Phonological and Working Memory and L2 Grammar Learning

by

Katherine I. Martin

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of Bachelor of Arts

With Honors in Brain Behavior and Cognitive Science from the

University of Michigan

2009

Advisor: Dr. Nick C. Ellis

Abstract

The current study analyzed phonological short-term memory (PSTM) and working memory (WM) and their relationship with vocabulary and grammar learning in an artificial foreign language. Non-word Repetition, Non-word Recognition, and Listening Span were used as memory measures. Participants learned the singular forms of vocabulary for an artificial foreign language and were introduced to plural forms in a sentence context. They were tested on their ability to generalize these grammatical forms to novel utterances. Results showed moderate correlations between both PSTM and WM and the vocabulary measures. In addition, PSTM and WM both correlated moderately with grammar scores. However, regression analyses demonstrated that a large portion of the relationship between grammar and memory measures was mediated by vocabulary knowledge. The relatively equal contributions of WM and PSTM to both vocabulary and grammar scores are considered within the context of the single-mechanism acquisition framework (Bates & Goodman, 1997).

Phonological and Working Memory and L2 Grammar Learning

It is common for adults to have differing success when attempting to learn a second language, and there has been a large amount of research investigating the various factors that underlie language learning success. Individual differences in many domains, including motivation (Fan, 2003; Sanaoui, 1995), age (Birdsong & Molis, 2001; Johnson & Newport, 1989), working memory capacity (Harrington & Sawyer, 1992; Sunderman & Kroll, 2009), and phonological short-term memory capacity (Gathercole & Baddeley, 1993; Service, 1992), have all been proposed as reasons for this differential success. This research has primarily focused on vocabulary learning, reading comprehension, and fluency development, and the results appear to be similar for both first and second languages. The acquisition of grammatical structures, however, has been studied much less extensively. The current study, therefore, intends to address the influences of working memory and phonological short-term memory on the acquisition of both vocabulary and grammar in a second language.

A Model of Working Memory

Working memory is the ability to mentally store and manipulate information (usually auditory, visual, or spatial) (Baddeley, 1998, 2003). Although two broad approaches to WM are used in the literature (Baddeley, 1998; van den Noort, Bosch, & Hugdahl, 2006), the distinction between them is not always made clear. The terms working memory, complex working memory, simple working memory, short-term memory, phonological short-term memory, and simple short-term memory are often used differently or interchangeably by various authors.

The “British” approach to working memory focuses on the storage of a specific type of information, such as auditory or visual input. This approach has been emphasized by a number of influential researchers, including Alan Baddeley and Susan Gathercole. The “American” approach to working memory emphasizes not only the storage of information, but also its processing and manipulation (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992). In this paper, "working memory" (WM) will refer to the American approach, defined as both storage and processing of information. Reading span, listening span, and some speech generation tasks have been used to measure this type of memory (Fortkamp, 1999; van den Noort et al., 2006). "Phonological short-term memory" (PSTM) will refer to storage alone, which is the approach used primarily by the British. Tasks such as non-word repetition, serial recall, non-word recognition, and digit span have commonly been used to measure working memory defined in this way (van den Noort et al., 2006). In addition, although a distinction is often made between language learning and language acquisition (see for example Krashen, 1981), the two terms will be used interchangeably in this paper.

Working memory is one of the individual difference factors that has been investigated most extensively. Baddeley's model of WM (Baddeley, 1998, 2003; Baddeley & Hitch, 1974) is one of the most influential models; a depiction can be found in Figure 1. The basic model is comprised of three parts: the central executive, the phonological loop, and the visuo-spatial sketchpad. The central executive is the task control center and is in charge of directing attentional processes and allocating cognitive resources. The phonological loop and the visuo-spatial sketchpad are considered slave systems which store the auditory and visuo-spatial information used by the central executive (Baddeley, 1992). Additionally, a third sub-component, the episodic buffer, has been recently proposed (Baddeley, 2003). It is responsible for combining the auditory and the visual codes and integrating them with representations from long-term memory.

