

8-14-2017

Influences on Nonword Repetition in Young Children

Dana Arthur

University of Connecticut - Storrs, arthur@newpaltz.edu

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Arthur, Dana, "Influences on Nonword Repetition in Young Children" (2017). *Doctoral Dissertations*. 1603.
<https://opencommons.uconn.edu/dissertations/1603>

Influences on Nonword Repetition in Young Children

Dana Thorne Arthur, PhD

University of Connecticut, 2017

Abstract

Nonword repetition (NWR) tasks have become popular in recent years as measures of phonological short term memory (PSTM) in research settings, and as potential markers of language impairment in clinical settings. This has increased interest in the subskills that potentially NWR performance. The current study investigated the influence of PSTM, vocabulary skills, articulatory output skills, literacy, and speech perception on NWR in children ages 4;0-6;5. Partial correlations and regression analysis were used to establish an in-depth picture of the skills that constrain NWR, and to differentiate the skills that underlie this measure when the language skills of test takers and the wordlikeness of NWR items were taken into account. Results indicated that both PSTM and vocabulary were independently related to NWR performance when the entire sample was considered. When the sample was divided by language scores, higher-scoring participants' NWR performance was influenced by PSTM, while only vocabulary scores were related to NWR performance in lower-language scoring children. Wordlikeness of nonword items also affected the skills related to NWR performance, in that scores on highly wordlike stimuli were principally related to vocabulary skills, while scores on less wordlike stimuli were significantly related to PSTM only. These results indicate that despite its popularity as a measure of PSTM, NWR cannot be considered a monolithic measure of this skill in all cases. In general, NWR performance in the current study were influenced by both PSTM and vocabulary skills, and that influence varied with the language skill of participants and the wordlikeness of NWR items.

Influences on Nonword Repetition in Young Children

Dana Thorne Arthur

B.A., Elms College, 2001

M.S. Boston University, 2003

Ed.M. Harvard University Graduate School of Education, 2004

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Influences on Nonword Repetition in Young Children

Presented by

Dana Thorne Arthur, B.A., M.S., EdM

Major Advisor _____
Bernard Grela

Associate Advisor _____
Tammie Spaulding

Associate Advisor _____
Jay Rueckl

University of Connecticut
2017

Acknowledgements

I could not have completed this journey without the untiring support of my family, friends, colleagues, and mentors, for whom I will be forever grateful. I would first like to thank my committee, Dr. Bernard Grela, Dr. Jay Rueckl, and Dr. Tammie Spaulding, for their guidance throughout the dissertation process. Their unique contributions helped me grow as a researcher and showed me clearly the kind of mentor that I wish to one day become. I also must thank my additional readers and supporters, Dr. Rachel Theodore, Dr. Emily Myers, and Dr. Ken Pugh, for taking time away from their busy schedules to lend me their expertise and help.

I would also like to acknowledge the tireless grad assistants in the Statistical Consulting Services program at UConn, for their help in planning my statistical approach. In addition, I am especially grateful to the participating schools, child care centers, and families who made this project possible.

This work could not have been carried out without the help of my excellent student research teams, both at UConn and SUNY New Paltz. Thanks to Andrew Dittus, Anna Grace Gerber, Ryan Buswell, Elissa Terach, Kayla Villante, James Paladino, Manuel Marte, Marcos Velasco, Amanda Starr, Maggie Bergin, Lachana Flanders, Michaela Caporaso, and Amanda Chambers for their assistance collecting and processing data, and their unending enthusiasm for even the most tedious parts of the research process.

I am more grateful than I can express for the support, insight, and friendship of my lab mates and colleagues at UConn and SUNY New Paltz, who both inspired me to continue my work and hang on to my sanity as I did so. I would particularly like to thank Beverly Collisson, F. Sayako Earle, Vanessa Harwood, Jen Mozeiko, Karen Aicher, Alexis Johns, and Ashlee Shaw Williams from UConn, and Wendy Bower, Andrea Abramovitch, Inge Anema, Jessica Welsh, and Stella Turk at SUNY.

Most importantly, I would like to thank my family for their love, support, and inspiration through this long and arduous journey. I would not be here without them.

Chapter I: Literature Review

1.1 The promise and challenge of nonword repetition as an assessment tool	1
1.2 Nonword Repetition and Language Impairment	2
1.3 Potential influences on Nonword repetition performance	
Phonological short term memory	5
Vocabulary	8
Literacy	12
Articulation	14
Speech Perception	17
Test-taker characteristics	19
Wordlikeness	21
1.4 Purpose and Hypotheses	24

Chapter 2: Methods

2.1 Participants	26
2.2 Testing Battery & Procedures	28
Nonword repetition testing	29
Assessment of potential influences	31
Assessment of participant language skills	34
Assessment of nonword characteristics	35
2.3 Data Analysis	36

Chapter 3: Results

3.1 Sample characteristics & variable transformation	38
3.2 Analysis 1: Significant Influences on NWR performance	42
3.3 Analysis 2: Differential Influences on NWR due to language skill	46
3.4 Analysis 3: Differences in NWR due to nonword characteristics	49

Chapter 4: Discussion

4.1 Summary	52
4.2 Research & Clinical Implications	58
4.3 Limitations and Future Directions	61
4.4 General Conclusion	62

Appendix A: Nonword Items	64
---------------------------	----

References	66
------------	----

Chapter I: Literature Review

1.1 The promise and challenge of nonword repetition as an assessment tool

Nonword repetition (NWR) is an evaluation task commonly used in language, reading, and cognitive evaluations. Tasks of this type require participants to repeat aurally presented, phonotactically legal nonwords of increasing length (e.g. meb, ballope, woogalamic). The ability to repeat novel phonological forms is strongly related to several areas of language development, such as word learning (e.g. Gathercole, Hitch, & Martin, 1997), and second language acquisition (Masoura & Gathercole, 1999).

Because this measure uses pseudo-word forms rather than real words, NWR was once thought to provide a relatively pure assessment of phonological memory, independent of lexical knowledge (Gathercole & Baddeley 1990a, b). However, it has since been observed that NWR performance does not depend solely on memory capacity (Archibald & Gathercole, 2006; Graf Estes, Evans, & Else-Quest, 2007). Several additional factors may influence an individual's NWR performance, including vocabulary, auditory processing, articulatory output processes, and literacy development. Further, the degree to which each of these factors influences NWR scores may vary in different populations, and may also be mediated by features of the nonwords themselves.

In recent years, research efforts have utilized NWR scores as solely a measure of phonological memory, in order to examine various aspects of learning in children (e.g. Bishop, Adams, & Norbury, 2005; Gathercole, Hitch, Service, & Martin, 1997; Masoura & Gathercole, 1999). If NWR performance is actually simultaneously influenced by multiple other skills and constructs in addition to phonological short term memory (PSTM), validity concerns may arise.

Any conclusions drawn from the use of NWR as a simple representation of PSTM, could be confounded by whichever other skills contribute significantly to NWR in the tested population. While it may be argued that no measure is pure, and therefore such validity issues are always present, researchers must have sufficient information about NWR as a measure in order to design precise protocols and minimize threats to validity.

Clinically, NWR has become popular in recent years as a diagnostic tool. NWR deficits have been observed in several disordered populations, including children with low reading ability, and neuropsychological patients showing acquired disorders of language processing (Gathercole, Willis, Baddeley, and Emslie, 2004). However, NWR has garnered particular attention as a potential marker of language impairment in children.

1.2 Nonword Repetition and Language Impairment

Deficits in NWR skill have been noted in several disordered populations, including individuals with autism (e.g. Riches, Loucas, Baird, Charman, & Simonoff, 2011), children with cochlear implants (Dillon, Cleary, Pisoni, & Carter, 2004), and children with reading problems (e.g. Gathercole, Willis, Emslie & Baddeley, 1991). Importantly, NWR deficits are also consistently associated with specific language impairment (SLI; e.g. Archibald & Gathercole, 2006; Graf-Estes, Evans, & Else-Quest, 2007; Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, & Jones, 2000).

SLI is a developmental disorder in which children's oral language skills lag behind those of their same age peers, despite unimpaired hearing, age appropriate nonverbal IQ, and the absence of overt neurological problems. Estimates of the prevalence of SLI in monolingual

English-speaking kindergarten children center around 7.4% (Tomblin, Records, Buckwalter, Zhang, Smith, & O'Brien, 1997).

Children with SLI frequently begin to talk later than their peers, and may produce fewer utterances as toddlers (e.g. Joanisse & Seidenberg, 1998; Leonard, 1998). As these children enter the preschool years, the language deficits of children with SLI become more obvious, as they tend to exhibit particular difficulty with grammatical morphology. Grammatical morphemes such as verb tense markers (e.g. Rice & Wexler, 1996) plural markers, or articles (e.g. Leonard, 1998) may be absent or used in error. In addition to this, children with SLI may show decreased language comprehension (Leonard, 1998; Montgomery, 1995), phonological impairments (Roberts, Rescorla, Giroux & Stevens 1997), word learning difficulties (Kan & Windsor, 2010), and motor difficulties (Zelaznik & Goffman, 2010).

During the school years, the difficulty most closely associated with SLI is a deficit in word finding, often coupled with naming errors and low reading achievement (Lahey & Edwards, 1996; Leonard, 1998; Stothard, Snowling, Bishop, Chipchase & Kaplan, 1998). Though limited information is available on the lifelong course of SLI, persistence of this disorder into adolescence and adulthood has been associated with reduced psychosocial functioning and limited literacy and academic achievement (Clegg, Hollis, Mawhood, & Rutter, 2005). SLI is therefore best described as a disorder with its roots in early language, but which has the potential to affect the lifespan.

Deficits in NWR in children with SLI have been found relative to same-age peers, as well as younger, language-matched children, indicating that the deficit cannot be attributed to language experience (Montgomery, 1995). Decreased NWR performance in SLI has been found through twin studies to be highly heritable, and NWR deficits have proposed as a phenotypic

marker of the disorder (Bishop et al., 1996). In a meta-analysis of 23 studies comparing the NWR performance of children with SLI to typically developing peers, Graf Estes, Evans, and Else-Quest (2007) found that on average, children with SLI performed more than a standard deviation worse than typically developing children across NWR tasks.

Archibald (2008) summarizes several advantages of NWR over traditional language composite measures for the assessment of language impairment. The first is the comparative reliance of traditional language tests on prior knowledge, including vocabulary. In contrast, according to Archibald, NWR represents a processing-based measure that measures a child's ability to react to new information. Further, NWR tasks appear to be less culturally biased than other language measures. Rodekohr & Haynes (2001) compared the scores of white children with speakers of African American English (AAE) on NWR and an omnibus standardized test of language development. It was found that AAE speakers with typically developing language scored significantly lower than white participants. However, the two typical-language groups were not differentiable on the NWR task. Lastly, Archibald points out that NWR tasks are simple to administer, and can be adapted to multiple populations, including children as young as 2 years of age (e.g. Roy & Chiat, 2004).

NWR has the potential to provide critical diagnostic information in evaluating SLI. However, the factors underpinning children's performance on NWR tasks remain unclear. This project examined the skills that underpin nonword repetition task performance in young children with and without language impairments.

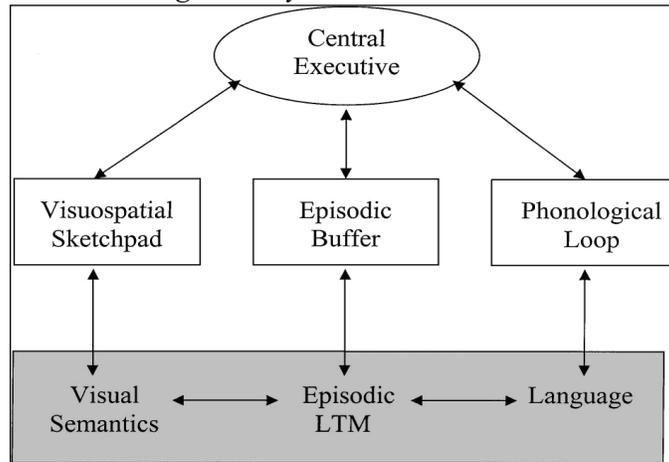
1.3 Potential influences on Nonword repetition performance

1.3.1 Phonological short term memory (PSTM)

In order to repeat an orally presented nonword, it is necessary to store the constituent sounds of that word. Consequently, much of the discussion surrounding NWR tasks deals with their use as a measure of phonological memory. Specifically, NWR tasks have been widely theorized to measure the phonological loop, a component of Baddeley and Hitch's (1974) multi-part model of working memory that deals with the temporary retention of verbal information (Baddeley, Gathercole, & Papagno, 1998). This component of memory is hypothesized as essential to establishing lasting phonological representations, thus contributing to overall language development.

Baddeley (1986) hypothesized a multi-part architecture functioning in concert to support cognitive function (see figure 1). This model involved separate rehearsal functions for visual information, maintained in the visual-spatial sketchpad, and phonological information, which was maintained in the phonological loop – also called the articulatory loop, or the phonological working memory system (Montgomery, 1995). According to this model, information in these two systems is subject to rapid decay over time, without the use of rehearsal. More recently, Baddeley added a 3rd slave system to his model of working memory, namely, the episodic buffer (Baddeley, 2000). This system provides connections to stored semantic meaning and long term memory. The episodic buffer is also purported to form links across the phonological loop and visuospatial sketchpad, creating multimodal units of information comprised of both visual and auditory information. In addition to the episodic buffer, visual-spatial sketchpad, and phonological loop, Baddeley's model of working memory (WM) involves a central executive, which functions to delegate tasks and allocate energy among the other systems within WM.

Figure 1
Baddeley's (2000) Revised working memory model



Note: LTM = Long term memory

Recent interest in NWR as a marker of language impairment sprang from a program of research by Gathercole and colleagues (e.g. Gathercole & Baddeley, 1989, 1990; reviewed in Baddeley, Gathercole, and Papagno, 1998). This research originally posited that NWR tasks represent a relatively pure, monolithic measure of phonological loop function. It was thought that the use of novel wordlike forms in the NWR would remove the opportunity for support from long term memory. Thus, NWR should be regarded as a more sensitive measure than alternate memory span measures, which rely on lexical items.

Several lines of research have found that NWR correlates highly with other measures of phonological short-term memory (PSTM) such as digit span (e.g. Gathercole, Willis, Emslie, & Baddeley, 1994; Baddeley, Gathercole, & Papagno, 1998; Dollaghan & Campbell, 1998). It has also been consistently observed that children with SLI perform most poorly on longer NWR items (Archibald & Gathercole, 2006), indicating that as phonological memory is taxed, performance decreases. Further, lower repetition performance has also been observed in

neuropsychological patients whose left hemisphere lesions result in verbal storage deficits (Trojano & Grossi, 1995). This confluence of evidence suggests that PSTM plays an essential role in NWR ability.

