley A. Recall of semantic clusters in primary memory. *Q J Exp Psychol — A.* 1971;23(1):8–13.
43. Nation K, Adams JW, Bowyer-Crane CA, Snowling MJ. Working memory deficits in poor comprehenders reflect underlying language impairments. *J Exp Child Psychol.* 1999;73(2):139–58.
44. Martin RC, He T. Semantic short-term memory and its role in sentence processing: a replication. *Brain Lang.* 2004;89(1):76–82.
45. Nichols S, Jones W, Roman MJ, Wulfeck B, Delis DC, Reilly J, et al. Mechanisms of verbal memory impairment in four neurodevelopmental disorders. *Brain Lang.* 2004;88(2):180–9.
46. Kittler P, Krinsky-McHale SJ, Devenny DA. Semantic and phonological loop effects on verbal working memory in middle-age adults with mental retardation. *Am J Ment Retard.* 2004;109(6):467–80.
47. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test.* 3rd ed. Circle Pines, MN: American Guidance Service; 1997.
48. Elliott CD. *The Differential Ability Scales.* San Antonio, TX: The Psychological Corporation; 1990.
49. ASHA. Guidelines for screening for hearing impairment and middle-ear disorders. *Asha.* 1990;32(Suppl. 2):17–24.
50. Cunningham CC, McArthur K. Hearing loss and treatment in young Down's syndrome children. *Child Care, Health Dev.* 1981;7(6):357–74.
51. Dahle AJ, McCollister FP. Hearing and otologic disorders in children with Down syndrome. *Am J Ment Defic.* 1986;90(6):636–42.
52. Conrad R. The chronology of the development of covert speech in children. *Dev Psychol.* 1971;5(3):398–405.
53. Hulme C. Developmental differences in the effects of acoustic similarity on memory span. *Dev Psychol.* 1984;20(4):650–2.
54. Cortese MJ, Khanna MM. Age of acquisition ratings for 3, 000 monosyllabic words. *Behavior Research Methods.* 2008;40:791–4.
55. Kucera H, Francis WN. *Computational Analysis of Present-Day American English.* Providence: Brown University Press; 1967.
56. Coltheart M. The MRC Psycholinguistic Database. *Q J Exp Psychol.* 1981;33A:497–505.
57. Cortese MJ, Fugett A. Imageability ratings for 3, 000 monosyllabic words. *Behav Res Methods Instrum Comput.* 2004;36(3):384–7.
58. Condouris K, Bermis R, Evancie L, McGrath L, Connolly C, Tager-Flusberg H. Discrimination and repetition of nonwords in autism and specific language impairment. Submitted.
59. Dollaghan C, Campbell TF. Nonword repetition and child language impairment. *J of Speech Lang Hear Res.* 1998;41(5):1136–46.
60. Poirier M, Saint-Aubin J. Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Q J Exp Psychol - A.* 1995;2:384–404.
61. Reutener DB, Fang J. Encoding processes and release from proactive interference in short-term memory of preschool children. *J Gen Psychol.* 1985;112(4):343–8.
62. Reutener DB, Rubenstein C. Release from proactive interference in short-term memory in mentally retarded persons. *Am J of Ment Defic.* 1976;81(1):102–5.
63. Battig WF, Montague WE. Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *J Exp Psychol Monograph Suppl.* 1969;80(3):1–46.
64. Peterson L, Peterson MJ. Short-term retention of individual verbal items. *J Exp Psychol.* 1959;58(3):193–8.
65. Cedrus. *SuperLabPro 2.0.4:* Cedrus Corporation; 2003.
66. Goldwave. *Goldwave 5.10:* Author; 2005.
67. Nadel L. Down's syndrome: a genetic disorder in biobehavioral perspective. *Genes Brain Behav.* 2003;2(3):156–66.
68. Jernigan TL, Bellugi U. Anomalous brain morphology on magnetic resonance images in Williams syndrome and Down syndrome. *Arch Neurol.* 1990;47(5):529–33.
69. Pinter JD, Eliez S, Schmitt JE, Capone GT, Reiss AL. Neuroanatomy of Down's syndrome: A high-resolution MRI study. *Am J Psychiatry.* 2001;158(10):1659–65.
70. Kates WR, Folley BS, Lanham DC, Capone GT, Kaufmann WE. Cerebral growth in Fragile X syndrome: review and comparison with Down syndrome. *Microsc Res Tech.* 2002;57(3):159–67.
71. Jernigan TL, Bellugi U, Sowell E, Doherty S, Hesselink JR. Cerebral morphologic distinctions between Williams and Down syndromes. *Arch Neurol.* 1993;50(2):186–91.
72. Pinter JD, Brown WE, Eliez S, Schmitt JE, Capone GT, Reiss AL. Amygdala and hippocampal volumes in children with Down syndrome: A high-resolution MRI study. *Neurology.* 2001;56(7):972–4.
73. Martin RC. Components of short-term memory and their relation to language processing: Evidence from neuropsychology and neuroimaging. *Curr Dir Psychol Sci.* 2005;14:204–8.
74. Crosson B, Rao SM, Woodley SJ, Rosen AC, Bobholz JA, Mayer A, et al. Mapping of semantic, phonological, and orthographic verbal working memory in normal adults with functional magnetic resonance imaging. *Neuropsychology.* 1999;13(2):171–87.
75. Martin RC, Wu DF M, Jackson EF, Lesch M. An event-related fMRI investigation of phonological versus semantic short-term memory. *J Neurolinguist.* 2004;16:341–60.