The phonological loop is the most extensively studied component of Baddeley's working memory model, and is comprised of two additional subsystems: a phonological store and a sub-vocal rehearsal process (Baddley, 1992). The phonological store maintains representations of auditory information for approximately 1.5 seconds (Baddeley, Thomson, & Buchanan, 1975), after which they begin to decay. The sub-vocal rehearsal process is capable of rehearsing and refreshing the information in the phonological loop, allowing it to be maintained beyond the base capacity of the system. The visuo-spatial sketchpad has been studied less than the phonological loop, but is believed to be responsible for maintaining mental representations of visually presented information, such as patterns and spatial relationships (Farah, 1988; Farah, Hammond, Levine, & Calvanio, 1988).

The central executive and the phonological loop have both been implicated in various aspects of language learning, with more focus on the phonological loop. Baddley (2003; Baddeley, Gathercole, & Papagno, 1998) has proposed that the function of the phonological loop is to support language acquisition. It is associated with PSTM and has been implicated in a number of aspects of language acquisition, including vocabulary development in both the first language (L1) and second language (L2), fluency, and some measures of comprehension (e.g., French, 2006; Service, 1992). The central executive is associated with working memory and seems to be related to reading comprehension and global verbal abilities (e.g., Daneman & Carpenter, 1980; Turner & Engle, 1989). Harrington and Sawyer (1992) and Berquist (1997) have concluded that it also accounts for individual differences in the efficiency of language processing. These findings will be reviewed in further detail below.

Working Memory and First Language Skills

Working memory has been strongly implicated as a predictor of L1 comprehension, reading ability, and traditional measures of verbal abilities, including the Verbal portion of the Scholastic Aptitude Test (VSAT) (Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Friedman & Miyake; 2004; Marton & Schwartz, 2003; Masson & Miller, 1983; Turner & Engle, 1989). A large portion of the literature focusing on working memory began with Daneman and Carpenter's (1980) development of the Reading Span Test and the Listening Span Test. In these tests, the participant either reads or hears a sentence and must decide whether it is grammatically sound. Sentences are presented in sets of increasing length, usually beginning with two or three and continuing until five to seven sentences long. At the end of each set of sentences, participants must recall the last word of each sentence in that set. It is considered a test of working memory because it requires both storage of the individual items and processing of the sentence content (Daneman & Carpenter, 1980; van den Noort et al., 2006). Daneman and Carpenter (1980) found that their Reading Span task correlated highly with measures of reading comprehension, with correlations as high as r = .70-.90. It also correlated with self-reported Verbal SAT scores (r = .59). They went on to demonstrate that their auditory version, Listening Span, also correlated highly with the same measures (*r*s between .49 and .86). Each task also correlated with both listening and reading comprehension. Friedman and Miyake (2004) also found that the Reading and Listening Span tasks correlated with reading comprehension (r = .55) and VSAT scores (r = .49). In contrast, simple word span (which requires storage, but not processing) only correlated with reading comprehension and VSAT at r = .45. This result supports Daneman and Merikle's (1996) meta-analysis finding that complex working memory tasks (involving both processing and storage) correlate more strongly with comprehension than do simple short-term memory tasks (involving storage only).

Reading Span also correlates with other complex working memory tasks, such as letter-number ordering (van den Noort et al., 2006). These complex span tasks require participants to mentally re-order strings of letters, digits, or words based on predetermined guidelines, or to perform mental operations, such as addition or subtraction (operations span). They are considered measures of WM because they require both processing and storage. Turner and Engle (1989) found that both complex operation span and complex word span correlated with reading comprehension, Quantitative SAT (QSAT), and VSAT scores. Since complex operation spans predicted verbal test scores, they concluded that the complex span tasks were not task-specific.

Marton and Schwartz (2003) tested children with normal language development and children with Specific-Language Impairment (SLI) on their working memory abilities. Specific Language Impairment is a developmental language disorder that is characterized by delayed vocabulary development and difficulty with complex syntactic structures (Cohen, 2002). It occurs without any obvious contributing factors, such as physical disability, or hearing, emotional, or social difficulties (Stark & Tallal, 1981; Watkins, 1994). Marton and Schwartz (2003) found that children with SLI performed significantly worse on Listening Span-type tasks than did those with normal language development, indicating that they had limited WM capacities. From the pattern of errors across a variety of different tasks, including reduced primacy and recency effects, the researchers concluded that the children with SLI had problems with processing stimuli rather than storing their structures. They also concluded that WM is important for comprehension during language learning because it assists the learner in segmenting and analyzing linguistic structures. Together, these studies demonstrate the importance of working memory for overall L1 verbal abilities.