However, the claim that NWR represents a pure measure of phonological short term memory, to the exclusion of other influences, has met with criticism. For example, Snowling, Chiat, and Hulme (1991) point out that in order to repeat nonwords, one must not only store them, but perceive them, segment them into appropriately ordered phonological units, and formulate and execute an articulatory motor plan to reconstruct them. Each of these activities may contribute to overall NWR performance.

In addition, long term knowledge has been observed to influence this supposedly short term memory measure, in the form of wordlikeness effects. Wordlikeness refers to the overall resemblance of a word form to other words in a given language. This can be quantified in several ways, including phonotactic frequency, the frequency with which certain phonemes and phoneme combinations appear in English, or neighborhood density, the number of real words that are phonologically similar to the nonword in question. It has been observed that nonwords which bear a strong resemblance to words in English, or those high in wordlikeness (e.g. “diller”), are more easily repeated than less typical nonword items (e.g. “wudoip”; Edwards, Beckman, & Munson, 2004). This differential difficulty of nonword stimuli based on their degree of overlap with known word forms indicates that long term knowledge, specifically vocabulary knowledge, has an impact on NWR ability.

1.3.2 Vocabulary

The relationship between vocabulary and short term memory has been the subject of intense debate in the literature, both in general and with regard to their subsequent relationship with NWR tasks. Vocabulary has been found to affect NWR scores (Metsala, 1999; Edwards, Beckman, & Munson, 2004), despite the non-lexical nature of the task items. Children with poorer vocabularies tend to perform worse on NWR than their peers. Interestingly, the relationship between NWR performance and vocabulary skills seems to apply to receptive vocabulary only, while expressive vocabulary has been found to be unrelated (Coady & Evans, 2008; Briscoe et al. 2001).

To explain the association between NWR and receptive vocabulary, two prominent theoretical positions have emerged. The first prioritizes memory processes, positing that PSTM plays a critical role in the process of acquiring new vocabulary, by providing for immediate recall of the phonological form of a word. This immediate recall allows in turn for the new vocabulary item to be transferred to long term memory over repeated exposures to this phonological form (Baddeley, Gathercole, & Papagno, 1988; Gathercole, 2006). This position holds that PSTM underlies both vocabulary and NWR, accounting for the apparent relationship between vocabulary skills and repetition skills.

However, contrasting results were found by Melby-Lervag and colleagues (2012). In a longitudinal study, they investigated Baddeley, Gathercole, and Papagno's (1988) claims about the function of the phonological loop as a vocabulary learning device. Melby-Lervag and colleagues followed 219 Norwegian children over the course of 4 years, testing their receptive vocabulary and nonword repetition ability at 4 points over the course of the study. Their analysis revealed that, despite a highly significant initial correlation, nonword repetition had no

longitudinal influence on vocabulary scores, and vice-versa. Vocabulary scores at one time point, age 4, tended toward prediction of NWR scores one year later, but the effect did not reach significance. These results cast doubt on the findings of Gathercole et al., and suggest that vocabulary and NWR may be unrelated concepts. However, two additional perspectives must be taken into consideration.

First, the link between NWR and vocabulary growth in the Melby-Lervag (2012) study may have been obscured by the cyclical relationship between these factors found by Gupta and Tisdale (2008), described below. Further, it may be that the longitudinal study's reliance on NWR measures as a pure measure of phonological working memory may have resulted in the above findings. NWR has been linked to multiple other constructs, several of which are explored in the present study. It is possible that the interaction of multiple interrelated factors obscured the longitudinal relationship between NWR and vocabulary.

An opposing theoretical position on the potential relationship among vocabulary, PSTM, and NWR, proposes the inverse of Gathercole's interpretation; meaning that as individuals acquire more vocabulary knowledge, their phonological short term memory becomes more finely tuned to the sounds of their own language. This process, known as lexical restructuring (Metsala, 1999), improves the ability of the phonological short term memory to respond to input. Essentially, this theoretical position contends that larger vocabularies contribute to more efficient and robust processes of short term memory.

The idea that vocabulary and PSTM are inextricably linked is carried further by proponents of MacDonald and Christiansen's (2002) connectionist view of the language and memory systems. This account contends that instead of an underlying memory capacity that constrains language processing, what appears to be the effect of PSTM is instead the product of

the network architecture of language processing itself. MacDonald and Christiansen argue that tasks such as the reading span are only measures of overall language processing, the system working as a whole, rather than the measure of a particular verbal memory store. This view of working memory contrasts with that of Baddeley and colleagues (2000) discussed above, which envisions disparate, interconnected systems interacting with long term language knowledge at specific points.

In MacDonald and Christiansen's view, language processing occurs through the differential activation of nodes within a multilayer network, not through maintenance and manipulation within a separate memory system. The structure and efficiency of this connectionist network varies from person to person, depending on that individual's language experience. For example, the frequency with which a word is encountered will affect the ease with which it is perceived (frequency effects, see for example Ellis, 2002). Additionally, the similarity of a word to other words that an individual has encountered will also affect how that word is processed (wordlikeness/regularity effects, see for example Cortese & Simpson, 2000). Both of these well documented phenomena can potentially be attributed to an individual's language experience. To summarize, while Baddeley and colleagues envision PSTM interacting with vocabulary in specific ways, MacDonald and Christiansen would argue that in reality there is no separation between these two constructs.

Recent investigations utilizing computational modeling techniques examined the potential contributions of PSTM and vocabulary to NWR, as well as their relationships with one another. Gupta and Tisdale (2009) constructed their model by exposing the system to linguistic input, and allowing that input to influence the connections between hidden units which initially processed words, and context units which held a copy of those words for future output

(simulating short term phonological representations). This mechanism operationalized the effects of long term vocabulary on short term memory by making context units more likely to retain faithful copies of word forms that the system had encountered multiple times. When completing a NWR task, this computational model was able to faithfully reproduce length and wordlikeness effects seen in behavioral NWR testing of young children. This led the authors to conclude that the model was providing a fairly faithful analog of the way in which real cognitive systems may approach a NWR task.

To formulate conclusions about the interaction of PSTM and vocabulary as they relate to NWR, Gupta and Tisdale manipulated the input of both of these factors within the system. They observed that, while basing short term recall fidelity on long term vocabulary knowledge resulted in a model that produced similar responses to those of young children, learning of new vocabulary words was impossible without adequate PSTM function. Thus, a feedback loop emerged in which vocabulary knowledge and PSTM capacity influence each other, and mutually influence NWR performance.

While this explanation does provide a resolution to the competing perspectives on vocabulary/PSTM interaction by demonstrating that the two major theoretical positions are not mutually exclusive, it does not fully explain their relationship to NWR. PSTM was found to have a direct impact on NWR performance within the model. However, vocabulary has two potential avenues of influence on NWR. One interpretation is that vocabulary levels have a direct impact on NWR scores; the other is that vocabulary has a solely indirect impact on NWR, influencing this task solely via its impact on PSTM. Thus, while it seems clear that both PSTM and vocabulary are related to NWR tasks, how interdependent their influences are remains to be seen.

1.3.3 Literacy

The relationship between literacy development and NWR tasks is a complex one. Children who are below average in reading ability are significantly less accurate when repeating nonwords than their peers of average reading level (Brady, Poggie, & Rapala, 1989). An important side note to this result is the lack of screening for oral language capabilities, making it impossible to rule out language impairment in the sample. Expanding on the idea of a connection between reading and NWR, Gathercole, Willis, Emslie, and Baddeley (1991) found that among multiple tests of linguistic and non-linguistic skills, NWR and rhyme awareness were the strongest contributors to reading skills. Gathercole and colleagues interpreted their findings to indicate a common phonological processing component underlying both phonological memory and rhyme awareness tasks.

Though several investigations have identified a relationship between NWR and literacy, one in particular theorized a causal connection between learning to read and improved NWR scores. Nation and Hulme (2011) used a longitudinal paradigm to test the directionality of the NWR-literacy relationship in a sample of 215 first and second grade children. Reading ability was found to be a significant predictor of NWR scores one year later. This finding was independent of both the longitudinal relationship between oral language skills and NWR, and the autoregressor effect of NWR at time 1. These findings strongly suggest that learning to read affects performance on NWR tasks.

Nation and Hulme (2011) frame these results in terms of increased specificity of phonological representations due to the contribution of orthographic information. This view is essentially an extension of the lexical restructuring hypothesis proposed by Metsala and Walley (1998). Although the original lexical restructuring hypothesis solely centered on oral vocabulary

development, a role of orthography in this process provides a plausible account of Nation and Hulme's results. In their view, orthographic information gleaned from learning to read influences the overall phonological representations of sounds, allowing for more accurate repetition of nonword items.

These results are supported by those of Petersson, Reis, Askelof, Castro-Caldas, and Ingvar (2000), who compared the word and nonword repetition of literate and functionally illiterate adults. Using structural equation modeling (SEM) to analyze behavioral and functional neuroimaging data, Petersson et al. found that while literate participants processed word and nonword repetition tasks similarly, there was a significant difference in the nature of neural network interactions when illiterate participants switched from word to nonword stimuli. By analyzing illiterate participants' performance via a neural network model of language processing, Petersson et al. hypothesized that the observed differences centered in general control and support systems, or more specifically within central executive and attentional modulation aspects of the model. Interestingly, there was an additional significant difference related to the organization of articulatory output within the model. Overall, the authors interpret their results as an indication that the development of alphabetic written language skills creates lasting changes on phonological processing. These changes are evident as differences in NWR performance.

The relationship between literacy skills and NWR performance becomes more critical in light of recent findings using NWR to differentiate between children with SLI and those with comorbid SLI and dyslexia. Although SLI is usually diagnosed in early childhood, the disorder also puts children at increased risk of problems in later life, including reading disabilities such as dyslexia. Co-morbidity of SLI and dyslexia has been estimated at 68% (Flax, Realpe-Bonilla,

Hirsch, Brzustowicz, Bartlett, & Tallal, 2003). There is therefore a significant but incomplete overlap between the two disorders.

Some recent investigations suggest that low NWR scores may be primarily associated with cases of SLI coupled with dyslexia, rather than with SLI alone, and that in school aged populations, NWR scores can differentiate children with SLI only from those with co-morbid dyslexia (Baird et al., 2011; Catts et al., 2005; Conti-Ramsden & Durkin, 2007; Rispens & Pariiger, 2010). Catts et al. (2005) interpret these data to mean that previous findings of NWR repetition deficits in the SLI population in general may be due to the presence of co-morbid dyslexia. If this is the case, carefully designed NWR tasks may have predictive validity for future reading problems in children with SLI.

This line of reasoning, however, contrasts with that of Nation and Hulme (2011), discussed above. The finding that NWR scores differentiate children with SLI from their peers with comorbid SLI and dyslexia may indeed reflect a deficit specific to NWR that can be found in the comorbid SLI/dyslexia populations. However, it may also reflect the influence of influence of growing literacy skills on NWR performance, as noted by Nation and Hulme. Stated another way, children with SLI only may be seeing a boost in NWR scores due to their intact literacy skill, rather than the NWR task revealing an underlying skill that gives them a boost in literacy. The resolution of this contrast may be significant to the early detection of literacy problems in the SLI population.

1.3.4 Articulation

Recently, articulatory output processes have also received attention as potential contributors to NWR. In a comparison of NWR task characteristics, Archibald and Gathercole

(2006) found that the articulatory complexity of nonword items negatively influenced the performance of children with language impairment. Although the children in this sample had no clinically relevant motor speech deficits, their NWR scores were reduced when the nonword stimuli placed particular demands on their articulation skills. This effect was seen even when the influence of nonword length and wordlikeness was taken into consideration.

Interestingly, this discrepancy in scores between high and low complexity nonwords was only seen in children with SLI, not in typically developing children. This suggests that the simple articulatory demands of repeating lengthening novel sequences are disproportionately taxing to children with SLI. A similar pattern of results was observed by Archibald, Joanisse, and Munson (2013), who investigated the effect of motor output on nonword repetition by constraining articulation with a bite block held between participants' side molars. They investigated this effect in three groups: typically developing children, children with SLI, and children with working memory deficits in the absence of language impairments. Motor constraint did not significantly impair the NWR performance of typically developing children or those with working memory impairment alone. However, children with SLI repeated complex phoneme sequences less accurately when their articulatory movements were constrained. The authors concluded that while the role of motor skill in nonword repetition may be relatively modest in the typically developing population, children with SLI may have reduced motor skills, making them less able to make the necessary motor adjustments to compensate for motor constraint. Further, the absence of motor effects in the working memory deficit group indicates that the effect of motor constraint in SLI is not due to the increased memory load of adjusting speech mechanics, but may instead be a direct motor effect.

These results are perhaps unsurprising, in light of previous evidence that articulatory stability decreases in both children and adults as task complexity is increased. Kleinow and Smith (2006) analyzed articulatory kinematic data from a group of adult and child (9-10 year old) participants as they repeated sentence stimuli of increasing length and syntactic complexity. Results showed that both sentence length and the addition of syntactic elements caused decreased articulatory stability during production across groups. Similar results were found in a group of 5 year old children (Jones Maner, Smith, & Grayson, 2000), whose articulatory stability decreased as they repeated a 6 syllable phrase in sentences of high and low syntactic complexity. The effect of increased sentence complexity on the articulatory stability of these young children was significantly greater than on that of an adult comparison group. Further, increases in cognitive complexity have also been found to destabilize articulation. Dromey and Benson (2003) found that complex cognitive tasks, such as mental arithmetic, increase variability of articulatory movements in young adults. These results were interpreted to mean that the finite neural resources available are allocated differently in differing contexts, resulting in reduced capacity for precise articulation in the face of increased competing demands.

The above results have a direct relationship with the findings of Archibald and Gathercole (2006) and Archibald, Joanisse, and Munson (2013), in that children with SLI are hypothesized to possess overall reduced cognitive processing capacity, which affects the speed and efficiency with which they are able to process mental tasks. Given that resources for articulation appear to be strained in complex cognitive and linguistic contexts, it follows that these resources are drawn from the same overall capacity pool. Children with SLI, whose capacity is already reduced, will have fewer overall resources to bring to the task of NWR. Their

performance on this task will likewise show the effects of articulatory complexity more readily than that of their typically developing peers.

Archibald and Gathercole (2006) suggest two possible interpretations of the disproportionate effect of articulation on the NWR of children with SLI: first that these children have less robust phonological representations of less common phoneme combinations, or second, that children with SLI have difficulty actually forming the novel articulatory sequences required by the NWR task. In support of the latter explanation, Goffman (2004) reported that 4-7 year old children with SLI were less able to produce organized, stable rhythmic articulatory movements. Stark and Blackwell (1997) also found that both isolated oral movement as well as sequences of such movements correlated with NWR.

Additionally, skill at producing sequences of non-phonological oromotor movements has been found to affect performance on NWR tasks (Krishnan, Alcock, Mercure, Leech, Barker, Karmiloff-Smith & Dick, 2013). Krishnan et al. found that even when the phonological aspect of these movements was removed, tasks assessing the accuracy of oromotor control accounted for significant unique variance in NWR scores. These results seem to indicate that motor planning and sequencing does not need to be linked to a phonological representation in order to affect NWR.

1.3.5 Speech Perception

The quality of auditory input processing has also been shown to have an impact on NWR tasks, presumably by affecting the analysis of each nonword's acoustic form. For example, in a study of the NWR performance of young children with concurrent bilateral otitis media (OM), Gathercole et al. (2006) found that children with OM were less able to repeat nonwords than

their healthy peers, even though there was not a significant difference in the groups' auditory digit span scores. Relatedly, Dillon, Cleary, Pisoni, and Carter (2004) reported that profoundly deaf children with cochlear implants correctly repeated only 5% of the stimuli in a NWR task. Dillon, Burkholder, Cleary and Pisoni (2004) similarly found that speech perception tasks explained a significant amount of variance in the NWR performance of children with cochlear implants, even when digit span, verbal rehearsal speed, and degree of exposure to oral communication had been taken into account. Clearly, deficits in hearing acuity and poor quality of auditory input impacts NWR performance.

Additionally, a relationship between higher level auditory processing skills and NWR has been noted. In an investigation of wordlikeness effects in the NWR performance of children with phonological disorders, Munson, Edwards, and Beckman (2005) found that a speech discrimination task accounted for significant variance in NWR accuracy in the phonologically impaired group, but not in typically developing controls. Also, in an investigation of the genetic origins of auditory processing impairments in children with SLI, Bishop et al. (1999) found a significant correlation between performance on Tallal's auditory repetition task (ART) and a NWR measure. The sample of children in this investigation had typical hearing overall, but varied in their ability to distinguish tones in the ART. Thus, even more subtle differences in auditory and speech processing can be related to performance on NWR tasks.

This relationship between auditory processing and NWR is perhaps unsurprising, considering that children with SLI have demonstrated deficits in both repeating nonwords and processing auditory stimuli. Children with SLI are noted to perform worse than typical peers on several types of auditory processing tasks, such as discriminating speech stimuli (especially stimuli requiring processing of rapidly changing auditory input) and backwards masking (Rosen,

2003). However, not all investigations have found significant differences on such tasks (Bishop, Carlyon, Deeks, & Bishop, 1999). It is possible that such auditory processing difficulties, if present, may contribute to the overall deficits in NWR performance in children with SLI.

Additionally, such perceptual deficits in SLI have been linked to other cognitive features of the disorder, namely working memory problems. Joanisse and Seidenberg (2003) suggest that speech perception deficits in children with SLI interfere with the formation of an appropriate phonological code, and that this in turn leads to the reduced PSTM capacity noted in this population. If this holds true, both auditory processing deficits may contribute to NWR performance in children with SLI, but the two factors may be inextricably linked.

1.3.6 Test-taker Characteristics

Not only has each of the above skills been found to influence the NWR performance of children, but the skills underlying NWR have been found to vary depending on the population of children being studied. Children with SLI may draw on different underlying skills than those used by typically developing peers, or by other clinical populations, when completing this task. For example, a qualitative analysis of NWR errors made by adolescents with SLI vs. adolescents with autism (Riches, Loucas, Baird, Charman, & Simonoff, 2011) revealed that although overall scores on the task were comparable between the two groups, there was a greater effect of nonword length in the SLI group than in the group with autism. Longer nonwords were more difficult to repeat for both groups, but the difference was significantly greater in for children with SLI.

This suggests that, while these groups both have problems with NWR tasks, these difficulties may stem from disparate underlying causes. The length effect seen in the scores of

the SLI group may indicate that phonological memory impairment had more bearing on the performance of this group than on the group with autism.

A similar pattern of results was found by Cairns and Jarrold (2005), who noted that NWR scores correlated with vocabulary knowledge, but not with PSTM, in individuals with Down syndrome. Individuals with this diagnosis have demonstrated particular and specific deficits in verbal memory span. The authors interpreted this finding to indicate that, at least in this population, specific deficits in PSTM prevent this system from supporting nonword repetition. In the absence of such support, the contribution of vocabulary knowledge was necessarily more influential on NWR scores in this group.

Factors related to NWR performance in bilingual children were investigated by Lee and Gorman (2012). The overall error rates of the bilingual groups did not differ significantly from the monolingual English speaking comparison group, due to the high English proficiency of the bilingual children in the sample. Group differences were observed in the relationships between underlying factors and NWR. For example, Korean-English bilingual children demonstrated a weaker correlation between vocabulary and NWR accuracy than monolingual English speakers. These results further illustrate the concept that influences on NWR performance are not the same in every population.

Clearly, more than the population varied between the studies considered above. Differences in population age, NWR task used, and a host of other factors may have influenced the differing patterns of results. However, in light of the number of findings that test-taker characteristics influence relationships between underlying factors and NWR, it bears consideration that differences in the skills that underpin NWR task performance may vary between children with higher and lower ability in the current study as well.

1.3.7 Wordlikeness

Similar consideration must also be given to the characteristics of individual nonword items. Although several characteristics such as articulatory complexity and stress patterns have been examined in NWR items, the metric of wordlikeness has received the most attention. Wordlikeness refers to the overall typicality of a word form, that is, how closely a word form's component phonemes resemble patterns found in the rest of the language. Wordlikeness judgments are based on multiple factors, including neighborhood density, or the number of words that share a close phonological resemblance to a given word form, and phonotactic probability, or the likelihood that phonemes found together within a given word form are also statistically likely to appear together across words in the language. One specific metric of this type, the frequency with which two phonemes are likely to occur together within a word in a given language is known as 'biphone probability'. For example, the phoneme sequence /ft/ is frequently found in English words (e.g. lift, heft, fifty), making it a high-probability sequence. The phoneme sequence /fk/ in contrast, appears infrequently in English, making it a low-probability sequence (Munson, Kurtz, and Windsor, 2005). Neighborhood density and phonological probability are highly correlated (Vitevitch, Luce, Pisoni, and Auer, 1999), and both have been shown to contribute to speakers' perceptions of wordlikeness (Frisch, Large, and Pisoni, 2000).

Similar patterns of typicality apply to nonword sequences of phonemes, making examples such as "ral", which have a sound sequence more typical of English, appear more wordlike than pseudowords such as "wudoip", which is made up of less common sound sequences. The extent to which the novel word form resembles known words affects how much word learners are able to use prior lexical knowledge when faced with a new word. Of particular interest here,

wordlikeness influences the amount of prior knowledge participants are able to bring to bear within a NWR task.

Sensitivity to the phonotactic regularities of language has been demonstrated in infants as young as nine months (Jusczyk & Luce, 1994), and phonotactic probability and neighborhood density have both been shown to have consistent effects on language processing throughout the life span (Newman & German, 2005). Vitevitch & Luce (1999) found that spoken nonword processing, as measured by same/different recognition task, was facilitated by high phonotactic probability. This pattern of more rapid and accurate processing of high-probability sequences has not been found in studies using real word stimuli. In the work of Vitevitch and Luce, words with higher phonotactic probabilities were processed more slowly than lower-probability words. To explain this contradiction, the authors point out that only words are subject to competition effects from other members of their phonological neighborhoods. Pseudowords have no lexical entries, and therefore it follows that they show no lexical competition effects. In the absence of these effects, higher phonotactic probabilities afford an advantage in language processing tasks.

Wordlikeness characteristics have similarly measurable effects on NWR performance. Wordlikeness of individual nonword items varies greatly between NWR tasks (Graf Estes, Evans, & Else-Quest, 2007). While some NWR tasks, such as the *Children's Test of Nonword Repetition* (CNRep; Gathercole, Willis, Baddeley, & Emslie, 1994), utilize nonword items that contain real words and word-parts within them (e.g. “glistening”, “trumpetine”, “woogalamic”), other NWR tasks such as the nonword repetition subtest of the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999), contain no such familiar lexical items. This basic contrast in item construction results in a vast difference in the overall wordlikeness of each NWR task. Dollaghan, Biber, and Campbell (1993) found that

lexicality of syllables within NWR items affected performance on the task. School age children with typically developing language were significantly more able to repeat multisyllabic nonwords when the stressed syllables of those nonwords had lexical status. Differences in wordlikeness, whether they are created by lexical syllables or more subtle differences, may affect the difficulty of a nonword item, and the underlying skills utilized in the completion of the NWR task. Stated another way, items with higher wordlikeness may draw upon different underlying skills than less familiar nonwords.

Gathercole (1995) investigated this possibility in a study of 70 children tested at 4 and 5 years of age. Results of this study found that vocabulary skills and short term memory constrained NWR performance differently dependent on wordlikeness characteristics. Children's PSTM scores were more closely linked to NWR performance on less wordlike stimuli than with highly wordlike stimuli. Vocabulary skills, however, were associated with both high- and low-wordlikeness items, depending on the age of the participants. A larger advantage in repeating highly wordlike nonwords was found at 5 years than at 4 years, indicating what the author describes as a developmental increase in the use of lexical background knowledge in the task.

Wordlikeness has been found to affect nonword repetition in both children with SLI and their typically developing peers, in that more wordlike nonwords tend to be repeated more accurately than less wordlike items (Edwards, Beckman, & Munson, 2004). Further, this effect has been found to disproportionately influence the scores of children with SLI (Munson, Kurtz, & Windsor, 2005). Coady, Evans, and Kluender (2010) also investigated phonotactic probability as they compared the NWR performance of children with SLI and their typically developing peers. Unlike Munson, Kurtz, and Windsor (2005) however, this investigation did not find an interaction between group and phonotactic frequency, indicating that both groups were

comparably affected by frequency differences in nonword stimuli. Taken together, these results indicate a high likelihood that wordlikeness characteristics, specifically the probability of phonological sequences within nonword items, will affect performance of the children in the current study. Further, manipulation of this variable may affect the underlying skills supporting the task.

1.4 Purpose & Hypotheses

NWR tasks have been the subject of increasing research attention and clinical popularity. However, there is no current consensus as to what precisely is measured by this task. It has been used clinically as an indicator of PSTM capacity, literacy, and articulatory motor skill (Coady & Evans, 2008), while additional lines of research acknowledge that vocabulary and auditory processing may also contribute to NWR performance. A deeper understanding of the skills underlying NWR could lead to the development of NWR tasks with increased utility in differential diagnosis, and increased predictive power. The current study intends to contribute to this deeper understanding by considering contributing skills, nonword item characteristics, and population differences in NWR performance.

The current series of experiments uses a regression approach to simultaneously examine the effects of multiple potential underlying factors on the NWR scores of young children. Specifically, the following research questions are addressed:

Analysis 1. When considering PSTM, vocabulary, literacy/reading readiness, speech processing, and articulatory output skills, what predicts NWR performance in 4-6 year old children?

Analysis 2. Do children with high and low language abilities differ in the skills that underpin NWR performance?

Analysis 3. Does the wordlikeness of individual NWR items affect the skills that support this measure?

It is hypothesized that, contrary to the original characterization of NWR tasks as relatively pure measures of phonological memory (Gathercole & Baddeley, 1990a; Gathercole & Baddeley, 1990b; Henry & Millar, 1991), multiple cognitive skills are related to NWR performance in children at this age. In particular, vocabulary skills are likely to have an independent relationship with NWR in addition to that of PSTM, as hypothesized by Gupta & Tisdale (2009). If the influences of vocabulary and PSTM on this measure appear highly interdependent, it may be indicative of a closer relationship between these constructs, as posited by MacDonald and Christiansen (2002).

It is also hypothesized that test taker characteristics, specifically language skill, will affect the underlying skills associated with NWR performance. It has been found across varying populations that factors related to NWR, such as vocabulary, phonological awareness, articulatory skill, can vary in their influence on this task. In this population, it is expected that the influence of articulatory output measures may increase in the lower language score group. This would coincide with the findings of Archibald and Gathercole (2006), who observed that NWR performance of children with SLI was negatively affected by increasing the articulatory complexity of nonwords. Typically developing children, however, were unaffected by change in this variable. This finding may be replicated in the current study, despite the overall typicality of the sample.

In analysis 3, it is expected that differences in wordlikeness characteristics, specifically biphone probability, will create differences in the underlying skills which support the NWR task. It is expected that vocabulary will predict NWR scores on both low and high probability nonword lists, while PSTM as measured by digit span will be related to high probability nonwords only. This pattern of findings concurs with that found by Gathercole (1995), in children of similar age.

Chapter 2: Methods

2.1 Participants

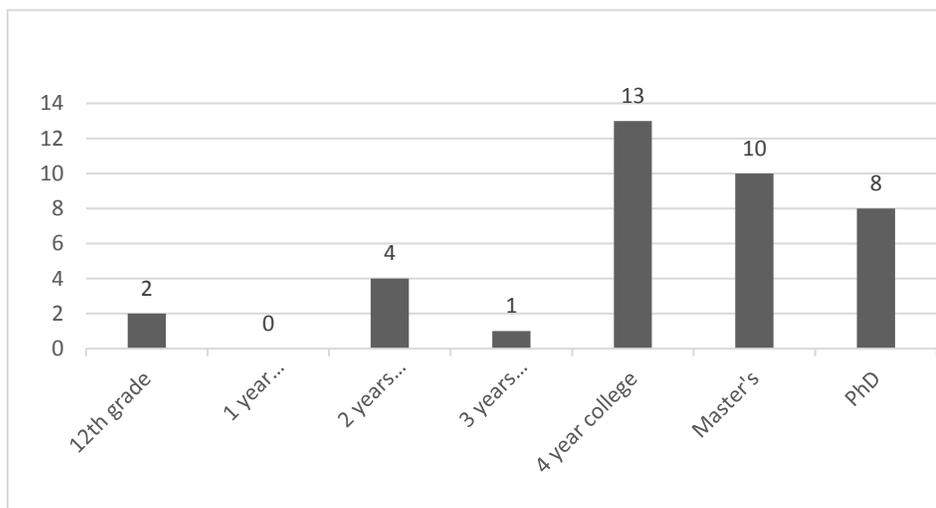
For this series of experiments, 41 children between the ages of 4;0 and 6;5 were recruited from preschool and day care programs in Tolland County, CT and Ulster County, NY. Of these, one participant did not meet screening criteria. Three additional participants were unable to complete the testing due to scheduling constraints, leaving 37 children who completed the testing protocol. Of these, 17 were female. The mean age of the sample was 5.1 (range 4.0-6.3) years.