Working Memory and Second Language Development

Working memory has also been implicated in comprehension, reading, and fluency in second language (L2) development. Besides the Reading and Listening Span tests, another relatively common WM test is backwards digit span. In this task, participants hear a string of digits and must repeat them back in the order opposite that in which they were presented. For example, a participant may hear “4, 7, 8, 1” and must produce “1, 8, 7, 4”. This task requires both remembering the individual items as well as manipulating their order of recall. Kormos and Sáfár (2008) found that backwards digit span correlated with five out of their six measures of L2 (English) ability, including reading (r = .31), listening (r = .37), speaking (r = .33), use of English (r = .47), and total proficiency (r = .55). They argue that this finding demonstrates a direct connection between WM and vocabulary and grammar acquisition that is based on WM’s ability to regulate attention. This function is important for noticing and encoding new information, and may be a basic mechanism for learning new patterns and rules in an L2.

Harrington and Sawyer (1992) developed an L2 Reading Span test based on Daneman and Carpenter (1980), simplified for use with non-native English speakers. They found that this Reading Span task correlated with both L2 reading scores and L2 grammar scores. Fortkamp (1999) replicated these results and also developed a new WM task: Speaking Span. In this task, participants were briefly presented with a list of words and then had to generate one sentence that incorporated each word. Scores on this task also correlated with scores of L2 fluency (rs between .61 and .64).

Finally, Sunderman and Kroll (2009) used a Reading Span test to investigate the working memory capacities of college students who studied abroad and those who did not. WM scores correlated significantly with performance on a translation comprehension task. The researchers calculated that a 10-unit increase in working memory capacity led to an 80 ms increase in comprehension processing speed. However, they also found that beneath a minimum WM capacity threshold, the students who studied abroad were not able to benefit from that experience. The authors concluded that WM is an important internal resource, necessary for the individual to benefit from studying abroad. Those individuals with higher WM may be able to attend to more linguistic factors at once, thereby increasing their ability to parse grammatical structures. Overall, the research in L2 studies complements the first-language research in demonstrating the important role of working memory in comprehension and reading abilities, as well as language skills more generally.

Phonological Short-Term Memory

Compared to working memory, phonological short-term memory is a simpler component of memory in that it considers only storage capacity and ignores processing efficiency. Two main types of tasks are used to asses PSTM: word and non-word span tasks. Word span tasks consist of hearing a list of words in a specific language and immediately repeating them back in the same order. Letters, digits, and concrete nouns are commonly used as the stimuli for this type of task, and these tasks are also commonly found on IQ tests (Wechsler, 1944). Non-word span tasks consist of hearing a string of nonsense words or syllables and immediately repeating them back in the same order. The non-words may vary in syllable length, frequency, wordlikeness, etc., and these considerations are reviewed below. Many different variations have been used, including the Non-Word Repetition Task (NRT; Archibald & Gathercole, 2006, 2007), Serial Recall (Archibald & Gathercole, 2007), the Children’s Test of Non-Word Repetition (CNRep; Gathercole, Willis, Baddeley, & Emslie, 1994), and Non-word Recognition

(Gathercole & Pickering, 1999; O'Brien, Segalowitz, Collentine, & Freed, 2006; O'Brien, Segalowitz, Freed, & Collentine, 2007).

Although both word and non-word span tasks are commonly used, non-word span tasks tend to be better predictors of language learning abilities than word span tasks (Baddeley, 2003). Words are recalled better than non-words because information in long-term memory can interact with the short-term store and support their representations (Gathercole, Frankish, Pickering, & Peaker, 1999). "Redintegration", proposed by Schweickert (1993), is the process by which long-term memory information supports and reconstructs the representations held in the short-term store. Non-words are less similar to the information in long-term memory, so their representations cannot be supported by long-term lexical knowledge. As a result, non-words test the capacity of the short-term store itself without any outside influence or support (Gathercole & Baddeley, 1993).