Four to six year olds were chosen for this study for two reasons. First, there is evidence that children of this age are not reliably able to use memory support strategies such as subvocal rehearsal (Gathercole & Adams, 1994). This increases the likelihood that the participants relied on PSTM, rather than conscious memory support strategies, to complete memory tasks. Second, this is the age when many children with suspected language impairments undergo extensive testing as they move from preschool, to kindergarten, and into first grade. It is therefore clinically important to understand what is being measured by NWR tasks at this age.

All participants were monolingual English speakers with hearing within normal limits according to pure tone hearing screening, and typical nonverbal reasoning skills as assessed by the *Columbia Mental Maturity Scale* (CMMS; Burgemeister, Blum, & Lorge, 1972). Children with existing diagnoses of autism spectrum disorder, seizure disorders, or overt neuropsychological disorders were excluded from participation. Testing was performed at university speech and hearing clinics at the University of Connecticut or the State University of New York at New Paltz, or at the participant's school or child care facility.

Socio-economic status (SES) was measured using mother's/primary caregiver's highest achieved education level as a marker. The sample was high-SES overall, with most mothers reporting a 4-year college degree or higher. All mothers reported achieving at least a 12th grade education (see Figure 2).

Figure 2:
Mother's (or Primary Caregiver's) education level



Despite efforts to recruit children with language impairments into the study, the majority of the participants who completed the test battery had typically developing language. Five children (4 male, 1 female) scored below the recommended cutoff score of 87 on the SPELT-P2

(Greenslade, 2009), and could be considered language impaired. Because this group of impaired children was too small to study separately, they were included in the sample as a whole and a median split was used to divide higher from lower language skill. This allowed the impaired children's scores to be considered while maintaining statistical power.

2.2 Testing battery & procedures

Over the course of 3 visits of approximately 1 hour each, children participated in a battery of cognitive, language, and behavioral evaluations in order to assess their PSTM, articulatory output skills, literacy skills, vocabulary, and speech perception. Prior to the full battery, each child was given a pure tone hearing screening and a measure of nonverbal reasoning as screening measures. Expressive language skills were also measured, using the *Structured Photographic Expressive Language Test –Preschool, 2nd edition* (SPELT-P2; Dawson et al., 2005). In addition, participants completed 2 NWR activities, the modified *Children's Test of Nonword Repetition* (CNRep; Gathercole, Willis, Baddeley, & Emslie, 1994), and the *Nonword Repetition Subtest of the Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999). Evaluation measures are discussed in further detail below.

2.2.1 Screening Measures:

A background questionnaire was completed by the parents/guardians of each participant. The questionnaire gathered information regarding educational history, language exposure, socioeconomic status, and reading/ home literacy environment. Responses on the questionnaire was used to rule out frank neurological impairments, autism, and seizure disorders. Socioeconomic status was determined by the mother's education level. A pure-tone hearing

screening (500, 1000, 2000, and 4000Hz) was completed by each participant on their first visit, to rule out overt hearing problems. Lastly, the *Columbia Mental Maturity Scale* (CMMS; Burgemeister, Blum, & Lorge, 1972), an age-appropriate measure of nonverbal reasoning ability, was administered to assess nonverbal intelligence in this sample. A standard score of 85 or above was considered typical, and children with lower scores were excluded from the sample.

2.2.2 NWR Measures:

The NWR tasks that make up the focus of the current project are modified versions of two popular published NWR measures: the *Children's Test of Nonword Repetition* (CNRep; Gathercole & Baddeley, 1996), and the Nonword repetition subtest of the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte 1999). To create a more child-friendly experience, recorded nonword stimuli from each measure were embedded in a game context. Participants were presented with cartoon “aliens” whose names had to be repeated before they could board their “spaceship” (see Figure 3). To reduce practice effects between the two NWR measures, order of presentation was counterbalanced between participants. All nonword items were administered to all participants.

Participant responses were recorded as .WAV files, and later transcribed using the International Phonetic Alphabet (IPA). NWR performance was scored at the phoneme level, in terms of percent phonemes in error. Responses were scored as correct or incorrect in relation to phonemes in the target nonword. Scoring procedures utilized performance on the GFTA to account for the phonological limitations and misarticulations common to children in this age group. Specifically, substitutions that appeared on the GFTA were scored as correct responses within nonwords on the CNRep and CTOPP-NR. Similar scoring systems have been used previously (e.g. Gray, 2003; Deevy, Weil, Leonard, & Goffman, 2010) to compensate for the

phonological limitations often seen in children whose speech sound systems are continuing to develop. Each NWR sample was transcribed by the primary investigator and a trained student research assistant. In the event of a scoring discrepancy, differences were examined and samples re-scored independently. Inter-rater reliability of at least 90% was maintained across all samples.

Figure 3.

Examples of graphics and prompts from the NWR “Alien Game”



“Diller”

“There he is! You found Diller!”

These particular NWR tasks were selected to represent a cross-section of differences often found in NWR measures. The items vary widely in terms of wordlikeness, length, and articulatory complexity. The CNRep was developed as a stand-alone NWR measure comprised of 40 nonwords, ranging from two to five syllables and conforming to English phonotactic and prosodic rules. Wordlikeness of the CNRep nonwords is relatively high, and a number of real English morphs (e.g. *altupatory*, *pristoractional*) appear in the three, four, and five syllable items. Due to concerns about the length of the task and fatigue in the participants, a subset of 20 of the CNRep items was used for this set of experiments. This shorter list of nonwords included five items each of 2-5 syllables, ranging for 4 to 13 phonemes. A similar subset of CNRep nonwords was used by Grey (2003), and Archibald and Gathercole (2006).

The CTOPP was originally published in 1999 as a norm-referenced test measuring aspects of phonological coding. It includes subtests in three composite areas: phonological awareness, phonological memory, and rapid naming. The nonword repetition task makes up part of the CTOPP's phonological memory composite score. To create the nonword items, the authors randomly combined phonemes to fill positions in syllables in order to avoid items analogous to real words (Lennon & Slesinski, 2002). The resulting list contains 18 nonword items ranging from 1 to 9 syllables (3-15 phonemes) in length. The CTOPP nonword items are low in wordlikeness compared to those used on the CNRep, with no English morphs present and low phonotactic probability.

2.2.3 Measures of Contributing Skills:

Phonological Short-Term Memory:

For the present study, PSTM was measured with the Digit Span subtest of the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999). Digit Span tasks are widely used to estimate PSTM capacity; this particular task uses verbal forward digit span, meaning that participants hear recorded series of randomly-ordered digits, and are asked to repeat them back in the same order. Number series increase in length until ceiling level (three consecutive errors) is reached.

Vocabulary:

Measures of both receptive and expressive vocabulary was included in the assessment battery in order to get a fuller picture of participants' language abilities. *The Expressive Vocabulary Test- 2nd Edition* (EVT; Williams, 1997) provides a measure of word retrieval and

expressive vocabulary. It is appropriate for 2.5-90+ years of age and can be completed in 15 minutes. Test items for this measure require the participant to label the picture, answer a certain question, or think of a synonym. Test/retest reliability and internal consistency on this measure have been found to be high (.94-.97 and .93-.94, respectively).

The Peabody Picture Vocabulary Test, 4th Edition (PPVT; Dunn & Dunn, 2007) is a widely used receptive vocabulary measure co-normed with the EVT. It is appropriate for 2;6-90+ years, and takes between 10 and 15 minutes to administer. The test consists of panels showing 4 line drawings; participants are asked to indicate via pointing which of the pictures corresponds to a particular vocabulary item. Test/retest reliability correlations on this measure have been found to be very high (.92-.96) with internal consistency correlations of .94-.95. Convergent validity with the EVT was also found to be high (.80-.84).

Literacy skills:

The Rosner Test of Auditory Analysis (TAAS; Rosner, 1993) is a criterion referenced measure of phonological awareness designed for use with kindergarten and preschool children. It evaluates the participant's ability to isolate and manipulate individual sounds within words. The child is presented with a spoken word (e.g. "say cowboy"), and asked to repeat it while leaving out part of the word (e.g. "now say it again, but don't say 'boy'"). Items on the task vary from deletion of syllables, to deletion of one consonant within a consonant cluster (e.g. "say stale without the /t/"). This task provides an age-appropriate measure of phonological awareness, a critical pre-literacy skill.

The Test of Preschool early Literacy (TOPEL; Lonigan, Wagner, Torgeson, & Rashotte, 2007) is an assessment of emergent literacy skills for ages 3 - 6. It consists of 3 subtests: Print

Knowledge, Definitional Vocabulary and Phonological Awareness. The TOPEL reports high convergent validity with other measures of similar constructs, such as the *Test of Early Reading Ability* (TERA-3). For the present study, the Print Awareness subtest was administered to gather information on participants' knowledge of print conventions and the alphabetic principle. This subtest includes letter-name identification, letter-sound identification, and word-picture discrimination activities.

Speech Perception:

The *SCAN-C Test of Auditory Processing Disorders in Children* (SCAN-C; Keith, 2000) is the most widely used test for auditory processing disorders used in the US and UK (Dawes & Bishop, 2007). It is normed for children as young as 5, and has a complete testing time of approximately 20 minutes. For the current project, two of the four SCAN-C subtests were administered: Auditory Figure/Ground (AFG) and Filtered Words (FW). In the AFG subtest, children are asked to repeat monosyllabic words that are presented against a background of multi-talker babble. In the FW subtest, the child is asked to repeat monosyllabic words that have been distorted with a low-pass filter of 1000 Hz. These subtests have been selected both to shorten the overall testing time associated with the SCAN-C, and to minimize the confounding influence of phonological short term memory that has been observed in the other SCAN-C subtests, which involve longer stimuli (Lum & Zarafa, 2010). Test-retest reliability was found to be 0.65 to 0.82 for these subtests. Due to the age variation of the present sample, raw scores for this task were used in analysis.

Articulatory Output:

Diadochokinetic rate (DDK) is a measure of articulatory motor coordination and stability, consisting of speeded repetitions of alternating syllables (e.g. /pa/, /ta/, and /ka/) followed by multiple contiguous repetitions of the resulting trisyllable nonword (/pataka/. Researchers demonstrated each syllable to the participant, and encouraged the participant to continue until 20-30 repetitions have been produced. In several cases, however, repetitions were halted as the participant became frustrated or fatigued. The percentage of correct repetitions of the trisyllable was calculated for analysis.

In addition to articulatory stability data with the DDK, standardized articulation data were also collected. The Goldman-Fristoe Test of Articulation 2 (GFTA; Goldman, 2000) is a standardized measure of articulation of consonant sounds and clusters. The age range of this test includes ages 2-21 years old with standard scores based on age and normative information for males and females. The participant is asked to respond to verbal cues and picture plates with single-word answers that provide common speech sounds.

It is important to note that information from the GFTA was also utilized in the scoring of NWR measures. It was not the intention of this study to penalize children for habitually misarticulated sounds, but instead to score NWR responses in a way that reflects their phonological system. Thus, if a participant substitutes /w/ for /r/ on the GFTA, then she was not penalized for producing /wal/ for /ral/ on a NWR task.

2.2.4 Assessing participant language skills

For analysis 2, comparing the predictor variables for NWR scores in participants with high and low language skill, the *Structured Photographic Expressive Language Test – Preschool*

2 (SPELT-P2; Dawson, Eyer, Fonkalsrud, 2005) was employed as an expressive language measure. The SPELT-P2 is designed to elicit morphological and syntactic structures in children 3-0 through 5-11. The test consists of 44 full color photographs of typical scenarios and objects, paired with direct questions and statements. The SPELT-P2 assesses a child's strengths and weaknesses using structures such as prepositions, articles, plurals, negatives, third person markers, and conjoining. The test has shown good excellent sensitivity and specificity in identifying children with language impairment (Greensalde, Plante, & Vance, 2009). It is notable that despite being older than the normed sample for this test, there was no notable ceiling effect in the current study.

2.2.5 Assessing Nonword characteristics

Analysis 3 compares the predictor variables for NWR scores on nonword items that are low and high in wordlikeness. To carry out this comparison, individual NWR items were used to construct the high and low wordlikeness lists for analysis. Wordlikeness in this case was equated with average biphone probability according to the Storkel and Hoover (2010) online phonotactic probability calculator (<http://wordlearning.ku.edu/child-calculator>). Average biphone probability is calculated by calculating the likelihood of two phonemes occurring together, and averaged across each phoneme pair in a given nonword. This particular probability calculator assess biphone probability based on child corpora of American English.

Since very few of the tested nonwords had any neighbors, neighborhood density for items in each list was held at zero. High- and low- biphone probability wordlists were balanced for length in syllables and in phonemes, and were found to be statistically different on biphone probability via t-test ($t(14)=7.4217$, $p<0.0001$). The high and low wordlikeness lists, with average biphone probabilities, are listed in Table 1 below.

Table 1

High and Low wordlikeness nonword lists with Mean Biphone Probabilities

Low Wordlikeness Nonwords		High Wordlikeness Nonwords	
	Biphone Mean		Biphone Mean
wudoip	0.0006	balope	0.0039
wulanawup	0.0007	stoppagratic	0.0035
Nigong	0.0009	sladding	0.0057
Rubid	0.0012	hampent	0.0053
vozetoov	0.0012	glistering	0.0047
tibudieshalt	0.0013	penneriful	0.0032
viversoomauge	0.0015	rooterpation	0.0027
versatrationist	0.0013	contramponist	0.0033
Average	0.0010875	Average	0.0040375
<i>Note.</i> Mean biphone probabilities according to Storkel & Hoover's (2010) child phonotactic calculator			

2.3 Data analysis and Model Selection

Given the large number of potential predictive variables relative to the n for this study, several steps were taken to maximize its explanatory power. First, attention was given to dimension reduction by combining appropriate variables. The two subtests of the SCAN-C were combined into a single speech perception measure. Using this new combined variable among the predictors, partial correlations controlling for age were performed to explore relationships between the raw scores of potential underlying factors and NWR percent error. Correlation analysis also revealed any potential collinearity among the predictor variables which may influence the subsequent regression analysis.

To further explore the relationships between potential underlying factors and NWR scores, best subsets regression analysis was used. Best subsets regression procedure is used to help with model selection, by identifying all possible regression models from combinations of all

candidate variables. This naturally results in a large number of possible models, from which the most appropriate model can be chosen. The best fitting model for each analysis was chosen using several criteria: adjusted R-squared, Mallow's Cp, and overall number of predictors.

R-squared, or the coefficient of determination, measures the degree to which change in a regression's response variable can be explained by changes in the predictor variables. However, one difficulty with using the R-squared to judge the goodness of fit of a regression analysis is the automatic increase of this statistic with the addition of more predictors to the model. It is therefore tempting to fit larger models, which will appear to have a better fit when they in fact simply have more terms. The adjusted R-squared takes into account the number of predictors in the model, and presents a clearer picture of the model's explanatory power. Unlike the R-squared, adjusted R-squared will only increase with the addition of a new predictor, if that new term increases the fit of the model more than would be expected by chance. For the current study, choosing the set of predictors with the largest adjusted R-squared was the first criterion for model selection.