Real words represented in the short-term store can also be supported by semantics. Poirer and Saint-Aubin (1995) and Wetherick (1975) demonstrated that serial recall for words was facilitated when all the words belonged to the same semantic category. In addition, Walker and Hulme (1999) found that concrete words were recalled significantly better than abstract words, demonstrating an average advantage of 8.6% across all serial positions. They concluded that the additional semantic information attached to concrete words helps support their representations in memory. This support means that some word span tasks may actually measure more than just the capacity of the phonological loop, complicating the correct interpretation of results.

While non-word spans do not allow for semantic effects on recall, phonotactic and frequency effects can still be found. Gathercole et al. (1999a) tested the serial recall of words, high-probability non-words, and low-probability non-words. High-probability non-words were created to maximize biphone frequencies, while low-probability non-words were created to minimize them. Word recall was superior to non-word recall, as expected, but high-probability non-words were recalled significantly better than low-probability non-words. In analyzing the error patterns from the participants, the biggest difference between the recall of high- and low-probability non-words was that high-probability non-words had more correct recalls and fewer partial recalls than low-probability non-words. This indicates that the non-words also benefited from the process of redintegration, which used biphone frequency information to restore the mental representations of the high-probability non-words but not the low-probability non-words. The researchers also divided their participants into two groups based on their vocabulary abilities: a high vocabulary group and a low vocabulary group. The errors in both the high-vocabulary group and the low-vocabulary group showed the same influences from lexicality and phonotactic frequency, indicating that the difference between the groups was in their storage capacities and not in their ability to use the redintegrative process.

The Non-word Recognition task has also been used as a measure of PSTM capacity (Gathercole & Pickering, 1999; Gathercole, Pickering, Hall, and Peaker, 2001; Gathercole, Service, Hitch, Adams, & Martin, 1999; Walker & Hulme, 1999). In this task, participants hear two strings of non-words. The strings are either exactly the same, or two adjacent items are switched in their order. Participants must decide whether the two strings were the same or different. This task has been shown to correlate with children’s vocabulary scores as highly as digit span and Non-word Repetition (Gathercole et al., 1999b). It also demonstrates the same the same word length effect as repetition tasks, in which recognition scores for lists of short stimuli are higher than for lists of long stimuli (Gathercole et al., 2001; Walker & Hulme, 1999).

Overall research on this method has concluded that it is at least as good of a measure of PSTM as repetition tasks. There are also some advantages to using the Non-word Recognition task compared to Non-word Repetition tasks. It does not require the production of unfamiliar sounds and also avoids the scoring issues raised by differences in regional accent (Wells, 1995). In addition, recognition tasks have been shown to be less influenced by word concreteness than repetition tasks (Walker & Hulme, 1999) and they demonstrate a reduced lexicality effect (Gathercole & Pickering, 1999; Gathercole et al., 2001; Jeffries, Frankish, & Ralph, 2006). Phonotactic frequencies therefore have less influence on Non-word Recognition than Non-word Repetition tasks, making it an increasingly popular measure of PSTM.

Phonological Short-Term Memory and First Language Development

PSTM has repeatedly been shown to correlate with first language vocabulary abilities in children. Gathercole et al. (1999a) found that their high-vocabulary group of children had significantly higher recall scores for both individual phonemes and for nouns. Gathercole and Pickering (1999) also found higher levels of Non-word Repetition and Non-word Recognition for their high-vocabulary group than their low-vocabulary group. Adams and Gathercole (1995) found that 3-year-olds with higher PSTM (measured by Non-word Repetition and digit span) produced more unique vocabulary words, had longer utterances, and displayed more varied and more complex syntactic structures in their spontaneous speech. As an example, they found a correlation of $r = .41$ between the children's repetition of three-syllable non-words and the number of unique words they produced. Adams and Gathercole (1996) also found that non-word repetition scores in 4- and 5-year-old children correlated at $r = .35$ with vocabulary scores and $r = .42$ with their mean length of utterance (MLU). They concluded that Non-word Repetition scores were good predictors of children's vocabulary production and their expressive language performance. In 2000, Adams and Gathercole replicated their results from 1995, this time with 4-year-old children. Non-word Repetition correlated positively with word production, MLU, and variation in syntactic constructs.

Blake, Austin, Cannon, Lisus, and Vaughan (1994) found similar results to Adams and Gathercole (1995, 2000). In their study, word span was a good predictor of spontaneous language complexity, better than either mental or chronological age. Word span also accounted for 44% of the variance in MLU for their 3-year-old participants. In a review of the literature to that point, Gathercole and Baddeley (1993) performed cross-lagged correlations on measures of phonological memory and vocabulary development in children ages 4-6. They found that the correlation between Non-word Repetition at age 4 and vocabulary at age 5 was stronger than the correlation between vocabulary and age 4 and Non-word Repetition at age 5. This result strongly supports a causal relationship of phonological memory on vocabulary acquisition, at least between the ages of four and five. Gathercole et al. (1999b) found that scores on their vocabulary measure correlated significantly with three different phonological memory measures (Non-word Repetition, r = .54-.61; digit span, r = .44-.67; and Non-word Recognition, r = .73) in both younger (4-year-old) and older (5-6-year-old) children. Non-word repetition still correlated with vocabulary at r = .49, even after articulation rate was accounted for. They concluded that the phonological loop is critical for the long-term learning of novel phonological forms in both younger and older children.

In a classic study, Daneman and Case (1981) created novel utterances that described various interactions between bug-like creatures. The researchers taught the actions and their non-word labels to children aged 2-6 and then measured their ability to comprehend and produce the utterances describing these actions. They found that word span correlated significantly with both production and comprehension of the action descriptions and that word span was a better predictor of performance than age. In another unique study, Baddeley and Gathercole (1990) taught children both common and phonologically novel names for various toys. They classified the children into high- and low-non-word repetition ability groups and found that those with low repetition skills had more difficulty learning the phonologically unfamiliar names. In contrast, both groups performed equally well learning common names for toys, indicating that the difference in ability between the two groups only held for phonologically unfamiliar sequences. In terms of production measures, Willis and Gathercole (2001) found that children with higher scores on digit span and the CNRep performed better on sentence comprehension and repetition measures. These results emphasize the importance of PSTM in language production abilities, as well.

Like WM, PSTM has also been shown to be an effective predictor of language abilities in populations of children with delayed or impaired first-language development. Thal, Miller, Carlson, and Vega (2005) found a significant difference in the number of phonemes correctly imitated by children with typical language development and those with a history of language delay. Children with SLI also do more poorly on Non-word Repetition tasks than those with normal language abilities (Archibald & Gathercole, 2006; Marton & Schwartz, 2003). Robinson, Mervis, and Robinson (2003) have studied children with Williams' syndrome, which is characterized by mild to moderate mental retardation but relatively spared language abilities. They found that backwards digit span, forward digit span, and Non-word Repetition all correlated significantly with these children's scores on the Test for the Reception of Grammar (Bishop, 1989). Based on their work, Robinson et al. (2003) argue that the ability to parse and analyze phonological representations is supported by WM. They also argue that this analytical ability is important for grammar learning, because individuals must be able mentally track and manipulate multiple aspects of an utterance simultaneously.

Taken together, these results indicate that phonological memory abilities are important across a large range of language skills and in populations with a variety of cognitive abilities. This is likely because phonological skills in memory allow for the development of representations of novel phonological content (Baddeley, Papagno, & Valler, 1988). This is important not only for vocabulary, but also for language expression more generally. Without the ability to remember new auditory information, language learning would be almost impossible.

Phonological Short-Term Memory and Second Language Development

In addition to its strong influences on first language acquisition in children, phonological memory has also been shown to be an important factor for both vocabulary and grammar development in second language learning. Ellis and Beaton (1993a) tested the effectiveness of various vocabulary learning strategies and found that repeating foreign-language words helped participants learn them better. This suggested that the phonological loop, used for repetition, was relied upon for phonologically unfamiliar material. Atkins and Baddeley (1998) measured adults' verbal span using both digit and letter span tasks and found a correlation between those scores and the speed of learning English-Finnish vocabulary pairs. Masoura and Gathercole (1999) also found that both foreign- and native-language Non-word Repetition scores correlated significantly with measures of second language (L2) vocabulary. This relationship stayed significant even after taking into account the participants' native-language vocabulary levels (r = .32).