Mallow's Cp is a criterion for model selection that assesses the relative fits of models with different numbers of predictors. This statistic compares the mean squared error (MSE) of each proposed model to the MSE of a model including all potential variables. Ideally, the Mallow's Cp value should be close to the number of predictors in the proposed model. A low Mallow's Cp value, close to the number of predictors in the potential model, was used as an additional criterion for model selection.

The number of predictors in the proposed model was considered as a final criterion for selection. Given the relatively low n , a more parsimonious model is desirable. In the event that

two models proposed by the best subsets selection procedure had similar or identical adjusted R-squared and Mallows' Cp values, the model with fewer predictors was chosen for analysis.

Once a combination of predictors has been selected using these criteria, regression analysis was used to test the resultant model. In the event that variables in the model are non-significant, they were removed and the resulting, more parsimonious model was retested. This model was then evaluated using the F-test of overall significance, which compares the fit of the proposed combination of predictors to a model with no predictors (intercept-only model). A significant F-test of overall significance indicates that the model provides a better fit than a model including only the intercept. Adjusted R-squared values and standard error of the regression (S) were used to evaluate the strength of the relationship between the model and the dependent variable. Residuals plots were examined to ensure that the observed error is random with approximately normal distribution. Non-random patterns in residuals were investigated for evidence of autocorrelation or missing elements.

Chapter 3: Results

For this series of experiments, 41 children between the ages of 4;0 and 6;5 (mean age 5;1) were recruited. Over the course of testing, 3 families dropped out due to scheduling constraints, and 1 child did not meet the *CMMS* criterion score to participate. This left a total of 37 participants for analysis. Data were analyzed pairwise to maximize available scores. Descriptive statistics of the sample for each measure are provided in Table 2.

Shapiro-Wilk Tests showed that several of the collected variables were non-normally distributed. This was confirmed with visual inspection of each variable's histogram. To avoid violating the normality constraint in subsequent regression analyses, transformations were

applied to non-normally distributed values according to the direction and severity of the skew. High- and Low-wordlikeness NWR scores, as well as GFTA Raw scores, required a square root transformation. The transformed, normally distributed variables were used for all subsequent analyses. Unfortunately, TAAS scores exhibited a strong floor effect that resisted transformation by both Log10 transformations, and had to be removed from further analysis.

Table 2.

Descriptive statistics of the sample, using the raw scores of each measure.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Age	37	4.000	6.420	5.08565	.709875
CMMS	37	25.0	52.0	39.676	6.7001
Spelt-P2	37	58	125	29.135	5.889
GFTA 2	37	.0	22.0	7.892	6.0818
DDK Rate	34	2.2	4.5	3.297	.5813
DDK% Correct	33	.0	100.0	74.391	23.4479
SCAN-C	30	20.0	33.5	27.233	2.8398
TOPEL	36	8	36	28.028	10.829
PPVT	37	53.0	145	113	23.624
EVT	36	46.0	119	82.972	21.894
Digit Span	36	6.0	18.0	10.417	3.1385
Combined NWR % Error	36	3.06748466	27.91411043	15.1414451261	6.97113052995

This sample was also for missing values on several tasks, owing to refusal on the part of some participants. In particular, the SCAN-C task was refused by seven of the 37 participants. In addition, while the missing values on other measures were distributed throughout the sample, children with lower language skill were more likely to refuse this task. These non-random

missing values made it inappropriate to impute the overall sample mean for replacement, nor could multiple imputation procedures be employed. No combination of other variables significantly predicted SCAN-C performance, preventing the use of a regression equation for imputation. In order to preserve overall explanatory power, the SCAN-C was omitted from subsequent regression analyses. It was included in partial correlations controlling for age, which preserved cases by eliminating missing data pairwise.

Partial correlations controlling for age revealed significant relationships between overall NWR performance and several variables. After Bonferroni-Holm correction procedure was applied to account for multiple tests (Gaetano, 2013), SPELT-P2, digit span, receptive and expressive vocabulary scores, and TOPEL scores correlated significantly with NWR Error percentage.

Table 3.
Partial correlations of the whole sample, controlling for age.

Measure	1	2	3	4	5	6	7	8	9	10
1. SPELTP-2	1.000									
2. DDK Rate	.011	1.000								
3. DDK%Correct	-.399	-.174	1.000							
4. GFTA	-.435	.100	.090	1.000						
5. SCAN-C	.366	-.030	-.252	-.026	1.000					
6. PPVT	.759**	-.111	.445*	-.353	.121	1.000				
7. EVT	.620**	-.160	-.311	-.359	.255	.755**	1.000			
8. TOPEL	.611**	-.070	-.350	-.318	.359	.695**	.736**	1.000		
9. Digit Span	.466*	-.228	-.216	-.220	.250	.513*	.411	.548*	1.000	
10. NWR % Error	-.490*	.060	.381	.385	-.147	-.605**	-.450*	-.430*	-.683**	1.00

Note. Marked correlations are significant after Holm-Bonferroni correction.

Partial correlations also revealed statistically significant relationships among several potential predictors. This was further investigated using variance inflation factor (VIF), as

multicollinearity among predictors can have deleterious effects on regression accuracy. VIF for each variable is shown in Table 3. TOPEL scores showed an unacceptably high VIF, and therefore had to be removed from the model. The multicollinearity of the remaining variables was low enough to proceed with regression modeling (see table 4).

Table 4.

Colinearity statistics for predictor variables, before and after removing TOPEL scores

Colinearity including TOPEL-TAAS scores

Colinearity after removing TOPEL-TAAS scores

Model		Collinearity Statistics		Model		Collinearity Statistics	
		Tolerance	VIF			Tolerance	VIF
1	(Constant)			1	(Constant)		
	DDK Rate	.739	1.353		DDK Rate	.769	1.301
	DDK % Correct	.561	1.784		DDK % Correct	.580	1.725
	SCAN-C	.665	1.504		SCAN-C	.691	1.446
	Digit Span	.247	4.051		Digit Span	.555	1.803
	GFTA	.734	1.362		GFTA	.834	1.199
	PPVT	.314	3.188		PPVT	.315	3.173
	EVT	.296	3.383		EVT	.427	2.340
	Spelt-p2	.470	2.127		Spelt-p2	.492	2.033
	TOPEL	.142	7.045				

Analysis 1. What underlying skills are related to NWR Performance?

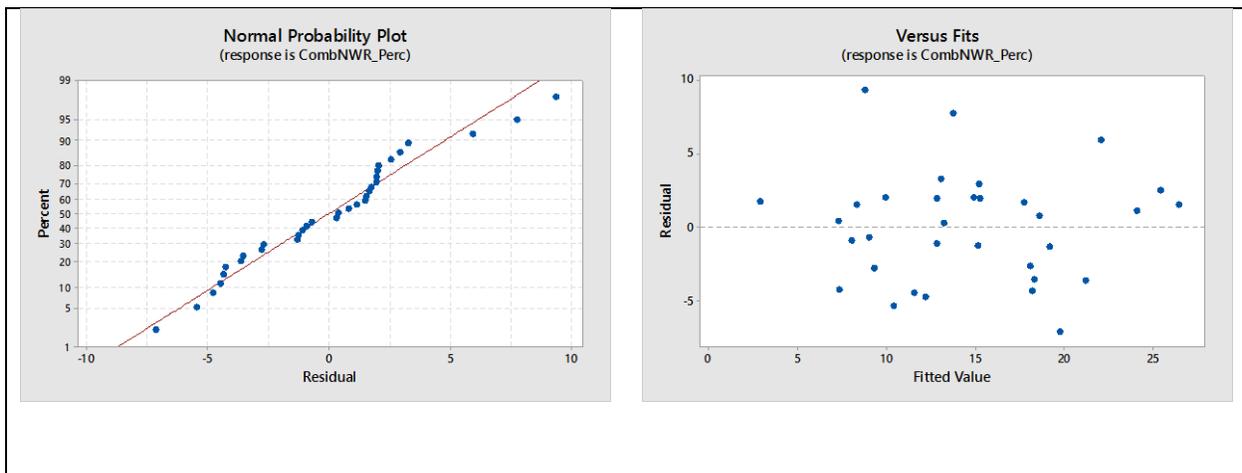
Best subsets regression was initially used to determine the optimal combination of variables to predict overall NWR percent error. Adjusted R-Squared, Mallow’s Cp, and standard error were used for model selection. To predict NWR error in the overall sample, the optimal model contained 5 predictors: Age, average DDK rate, GFTA score, PPVT score, and digit span. This model had an adjusted R-squared of 64.6 and a Mallow’s Cp value of 5.3, indicating that it was likely to be a reasonably good fit for the data.

Multiple regression analysis was used to test whether this combination of scores performed as predicted in the best subsets regression. The overall F-statistic indicated that the model itself was significant, as it explained more of the variance in NWR performance than the intercept-only model. The adjusted R-squared of the overall model was 64.59% and the overall model error was 4.05. Coefficients for this model are listed in Table 5.

Table 5.
Coefficients for the 5 variable model predicting overall NWR performance.

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	28.40	8.32	3.41	0.002	
Age	3.33	1.23	2.70	0.012	1.50
DDKRate	-1.35	1.32	-1.03	0.313	1.05
DigitSpan	-1.158	0.289	-4.01	0.000	1.67
GFTA	1.052	0.671	1.57	0.129	1.31
PPVT	-0.1444	0.0441	-3.27	0.003	1.85

Figure 3:
Normal probability plot and residuals-versus fits plot for the 5 variable model predicting overall NWR performance



In this model, digit span, PPVT, and age were the only significant predictors of NWR performance at the p 0.05 level. Examination of residual plots for this model revealed

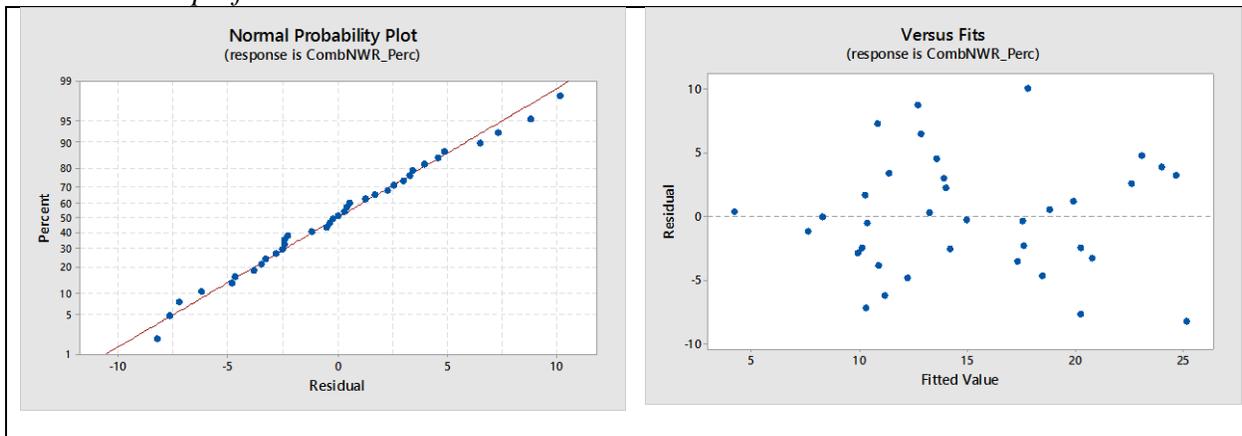
symmetrically distributed residual values clustered under 10, and an approximately normal distribution (See figure 3).

This model was refined by removing average DDK rate and GFTA score as predictors. The resulting 3-variable model was also significant overall, with a comparable adjusted R-squared of 53.96% and overall error value of 4.73. This more parsimonious model indicated that digit span and PPVT scores were robust predictors of NWR performance. Age was no longer a significant predictor in this model. Residuals for this model were clustered below ten and normally distributed (see Figure 4). Coefficients for the more parsimonious 3-variable model are shown in Table 6.

Table 6.
Coefficients for the 3 variable model predicting overall NWR performance.

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	32.94	6.11	5.39	0.000	
Age	1.32	1.26	1.04	0.304	1.27
DigitSpan	-1.213	0.330	-3.67	0.001	1.68
PPVT	-0.1051	0.0418	-2.51	0.017	1.57

Figure 4:
Normal probability plot and residuals-versus fits plot for the 3 variable model predicting overall NWR performance.



Analysis 2. Do factors relating to NWR performance differ between children with high and low language skills?

To answer this question, the sample was split at the median SPELT-P2 score. High and low language score groups were confirmed to have different SPELT-P2 scores via an independent samples T-test, $t(24) = -6.090$, $p < .001$. Descriptive statistics for the high and low language scoring groups are shown in Tables 7 and 8.

Table 7.
Descriptive statistics of the lower language scoring group

	N	Minimum	Maximum	Mean	Std. Deviation
Age	19	4.000	6.420	5.11411	.747594
CMMS	19	25.0	49.0	37.053	6.5530
Spelt-P2 SS	19	58	105	93.32	12.079
GFTA-2	19	.0	22.0	9.895	6.3587
DDK Rate	17	2.4	4.5	3.282	.5670
DDK % Correct	16	.0	100.0	64.187	27.9481
SCAN-C	14	20.0	33.5	26.179	3.6088
TOPEL	19	8.0	36.0	23.7	11.489
PPVT	19	53.0	144.0	103.89	27.52
EVT	19	46.0	104.0	77.37	18.77
Digit Span	19	6.0	15.0	9.105	2.8847
Combined NWR % Error	19	6.44171	27.91411	17.96900	6.07444

Given the small n of each language skill group, partial correlations were initially used to determine relationships between predictor variables and NWR scores. Regression analyses were then attempted for each group.

Table 8.

Descriptive statistics of the higher language scoring group

	N	Minimum	Maximum	Mean	Std. Deviation
Age	18	4.167	6.170	5.05561	.688111
CMMS	18	31.0	52.0	42.444	5.8130
Spelt-P2 SS	18	106	125	114.11	5.268
GFTA-2	18	.0	21.0	5.778	5.1284
DDK Rate	17	2.2	4.4	3.312	.6123
DDK % Correct	17	60.0	100.0	83.994	12.8545
SCAN-C	16	25.0	30.0	28.156	1.5353
TOPEL	17	20	36	32.88	4.4844
PPVT	18	104.0	145.0	122.61	13.76
EVT	17	62.0	119.0	89.24	13.59
Digit Span	17	7.0	18.0	11.882	2.8036
Combined NWR % Error	17	3.06748	27.91411	11.98123	6.68496

For the lower language group, partial correlations controlling for age indicated that only receptive and expressive vocabulary scores were significantly related to NWR performance. Correlation coefficients for all variables are shown in Table 9 below. Interestingly, both EVT and PPVT scores were found to be significantly correlated with NWR performance in the lower language score group. This contradicts the findings of Briscoe et al. (2001), who assert that expressive vocabulary is not significantly associated with repetition accuracy. To follow up on this finding, EVT was added to best subsets regression analysis for this group. This procedure determined that despite significant correlations between both expressive and receptive vocabulary and NWR, only receptive vocabulary scores had a strong enough independent relationship with NWR to appear in the most powerful regression model. This is most likely due to the high degree of overlap between the EVT and PPVT scores.