Baddeley, Papagno, and colleagues (Baddeley et al., 1988; Papagno, Valentine, & Baddeley, 1991) studied Patient PV, who suffered a stroke that selectively damaged her phonological loop. The researchers taught her pairs of words in her native language, which she learned without difficulty. However, she was greatly impaired when attempting to learn pairs of native language-foreign language words. They concluded that foreign language vocabulary learning depended upon the disrupted phonological loop, whereas native language words were learned by semantic mediation (Baddeley et al., 1988). Similar results have also been reported by Martin (1993), whose Patient EA had a severely limited phonological short-term memory span and difficulties learning foreign languages, and Baddeley (1993), whose patient NR had a relatively impaired digit span and a history of failing to learn foreign languages. The importance of the phonological loop in L2 vocabulary learning was further supported in a study by Papagno et al. (1991). Their participants were unable to learn foreign language words under an articulatory suppression condition, which involved them continuously repeating the syllable "bla" and therefore being unable to use their phonological loop to rehearse the words. This was especially the case when the words were very dissimilar from their native language. These results support the importance of the phonological loop for foreign language vocabulary learning and extend this finding to a normal adult population.

In addition to vocabulary, PSTM also seems to be important for L2 grammar and fluency development. Service (1992; Service & Kohonen, 1995) conducted a longitudinal study of foreign language (English) learning in Finnish-speaking primary school children. She found that English (L2) Non-word Repetition abilities at the beginning of primary education were a good predictor of success in English learning during the first 2-3 years formal education (Service, 1992). Service and Kohonen (1995) found that Non-word Repetition also correlated significantly with children's scores on both a communicative test of English and a more "traditionally-styled" (p. 161) test of their knowledge and use of English.

Ellis and Sinclair (1996) tested adults' ability to learn Welsh as a foreign language and found that participants who repeated the language aloud scored significantly higher on vocabulary, use of phrasal constructions, and correct sound mutations. They concluded that the more often foreign language structures are repeated, the easier it is to learn them and generalize rules from them. Sleve and Miyake (2006) found that phonological memory accounted for significant variance in both syntax (r = .48) and receptive phonology (r = .37), even after age of arrival, length of residence, and exposure to and use of the L2 were taken into account. In addition to their results with working memory, Kormos and Sáfár (2008) also found that non-word span correlated with writing scores, Use of English scores (which included knowledge of vocabulary and grammatical constructions), and overall proficiency scores in adolescent pre-intermediate learners of English.

French (2006) studied French-speaking children entering an intensive English program. His measures of English Non-word Repetition (ENWR) and Arabic Non-word Repetition (ANWR) showed significant correlations with L2 proficiency measures, with r values all at least .60. Phonological memory correlated with overall scores on syntax measures both at the beginning and end of the intensive language program, and this relationship was still significant after partialing out vocabulary knowledge. Phonological memory accounted for up to 67% of the variance in L2 proficiency at the end of five months, and it was more important for the participants with lower levels of proficiency. French and O'Brien (2008) found that Time 1 ANWR and ENWR predicted L2 productive vocabulary (*r*s between .48-.64) and grammar (*r*s between .79-.82) scores at Time 2, after a 5-month intensive study program. The phonological memory scores also explained up to 27.9% of the variance in grammar scores at Time 2, even after controlling for vocabulary knowledge at either Time 1 or Time 2

Speidel (1993) studied a brother and sister who were native speakers of German and L2 speakers of English. The brother, Mark, had difficulty with gender forms and case endings in German, and produced shorter and simpler utterances than his sister. He also tended to produce fewer unique words. Speidel discovered that he had difficulties with his PSTM, measured by a word span test. Based on Mark's language difficulties, Speidel concluded that phonological memory is important for creating stable representations of both vocabulary and grammatical structures. Difficulties in creating these representations would lead to problems building a storehouse of token phrases from which to generalize grammatical rules. Speidel suggested that phonological memory would be especially important for learning material not easily learned through another method, such as semantic mediation. Grammatical function words and morphemes are often devoid of clear semantic meaning (DeKeyser, 2005), suggesting that they should be especially dependent upon phonological memory. Ellis (1996) also argues that vocabulary and grammar learning may be dependent on similar methods of acquisition. Both processes involve the abstraction of patterns from unfamiliar phonological material; this occurs on the level of phonemes for vocabulary and on the level of morphemes for grammar. However, morphemes and morpheme combination sequences are much less fixed and often more abstract; because of this, learning these should rely even more heavily on memory abilities (Speidel, 1993).

O'Brien and colleagues also found important influences of phonological memory capacity on measures of adult L2 development, including vocabulary, correct use of grammatical structures, and fluency (O'Brien et al., 2006; O'Brien et al., 2007). O'Brien et al. (2006) found that phonological memory (measured by Non-word Recognition) correlated with vocabulary scores, narrative abilities, and use of free grammatical morphemes and subordinate clauses, both at the beginning and the end of a semester of Spanish learning (*r*s between .30-.41). The students with larger phonological memory capacities also made more gains in these measures than those with lower memory capacities. O'Brien et al. (2007) extended these findings to measures of overall fluency in adult L2 learners. Non-word Recognition correlated at $r = .34$ with the absence of filled pauses in speech, as well as with the development of this fluency skill ($r = .30$). PSTM also explained a significant amount of variance in the total number of words produced, the number of words in the longest turn without a pause, and the ability to speak without filled pauses. They concluded that phonological memory is important for the ability to imitate and retain utterances and the development of fluency, especially in less experienced learners.

Finally, Williams and Lovatt (2003) investigated the relationship between PSTM and the ability to generalize grammatical gender. They found that phonological memory correlated with vocabulary learning efficiency ($r = .51$) and with the number of determiner-noun pairs correctly recalled on the first test ($r = .53$). Using an artificial language, they found that phonological memory predicted grammatical generalization abilities at $r = .60$ and that this correlation appeared significant across multiple cycles of generalization tests. They interpreted this finding as indicating that PSTM correlated better with the rate of learning than its ultimate level of attainment. A new test measure, Memory for Morpheme Combinations, measured the ability to store combinations of familiar but semantically empty morphemes. This measure also predicted generalization scores, with *r* values ranging from .47-.64. Like Speidel (1993) and Ellis (1996), Williams and Lovatt (2003) also opined that grammar rules are generalizations of patterns across sequences of words. They argued that if PSTM is related to learning the words, it should also be related to learning the patterns among them. They suggested that since there appears to be a consistent relationship between PSTM and vocabulary (e.g. Service, 1992) and between vocabulary and grammar (e.g. Service & Kohonen, 1995), it is reasonable to assume a connection between PSTM and grammar, as well.

Summary of Previous Findings

As reviewed above, a large amount of research has indicated that PSTM and WM are important for various aspects of language ability. A significant amount of developmental work shows a consistent and strong relationship between PSTM and vocabulary abilities. Most researchers attribute the connection between PSTM and language learning to the importance of the phonological loop for forming stable, long-term mental representations of novel phonological material (such as new vocabulary words). These representations are important for item knowledge, as well as the ability to abstract patterns from this information. Some connections have also been found between WM and language abilities. These relationships are usually attributed to an individual's ability to parse, analyze, and effectively manipulate new linguistic items and structures. The attentional aspect of WM is also important because it allows learners to selectively attend to multiple aspects of the linguistic structure at once.

Baddeley and colleagues (Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998) have concluded that the evolutionary purpose of the phonological loop is to serve as a language acquisition device for both vocabulary and syntax. This view is based on the large body of evidence from child language acquisition and foreign language learning studies, reviewed above, that indicate the importance of phonological memory in vocabulary, grammar, and fluency development. This conclusion is also supported by computational models of language acquisition, in which the development of both individual vocabulary items and their morphological properties depends on a single underlying mechanism (Bates & Goodman, 1997; Plunkett & Marchman, 1993). The author will return to this viewpoint in the Discussion.