Table 9.

Partial correlations for the low-language score group, controlling for age.

Measure	1	2	3	4	5	6	7	8	9	
1. SpeltP-2	1.000									
2. DDK Rate	-.061	1.000								
3. DDK % Correct	-.139	-.289	1.000							
4. GFTA	-.342	-.152	-.179	1.000						
5. SCAN-C	.180	.062	-.001	.305	1.000					
6. PPVT	.768*	-.250	-.192	-.167	-.163	1.000				
7. EVT	.721*	-.221	-.257	-.201	.250	.847**	1.000			
8. TOPEL	.377	.022	-.259	-.033	.110	.648*	.626*	1.000		
9. Digit Span	.238	-.277	.093	.226	-.129	.529	.449	.359	1.000	
10. NWR % Error	-.459	.149	.137	.368	.485	-.634*	-.634*	-.240	-.314	1.000

Note. Marked correlations are significant after applying Holm-Bonferroni correction.

Follow-up best subsets regression for the lower language score group showed that the strongest model contained 3 predictors: Age, GFTA-2 raw scores, and vocabulary scores. Direct testing of this regression model indicated that only vocabulary was a significant predictor of NWR performance in this subgroup. The F-statistic of overall significance indicated that this model was significant as a whole. The adjusted R-squared for this model was 51.64%, and the error of the model was 4.22. Coefficients for this model are shown in Table 10 below.

Table 10.

Coefficients for the 3-variable model predicting NWR performance in the Low-language group

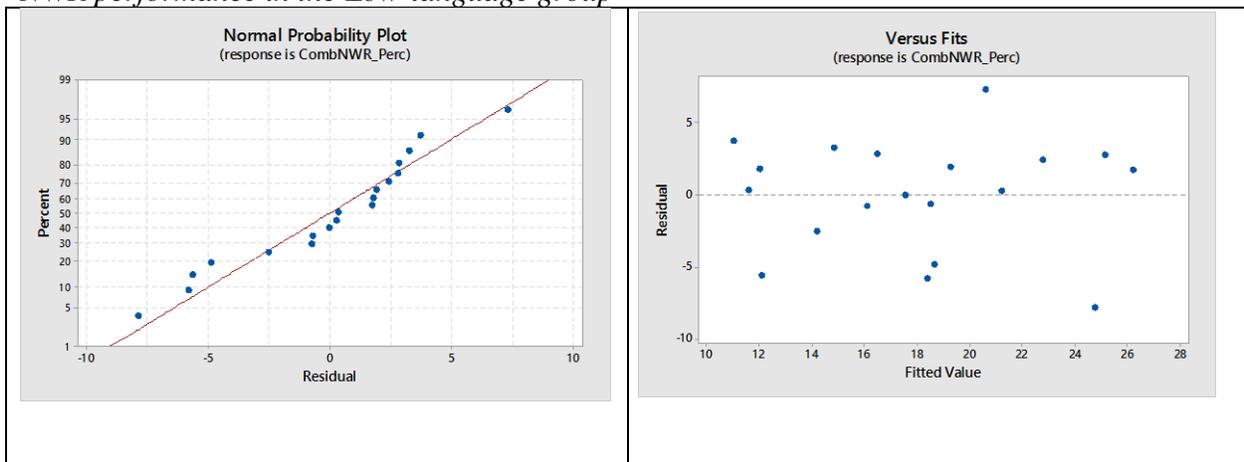
Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	29.98	9.49	3.16	0.006	
Age	-0.47	1.66	-0.28	0.782	1.55
GFTA	1.389	0.979	1.42	0.176	1.31
PPVT	-0.1319	0.0429	-3.08	0.008	1.40

In the interest of identifying the most parsimonious and effective combination of predictors, the insignificant variable Age was removed from the above model. This resulted in a slightly higher adjusted R-squared of 56.72%. The error of this new model was 3.99, however PPVT scores remained the only significant predictor of NWR performance. GFTA2 scores were near significant at $p = 0.70$. Examination of residuals showed a symmetrical distribution of residual values clustered under 10, and an approximately normal distribution (see figure 5). The coefficients for this model are shown in Table 11.

Table 11.
Coefficients for the 2-variable model predicting NWR performance in the Low-language group

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	28.77	4.74	6.07	0.000	
GFTA2	0.309	0.160	1.94	0.070	1.16
PPVT	-0.1334	0.0369	-3.62	0.002	1.16

Figure 5.
Normal probability plot and residuals-versus fits plot for the 2 variable model predicting NWR performance in the Low-language group



In the higher language scoring group, partial correlations controlling for age indicated that NWR performance was significantly related to digit span only. None of the other potential predictors were significantly correlated with overall NWR scores after the Holm-Bonferroni correction was applied. Correlation coefficients are displayed in Table 12.

Table 12.
Partial correlations for the high-language score group, controlling for age.

Measure	1	2	3	4	5	6	7	8	9	
1. SPELT P-2	1.000									
2. DDK Rate	.207	1.000								
3. DDK % Correct	-.207	-.242	1.000							
4. GFTA	-.057	.338	.039	1.000						
5. SCAN-C	-.472	.063	-.105	-.122	1.000					
6. PPVT	.520	.222	-.537	-.307	.001	1.000				
7. EVT	.000	-.014	.064	-.269	-.295	.345	1.000			
8. TOPEL	-.066	-.162	.390	-.395	-.138	.177	.759*	1.000		
9. Digit Span	-.124	-.243	-.021	-.243	.096	.081	-.112	.368	1.000	
10. NWR % Error	.162	-.057	.297	.162	-.357	-.324	.340	-.140	-.745*	1.000

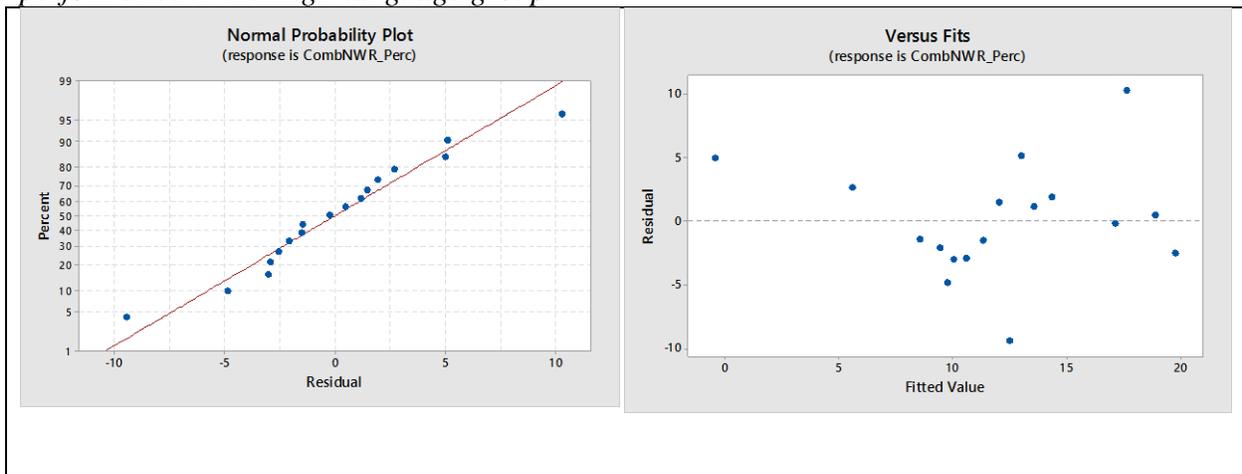
Note. Marked correlations are significant after applying Holm-Bonferroni correction.

Follow-up best subsets regression for this group indicated that the most effective model included three predictors: Age, DDK rate, and digit span. Direct testing of this model showed that age and digit span were significant predictors of NWR error. This model had an adjusted R-squared of 59.39%, a model error of 3.47, and a significant overall F-test. Coefficients for this model are shown in Table 13. Residuals were approximately normally distributed, and no particular pattern emerged on the residuals versus fits plot (see Figure 6).

Table 13.
Coefficients for the 3 variable model predicting NWR performance in the high-language score group

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	21.55	8.66	2.49	0.029	
Age	3.59	1.41	2.55	0.026	1.13
DDKRate	-2.81	1.69	-1.66	0.123	1.11
DigitSpan	-1.612	0.342	-4.72	0.001	1.06

Figure 6.
Normal probability plot and residuals-versus fits plot for the 3 variable model predicting NWR performance in the High-language group



In this case, removal of the non significant variable DDK rate reduced the adjusted R-squared of the model to 49.27%. In this more parsimonious model, digit span and age remained significant predictors.

Analysis 3: Do factors relating to NWR performance differ for nonword items of high and low wordlikeness?

For this set of analyses, performance on lists of nonword items with high and low phonological probability was measured for each participant. There was a significant difference in

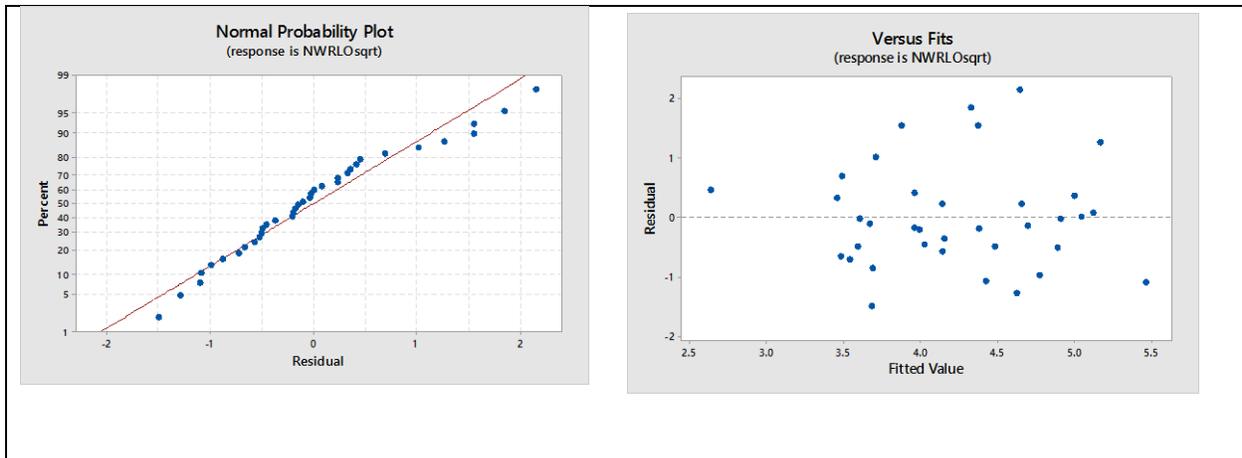
the percent error for nonwords that were high (M=.093, SD=.073) and low (M=.19, SD=.10) in phonological probability; $t(70)=4.703$, $p<0.001$.

Best subsets regression analysis of the entire sample’s performance on the **low** biphone probability list indicated that the optimal model would include three predictors: Age, digit span and PPVT scores. The overall F-test showed that this model was significant compared with an intercept-only model. The adjusted R-squared was relatively low at 27.03%, and the error of the model .921. However, only digit span was a significant predictor of NWR error on low wordlikeness items. Coefficients for this model are shown in Table 14 below.

Table 14.
Coefficients for predicting performance on low wordlikeness nonwords

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.28	1.19	4.44	0.000	
Age	0.363	0.246	1.48	0.150	1.27
DigitSpan	-0.1637	0.0643	-2.55	0.016	1.68
PPVT	-0.0107	0.00814	-1.32	0.198	1.57

Figure 7.
Normal Probability Plot and Residuals Versus Fits plots for the 3 variable model predicting performance on low wordlikeness nonwords.



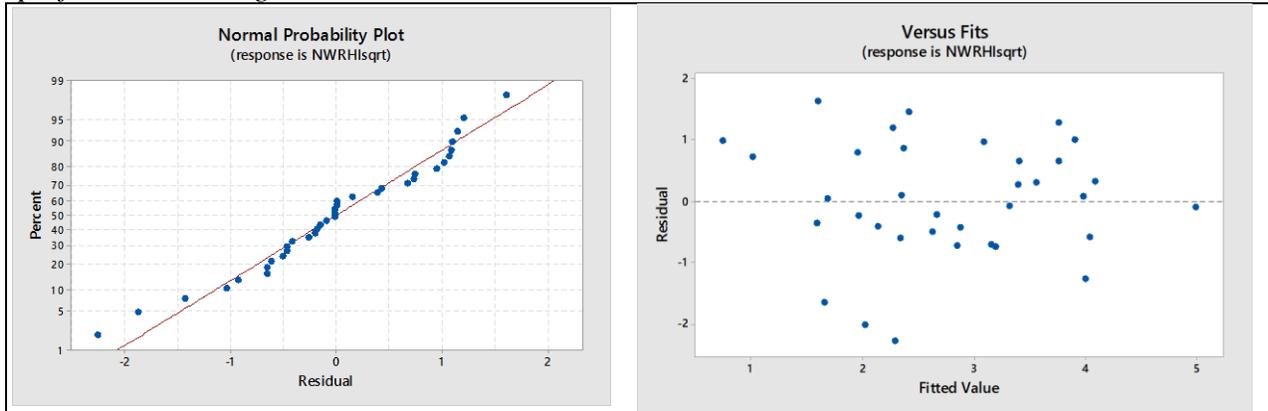
Removing the non significant predictor PPVT from the model reduced its fit slightly, lowering its adjusted R-squared to 25.42% and increasing model error to .931. In this more parsimonious model, digit span remained the only significant predictor.

For nonwords with a **high** biphone probability, best subsets regression revealed that the most likely model also included three predictors: DDK % correct, PPVT, and digit span. Unlike the previous analysis, only PPVT scores were significant within the model. DDK % correct and digit span scores were both close to significance at $p=0.069$ and $p=0.068$, respectively. The overall F statistic indicated that this model was significant, and the adjusted R-squared for the model was 52.53%. Residuals for this model were random, and approximately normally distributed (See Figure 8). Coefficients are displayed in Table 15.

Table 15.
Coefficients for predicting performance on highly wordlike nonwords

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.94	1.06	5.61	0.000	
DDK % correct	0.1452	0.0773	1.88	0.069	1.26
DigitSpan	-0.1170	0.0619	-1.89	0.068	1.53
PPVT	-0.02338	0.00883	-2.65	0.013	1.82

Figure 8.
Normal Probability Plot and Residuals Versus Fits plots for the 3 variable model predicting performance on high wordlikeness nonwords.



Removing the two non-significant terms from the model reduced its overall fit, decreasing the adjusted R-squared value to 45.72% (r-squared 47.27%).

Chapter 4: Discussion

4.1 Summary

The current study investigated the linguistic and cognitive factors that underlie and potentially constrain performance on NWR repetition in children. A regression approach was used to test the effects of multiple factors within the same sample. Unfortunately, missing data points and problems with collinearity of the variables made it necessary to remove SCAN-C scores and TOPEL-TAAS scores from the regression analysis, meaning speech perception and literacy skills could not be tested simultaneously with PSTM, vocabulary, and articulatory output factors. These scores were considered in initial partial correlations, however, allowing their potential relationships with NWR performance to be noted.

In Analysis 1, partial correlations using the entire sample indicated that several variables were significantly related to NWR performance overall. Measures of digit span, vocabulary, and

literacy were significantly related to NWR scores. Interestingly, both expressive and receptive vocabulary scores were significantly related to NWR, but this may be accounted for by their simultaneous strong correlation to each other.

In regression analysis, PPVT and digit span scores were the only significant predictors of overall NWR performance. Age was also a significant predictor in initial models suggested by best subsets regression, but lost significance as the model was refined.

The strength of the relationship between PSTM as measured by digit span and NWR scores is not surprising given the multiple previous findings of a similar link (Gathercole, Willis, Emslie, & Baddeley, 1994; Dollaghan & Campbell, 1998). In the current sample, this correlation was robust enough to remain significant even after the effects of all other variables were taken into account. The influence of vocabulary skill is also unsurprising, in light of the findings of Metsala (1999), which noted that children with lower vocabulary skills tend to have poorer NWR scores.

Overall, the results of Analysis 1 indicate that both long term lexical knowledge and short term phonological storage contribute independently to children's performance on NWR tasks. These findings coincide with the conclusions of Gupta and Tisdale (2009), which showed that in computational models, separate contributions of PSTM and vocabulary significantly influence NWR. Their simulation produced a cyclical relationship between vocabulary and PSTM, both of which affected NWR performance.

Gathercole (1995) suggests that the extent to which long term lexical knowledge affects NWR scores is mediated by the wordlikeness of the items in the task. The overall list of nonwords utilized in the current study did vary widely in phonotactic probability (See Appendix A for word lists with average biphone probability). Analysis of separate sublists constructed to

isolate high and low phonological probability nonwords did reveal that vocabulary skills were more strongly related to repetition of more wordlike items.

The results of Analysis 1 are relevant to the discussion of the overall nature of PSTM, specifically, to the contrast between the model of working memory espoused by Baddeley and colleagues (2000) and that put forward by MacDonald and Christiansen (2002). Recall that Baddeley's model presents the phonological loop as a rehearsal device separate from long term memory, yet influenced by LTM via the episodic buffer. MacDonald and Christiansen's conceptualization, however, incorporates PSTM as a function of overall language experience. Extending this difference to NWR, Baddeley's point of view claims that PSTM and vocabulary both influence NWR scores because the constructs are interrelated. MacDonald and Christiansen's model claims that they are essentially the same construct.

The results of the regression in analysis 1 show independent relationships between digit span and PPVT scores and NWR % error. That is, receptive vocabulary as measured by PPVT remains a significant predictor of NWR performance when the influence of digit span has been taken into account, and PSTM as measured by digit span also remains significant after accounting for PPVT scores. Digit span and PPVT were significantly related, according to partial correlations controlling for age, as would be expected from MacDonald and Christiansen's viewpoint. However their independent relationships with NWR in the regression model would seem to indicate that these measures are separable, rather than two overlapping constructions of the same concept.

Division of the sample using a median split in Analysis 2 had a decided effect on the underlying factors predictive of NWR performance. In the lower language score group, only vocabulary as measured by the PPVT predicted combined NWR percent error. In the higher

language score group, however, age and digit span were both significant predictors of NWR performance. Given the small n of each language score group, the results of the regression analyses should be interpreted with some caution. However, this pattern of findings was also replicated faithfully in partial correlations, which maximized statistical power by eliminating missing data pairwise. Taken together, the results of the correlation and regression analyses point to a pattern of relationships in which PSTM predicts change in NWR performance in children with higher language skills, while receptive vocabulary predicts change in NWR scores in those with lower language skills.

The results of analysis 2 indicate first that as language scores increase, NWR tasks may be supported by different underlying skills. This finding coincides with work in several different populations, suggesting that test taker skills and characteristics can affect the underlying skills related to NWR performance (Riches, Loucas, Baird, Charman, & Simonoff, 2011; Cairns and Jarrold 2005; Lee and Gorman 2012).

It is important to note that the lower language group also had significantly lower scores on digit span. It may be that children recruit alternate skills such as vocabulary to supplement lower PSTM. A similar pattern of results was found by Cairns and Jarrold (2005), in individuals with Down syndrome, noted to experience particular and specific deficits in verbal memory span. The authors interpreted this finding to indicate that, at least in this population, specific deficits in PSTM prevent this system from supporting nonword repetition. Without this support, the contribution of vocabulary knowledge was necessarily more influential on NWR scores in this group. Although the participants in the current study did not have Down syndrome, the application of this same principle would lead to the conclusion that those with lower PSTM

scores may rely more heavily on other factors to support their NWR performance. This effect may explain the observations made in Analysis 2.

The absence of an effect of articulatory output in the low language group is interesting, given that some evidence has been found of an increased articulatory complexity effect in children with SLI (Briscoe et al., 2001; Archibald & Gathercole, 2006). Briscoe et al. compared the performance of children with SLI and their typically developing peers on nonwords containing either single consonants or consonant clusters. Results indicated that the scores of the SLI group were more affected by articulatory complexity. Similarly, Archibald and Gathercole found that the presence of consonant clusters within nonwords differentially affected the NWR performance of children with SLI. The fact that this pattern of results was not replicated in the current study likely stems from sample characteristics. Specifically, only 5 of the 37 children in the current sample perform below cutoff scores for SLI; the rest of the lower language score group are typically developing. It is likely that the performance characteristics of this group do not reflect those of a true SLI sample.

Analysis 3 sought to differentiate underlying skills related to repetition of nonwords with varying degrees of wordlikeness. Specifically, scores on lists of high and low average biphone probability were isolated and analyzed for predictor variables. First, it is important to note that percent error was significantly higher for items of low wordlikeness, replicating the robust findings of Edwards, Beckman, and Munson (2004), Coady, Evans, and Kluender (2010), and others.

Regression analysis found a difference in significant predictor variables between the high- and low-wordlikeness lists. PSTM as measured by digit span was the only significant predictor of performance on low wordlikeness items. On high wordlikeness items, however,

PPVT scores were the only significant predictor of performance, although both digit span and DDK % correct were near significance.

These findings partially contrast with those of Gathercole (1995), who found that vocabulary skills affected performance on both high and low wordlikeness stimuli, but PSTM scores were more closely linked to high wordlikeness stimuli. One potential explanation of this difference is a contrasting operational definition of wordlikeness. Gathercole's word lists were compiled using a scale of adult wordlikeness judgment, rather than phonological probability or neighborhood density. The biphone probability mean used in the current study is a more objective measure. Further, both the high- and low-wordlikeness items in the Gathercole study were taken from the CNRep, a NWR task with relatively wordlike items. Thirteen of the twenty items on the low-wordlikeness list contained real English morphs. For example, "penneriful", "diller", and "blonderstaping" were considered low in wordlikeness. In the current study, however, only one of the low biphone probability nonwords contained a real English morph.

While phonotactic probability measures were not reported for Gathercole's (1995) nonword lists, it is reasonable to assume that the current low-wordlikeness items are much less typical than those used in the previous study. The effect of using such a very low wordlikeness list may have been to negate any possibility of vocabulary skill being utilized to support repetition of these nonwords. Therefore, the children in the current study instead supported their repetition of low wordlikeness items with PSTM alone.

Predictors of performance on the high-probability word list were also unexpected given Gathercole's (1995) results. In the present study, PPVT scores were the only significant predictor of performance on high-wordlikeness items. It is important to note, however that both digit span and DDK % correct were near significance in this model, and therefore the effects seen in

Gathercole's experiment may only have been suppressed by the current study's smaller sample size.

4.2 Implications

NWR tasks are consistently and increasingly popular in language research as a measure of phonological short term memory. The advantages of these tasks are many. They are relatively fast, engaging, and adaptable to multiple age groups, populations, and languages. These same advantages make NWR tasks popular in clinical settings, where they are used by reading specialists, speech-language pathologists, and school psychologists. However, as the results of the current project show, there is a danger in treating these tasks as monolithic measures of PSTM, and further danger in treating different NWR tasks as interchangeable.

The current results join the confluence of evidence that skills other than PSTM can and do affect performance on NWR tasks. For researchers wishing to utilize NWR to investigate child language and cognition, the implications of the current results are clear. NWR tasks cannot be regarded as monolithic tests of PSTM in all cases, and caution should be used when including NWR as a memory measure. Analysis 1 suggests that overall, both PSTM and vocabulary influence the NWR scores of children at this age. Investigators must take this into account when using NWR, and consider the implications of this influence for their findings. If a specific test of PSTM is needed, careful consideration of participant and nonword characteristics should be made.

The results of the current study suggest that influences on NWR performance vary according to the language skill of the participants, and the phonological probability of the nonword items. In children with higher language skill, PSTM as measured by digit span was the only candidate predictor significantly related to NWR, but for children with lower language

scores, long-term lexical knowledge as measured by PPVT was the sole significant predictor of NWR performance. Investigators should be aware of the language characteristics of their sample population, and the difference that might make to the underlying skills measured by NWR.

Further, the wordlikeness of nonword items influences the skills related to NWR performance. These effects have been investigated in the past (e.g. Gathercole, 1995), and the current results underscore the need to carefully evaluate nonwords for wordlikeness when utilizing NWR. In analysis 3, repetition of nonwords with low phonological probability was related to PSTM alone. However, repetition performance on nonwords with high phonological probability was influenced by long term lexical knowledge to such an extent that PPVT scores were the only significant predictor of NWR scores. Consequently, investigators intending to use NWR as a measure of PSTM would be better served by a nonword list low in wordlikeness.

Clinically, similar care must be taken when using NWR tasks as part of language evaluations, or when taking these measures into account as part of treatment planning. Although NWR tasks have shown promise as identifiers of language impairment in past investigations, it cannot be assumed that the low NWR scores common in children with language impairments are solely indicative of low PSTM capacity. In analysis 2, different predictors are related to NWR in children with different levels of language skill. If this pattern persists in the larger population of disordered children, it has serious implications for the interpretation of NWR tasks. NWR scores in the lower language scoring group were significantly predicted by receptive vocabulary as indexed by PPVT scores.

The current interpretation of this finding is that the lower language scoring group, which also had lower PSTM, relied upon their vocabulary skills to support a task for which their memory capacity was insufficient. In this view, NWR tasks are only an indirect measure of

PSTM for this population. Their low scores on NWR may be due to the fact that their PSTM is not sufficient to support the temporary retention and repetition of nonwords. However, clinicians should be aware that low NWR performance in this population is also reflective of receptive vocabulary.

Archibald (2008) suggests that NWR tasks have several advantages over traditional language tests, the first of which is the reliance of traditional tests on prior knowledge and vocabulary. According to the current results, vocabulary skill is related to NWR scores, particularly in children with lower language skills. While NWR does test children's ability to process new information, as stated by Archibald, prior knowledge in the form of receptive vocabulary skill, does influence that processing.

Delays in vocabulary acquisition have been noted in children with SLI (e.g. Hick, Conti-Ramsden, Serratrice, Faragher, 2002) However, in an investigation of four widely used vocabulary measures including the PPVT, Gray and colleagues (1999) found that vocabulary scores alone were not good identifiers of language impairment. In contrast to Archibald's (2008) assertion that NWR is advantageous because of its independence from vocabulary skill, it may be that the combination of vocabulary and PSTM tested by NWR is what makes it a powerful tool for detection of language impairments.

NWR tasks show great promise as tools in the clinic and as measures for research purposes. However, they are by no means simple measures. When utilizing NWR, care must be taken to ensure that the specific characteristics of nonword items are valid for the purposes for their user, and that their use in a given population will yield the information sought.

4.3 Limitations and Future Directions

This series of studies represents a largely exploratory examination of the underlying skills that influence NWR in children. Its potential weaknesses should be kept in mind by researchers making further attempts to pursue the same questions. First, a relatively small n in this study reduced its explanatory power and limited the statistical techniques that could be brought to bear on the sample. In future studies, a similar battery of tests on a larger group of participants might allow for more powerful analyses such as principle components analysis to further explore relationships among contributing skills.

A second limitation presents itself when one considers the overrepresentation of high SES and high language skill participants within the sample. It is likely that the time commitment involved with completing the test battery – at least three visits, often with an additional visit if the child was easily fatigued or distractible – discouraged the participation of families with multiple demands on their time. In addition, coordination of travel to a University clinic may have precluded participation of families without access to reliable transportation. Arranging for more community-based testing locations is one potential solution to this sampling issue.

The time commitment involved in a testing battery of this size is also a likely deterrent for families with children who have language problems. The demands of therapy appointments, as well as school and family activities may have made the additional commitment to participation burdensome. Streamlining the battery, as well as increasing community and/or school based testing, might increase the number of participants with lower language skills in future studies.

Missing data points were also a potential weakness of the present study, particularly because they were not randomly distributed among participant cases. The fact that children with lower language skills were more likely to refuse tasks such as the DDK and SCAN-C subtests should be taken into account in future studies. Replacing these tasks with less demanding or more child-friendly versions could potentially alleviate this problem. A larger N, including more children with lower language skills, would also be more robust to missing data.

4.4 Conclusion

The current results contribute to the growing body of evidence that NWR tasks are not monolithic measures of phonological loop function, but are instead influenced by multiple underlying factors, which may change depending on the characteristics of both nonword items and test takers. The present study extends the existing research on NWR in several ways. First, it used a regression approach to examine multiple potential predictive factors within the same sample. Although several potential predictor variables have been suggested to influence NWR, few have been tested simultaneously. Testing multiple potential predictors in the same sample allows for the identification of the most impactful variables. In this sample, PSTM and vocabulary skills were related to NWR performance, above and beyond the influence of GFTA and DDK. Although speech perception as measured by the SCAN-C could not be included in the regression analysis, these scores were not related to NWR according to partial correlations. Second, this study extends previous knowledge of NWR investigated whether differences in wordlikeness and test-taker language skill alter the skills related to NWR performance. While other studies have investigated such differences indirectly (Archibald & Gathercole, 2006) The present study systematically tested each proposed underlying factor and analyzed their influence

on NWR separately in children high and low in language skill, as well as their influence on performance on wordlike and unwordlike NWR items. In the current sample there were differences in the predictors of NWR in the different language groups, as well as differences in predictors for scores on differing nonword lists.

The results of the current project have some bearing on theoretical differences in the literature. Specifically, independent contributions of vocabulary and PSTM seem to indicate that these constructs are separable, though strongly inter-related, supporting the view of Baddeley and colleagues (2000) and Gupta and Tisdale (2009). Moreover, the current results have important implications for users of NWR tasks for research and clinical purposes. It is imperative that NWR tasks not be treated as measures of PSTM as a matter of course. The characteristics of the nonword items themselves, as well as those of the test taker, must be accounted for. Careful selection of nonword items/tasks, while keeping in mind the population to be tested, can maximize the clinical and research utility of this measure.

Appendix A

Nonword Items for CTOPP NR and CNRep Tasks

CTOPP NR		
Test Item	IPA	Mean biphone Probability
ral	ræl	0.0047
sart	sart	0.0098
ballop	bæloʊp	0.0039
jup	jup	0.0019
zid	zɪd	0.0024
pate	pɛt	0.0031
meb	mɛb	0.0029
wudoip	wudoɪp	0.0006
nigong	nɑɪgɑŋ	0.0009
chaseedoolid	tʃæsɪdʊlɪd	0.0016
bieleedoge	bɑɪlɪdoʊdʒ	0.0028
voesutoov	vʊʊzətʊv	0.0012
lisashrul	lɪsəʃrəl	0.0026
wulanuwup	wulænəwəp	0.0007
teebudieshawlt	tɪbədɑɪʃɑlt	0.0013
viversoomouj	vɪvəsumɑʊdʒ	0.0015
burloogugendaplo	bəlʊgədʒɛndɑplʊ	0.0006
gekieziesae kad	gɛkɪzɪsɛɪkæd	0.0019
mawgeebooshernoosiek	mɑgɪbʊʃənuʃɪk	0.0011
dookershatupietazawm	dʊkəʃætəpɑɪtəzɑm	0.0007
shaburiehuvoimush	ʃæbəɑɪhəvɔɪmʊʃ	0.0005

CNRep		
Test Item	IPA	Mean biphone Probability
diller	dɪlə	0.0091
hampent	hæmpɪnt	0.0053
rubid	rubɪd	0.0012
bannow	bæno	0.0085
sladding	slædɪŋ	0.0057
bannifer	bænɪfə	0.0054
trumpetine	trʌmpɪtɪn	0.0031
glistering	glɪstə-rɪŋ	0.0047
thickery	θɪkəri	0.0048
doppelate	dʌpələt	0.0026
stopograttic	stɒpəgrætɪk	0.0035
contramponist	kəntɹæmpənɪst	0.0033
woogalamic	wuɡələmɪk	0.002
loddenapish	lɒdenæpɪʃ	0.002
pennerriful	pənəəfʊl	0.0032
altupatory	æltʊpɪtɔri	0.0019
pristoractional	pɹɪstəækʃənəl	0.0024
reutterpation	ruðəpeɪʃən	0.0027
confrantually	kənfræntʃuəli	0.001
varsatrationist	vəsətɹæʃənɪst	0.0013

References:

- Archibald, L. M. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology*, 32(1), 21-28.
- Archibald & Gathercole, (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research : JSLHR*, 49(5), 970-983.
- Archibald, L. M., & Gathercole, S. E. (2006). Short- term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, 41(6), 675-693.
- Archibald, L.M.D., Gathercole, S.E. (2007) Nonword repetition in specific language impairment: More than a phonological short-term memory deficit *Psychonomic Bulletin and Review*, 14 (5), pp. 919-924.
- Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52(4), 899-914.
- Archibald, L. M., Joanisse, M. F., & Munson, B. (2013). Motor control and nonword repetition in specific working memory impairment and SLI. *Topics in language disorders*, 33(3), 255.
- Baddeley, A. D. (1986). *Working memory* (Vol. 11). Oxford: Clarendon Press.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, 4(11), 417-423.
- Baddeley, A.D., Hitch, G. (1974) *Working Memory Psychology of Learning and Motivation - Advances in Research and Theory*, 8 (C), pp. 47-89.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173.
- Baird, G., Slonims, V., Simonoff, E., & Dworzynski, K. (2011). Impairment in non- word repetition: a marker for language impairment or reading impairment?. *Developmental Medicine & Child Neurology*, 53(8), 711-716.
- Bishop, D. V., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short- term memory deficits: evidence from 6- year- old twins. *Genes, Brain and Behavior*, 5(2), 158-169.
- Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, 37, 391–403.
- Bishop, D.V.M., Bishop, S.J., Bright, P., James, C., Delaney, T., Tallal, P. (1999) Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study *Journal of Speech, Language, and Hearing Research*, 42 (1), pp. 155-168.
- Bishop, D. V. M., Carlyon, R. P., Deeks, J. M., & Bishop, S. J. (1999). Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children. *Journal of Speech, Language, and Hearing Research*, 42(6), 1295-1310.
- Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, 36, 421–432.

- Brady, S., Poggie, E., & Rapala, M. M. (1989). Speech repetition abilities in children who differ in reading skill. *Language and Speech*, 32(2), 109-122.
- Briscoe, J., Bishop, D. V., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(3), 329-340.
- Burgemeister BB, Blum LH, Lorge I. *The Columbia Mental Maturity Scale*. San Antonio, TX: Psychological Corporation; 1972.
- Cairns, P., & Jarrold, C. (2005). Exploring the Correlates of Impaired Non-Word Repetition in Down Syndrome. *British Journal of Developmental Psychology*, 23(3), 401-416.
- Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech, Language, and Hearing Research*, 48(6), 1378-1396.
- Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental language disorders—a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, 46(2), 128-149.
- Coady, J.A., Evans, J.L. (2008) Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders*, 43 (1), 1-40.
- Coady, J., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders*, 45(4), 494-509.
- Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they? *Memory and Cognition*, 28(8), 1269-1276.
- Dawes, P., & Bishop, D.V.M. (2007) The SCAN-C in testing for auditory processing disorder in a sample of British children. *International Journal of Audiology*, 46 (12), 780-786.
- Dawson J, Stout C, Eyer J, Tattersall P, Fonkalsrud J, and Croley K (2005) Structured photographic expressive language test: Preschool. DeKalb, IL: Janelle.
- Deevy, P., Weil, L. W., Leonard, L. B., & Goffman, L. (2010). Extending use of the NRT to preschool-age children with and without specific language impairment. *Language, Speech, and Hearing Services in Schools*, 41(3), 277-288.
- Dillon, C. M., Burkholder, R. A., Cleary, M., & Pisoni, D. B. (2004). Nonword repetition by children with cochlear implants: Accuracy ratings from normal-hearing listeners. *Journal of Speech, Language, and Hearing Research*, 47(5), 1103-1116.
- Dillon, C.M., Cleary, M., Pisoni, D.B., Carter, A.K. (2004) Imitation of nonwords by hearing-impaired children with cochlear implants: Segmental analyses *Clinical Linguistics and Phonetics*, 18 (1), pp. 39-55.
- Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech, Language, and Hearing Research*, 36(5), 1051-1054.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136-1146.
- Dromey, C., & Benson, A. (2003). Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance. *Journal of Speech, Language, and Hearing Research*, 46(5), 1234-1246.

- Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test: PPVT-4B. NCS Pearson.
- Durkin, K., & Conti-Ramsden, G. (2007). Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. *Child development*, 78(5), 1441-1457.
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421.
- Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 297-339.
- Ellis Weismer, S. et al. "Nonword repetition performance in school-age children with and without language impairment." *Journal of Speech, Language, and Hearing Research* 43.4 (2000): 865-878.
- 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(February), 177--195.
- Flax, J. F., Realpe-Bonilla, T., Hirsch, L. S., Brzustowicz, L. M., Bartlett, C. W., & Tallal, P. (2003). Specific language impairment in families: Evidence for co-occurrence with reading impairments. *Journal of Speech, Language, and Hearing Research*, 46(3), 530-543.
- Flynn J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin* 101: 171–191.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of memory and language*, 42(4), 481-496.
- Gaetano J. (2013). Holm-Bonferroni sequential correction: An EXCEL calculator (1.1) [Microsoft Excel workbook]. Retrieved from https://www.researchgate.net/publication/236969037_Holm-Bonferroni_Sequential_Correction_An_EXCEL_Calculator . doi:10.13140/RG.2.1.4466.9927
- Gathercole, S.E. (1995) Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition*, 23 (1), pp. 83-94.
- Gathercole, S.E. (2006) Nonword repetition and word learning: The nature of the relationship *Applied Psycholinguistics*, 27 (4), pp. 513-543.
- Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of memory and language*, 33(5), 672-688.
- 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 and language*, 28(2), 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(3), 336-360.
- Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, 81(4), 439-454.
- Gathercole, S. E., Hitch, G. J., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental psychology*, 33(6), 966.

- 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(2), 103-127.
- Goffman, L. (2004) Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research*, 47 (5), pp. 1088-1102.
- Goldman, R. (2000). Frisloe M. Goldman-Frisloe test of articulation, 2nd Edition.
- Gray, S. (2003) Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, 36 (2), pp. 129-151.
- Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools*, 30(2), 196-206.
- Greenslade, K.J., Plante, E., Vance, R. (2009) The diagnostic accuracy and construct validity of the structured photographic expressive language test-preschool: Second edition. *Language, Speech, and Hearing Services in Schools*, 40 (2), pp. 150-160.
- Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, 61(4), 481-502.
- Henry, L. A., & Millar, S. (1991). Memory span increase with age: A test of two hypotheses. *Journal of Experimental Child Psychology*, 51(3), 459-484.
- Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of communication disorders*, 38(3), 197-213.
- Hick R., Joseph K., Conti-Ramsden G., Serratrice L., Faragher B. (2002) Vocabulary profiles of children with specific language impairment. *Child Language Teaching and Therapy* 18: 165–180.
- Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: a deficit in grammar or processing?. *Trends in cognitive sciences*, 2(7), 240-247.
- Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and syntax in specific language impairment: Evidence from a connectionist model. *Brain and language*, 86(1), 40-56.
- Joanes Maner, K., Smith, A., & Grayson, L. (2000). Influences of utterance length and complexity on speech motor performance in children and adults. *Journal of Speech, Language, and Hearing Research*, 43(2), 560-573.
- Jusczyk, P. W., & Luce, P. A. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33(5), 630-645.
- Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53(3), 739-756.
- Keith, R.W. (2000) Development and standardization of SCAN-C test for auditory processing disorders in children. *Journal of the American Academy of Audiology*, 11 (8), pp. 438-445.
- Keith, R. W. (2000). SCAN C: Test for auditory processing disorders in children. Psychological Corporation.

- Kleinow, J., & Smith, A. (2006). Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental psychobiology*, 48(4), 275-287.
- Krishnan, S., Alcock, K.J., Mercure, E., Leech, R., Barker, E., Karmiloff-Smith, A., Dick, F. (2013) Articulating novel words: Children's oromotor skills predict nonword repetition abilities. *Journal of Speech, Language, and Hearing Research*, 56 (6), pp. 1800-1812.
- Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers?. *Journal of Speech, Language, and Hearing Research*, 39(5), 1081-1098.
- Lee, S., & Gorman, B. (2012) Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism*, 17(4), 479-495.
- Leonard, L. B. (2014). *Children with specific language impairment*. MIT press.
- Lennon, J. & Slesinski, C. (2002) *Comprehensive Test of Phonological Processing (CTOPP): Cognitive-linguistic assessment of severe reading problems*.
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *TOPEL: Test of preschool early literacy*. Pro-Ed.
- Lum, J. A., & Zarafa, M. (2010). Relationship between verbal working memory and the SCAN-C in children with specific language impairment. *Language, Speech, and Hearing Services in Schools*, 41(4), 521-530.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35-54.
- Masoura, E. V., & Gathercole, S. E. (1999). Phonological short- term memory and foreign language learning. *International Journal of Psychology*, 34(5-6), 383-388.
- Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological science*, 23(10), 1092-1098.
- Metsala, J.L. (1999) Young children's phonological awareness and nonword repetition as a function of vocabulary development *Journal of Educational Psychology*, 91 (1), pp. 3-19.
- Metsala, J.L., Chisholm, G.M. (2010) The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics*, 31 (3), pp. 489-506.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability.
- Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech, Language, and Hearing Research*, 38(1), 187-199.
- 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.
- Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 48(5), 1033-1047.

- Nation, K., Hulme, C. (2011) Learning to read changes children's phonological skills: Evidence from a latent variable longitudinal study of reading and nonword repetition. *Developmental Science*, 14 (4), pp. 649-659.
- Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language and Speech*, 48(2), 123-156.
- Petersson, K. M., Reis, A., Askelöf, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *Journal of cognitive neuroscience*, 12(3), 364-382.
- Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, 39(6), 1239-1257.
- Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2011). Non-word repetition in adolescents with specific language impairment and autism plus language impairments: A qualitative analysis. *Journal of communication disorders*, 44(1), 23-36.
- Rispens, J., & Parigger, E. (2010). Non- word repetition in Dutch- speaking children with specific language impairment with and without reading problems. *British journal of developmental psychology*, 28(1), 177-188.
- Roberts, J., Rescorla, L., Giroux, J., & Stevens, L. (1998). Phonological skills of children with specific expressive language impairment (SLI-E): Outcome at age 3. *Journal of Speech, Language, and Hearing Research*, 41(2), 374-384.
- Rodekohr, R. K., & Haynes, W. O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders*, 34(3), 255-272.
- Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything?. *Journal of phonetics*, 31(3), 509-527.
- Rosner, J. (1993). *Helping Children Overcome Learning Difficulties*, 3rd Ed. Walker and Company: NY.
- Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2-to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, 47(1), 223-234.
- Snowling, M., Chiat, S., & Hulme, C. (1991) Words, Nonwords, and Phonological Processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Linguistics*, 12 (3), pp 369-373.
- Stark, R.E., Blackwell, P.B. (1997) Oral volitional movements in children with language impairments *Child Neuropsychology*, 3 (2), pp. 81-97.
- Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, 44(6), 1321.
- 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(2), 497-506. <http://wordlearning.ku.edu/child-calculator>
- Stothard, S. E., Snowling, M. J., Bishop, D. V., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research*, 41(2), 407-418.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research*, 40(6), 1245-1260.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*.

Trojano, L., & Grossi, D. (1995). Phonological and lexical coding in verbal short-term-memory and learning. *Brain and language*, 51(2), 336-354.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and language*, 68(1), 306-311.

Williams, K. T. (1997). *Expressive vocabulary test*. American Guidance Service.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing: CTOPP*. Pro-Ed.

Zelaznik, H. N., & Goffman, L. (2010). Generalized motor abilities and timing behavior in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 53(2), 383-393.