However, there are still some unanswered questions. There is much less evidence for a relationship between memory measures and the development of grammatical abilities. In addition, some authors have suggested that the relationship between PSTM and grammar abilities is mediated by vocabulary knowledge (e.g., Service & Kohonen, 1995). Finally, the majority of the work that has been done with these memory measures and their relationship with language abilities has been developmental or observational, rather than strictly experimental. Further research is necessary in order to more clearly define the importance of PSTM and WM for language development. The current study intends to address these issues.

The Current Study

The purpose of the current study was to further investigate the relationship between both working memory and phonological short-term memory capacities and second language learning in adults. Three memory tests were used: Listening Span (adapted from Harrington & Sawyer, 1992), measuring working memory; and Non-word Repetition (Gathercole et al., 2001) and Non-word Recognition (O'Brien et al., 2006, 2007), measuring phonological short-term memory. An artificial language based on non-word stimuli was used for the language learning tasks. Participants learned the vocabulary and were then exposed to plural markings in a sentence context, without explicit instruction. Their use of plural markings on nouns, verbs, and adjectives was measured in a generalization test at the end of the study; these scores reflected their ability to generalize the plural structures on-line during the task. Based on the previous findings in the literature, the hypotheses were as follows:

1. There will be a significant positive correlation between the scores on the PSTM tasks and the vocabulary scores during the generalization test;
2. There will be a significant positive correlation between the vocabulary scores and the grammatical generalization scores;
3. There will be a significant positive correlation between the scores on all measures of memory capacity and the grammar scores from the generalization test;
4. The correlation between PSTM scores and grammar scores will be larger than the correlation between WM scores and grammar scores.

Method

Participants

Fifty monolingual native English speakers (40 females, 10 males) were recruited from the University of Michigan in the fall of 2008. Three participants were excluded because post-experimental questionnaires revealed that they grew up as bilinguals. Seven additional participants did not return for the second session of the study, leaving a total of 40 participants (36 females, 4 males) included in all the analyses. Data sets from the first session include scores from all 47 monolingual participants. Data sets from the second session include only the 40 participants who completed the entire study.

Individual Differences Measures

Non-word Repetition. Participants completed a Non-word Repetition task as a test of their phonological short-term memory. Participants heard a list of one-syllable non-words and were asked to repeat them as clearly and accurately as possible. The non-words were pre-recorded by a female native English speaker and were presented aurally through headphones, using the psychology software E-Prime (Schneider, Eschman, & Zuccolotto, 2002). There were four lists at each of four syllable lengths: three, four, five, and six, and each list was created to be as phonologically distinct as possible (Gathercole et al., 2001). All participants heard the lists in the same order, beginning with the lists with three stimuli and increasing to the lists with six stimuli. Participants' responses were recorded through a microphone using the sound-editing software Audacity (Audacity Team, 2008). The non-words used in this portion of the study were randomly selected from the non-word stimuli used by Gathercole et al. (2001); this list can be found in Appendix A.

Non-Word Recognition. Participants completed a Non-word Recognition task as an additional measure of their PSTM. They listened to two presentations of a list of non-word syllables and judged whether the two presentations were exactly the same or slightly different. These stimuli were also composed of one-syllable non-words taken from Gathercole et al. (2001) and also used by O'Brien et al. (2006). Because of the limited number of non-word stimuli, some items were repeated from the Non-word Repetition stimuli. Eight lists were used at each of three list lengths: five, six, and seven items. There were also four practice trials with four items each. Items in each list were chosen so that they were maximally phonologically distinct. Lists were pre-recorded by a female native English speaker and were presented through headphones, using the psychology software E-Prime. Of the eight lists at each length, four lists were created that were identical and four lists were created that were different. For the identical presentations, the same list was presented twice, with a short pause (1200 ms) between the two presentations. For the presentations that were different, the first presentation of the list was followed by a short pause (1200 ms) and a second presentation of the list with two adjacent items from the list transposed. The location of the transposed syllables within the list of items was randomized so as to be unpredictable, with the exception that the first and last syllables were not transposed. This was done to reduce the salience of the transposed items and encourage the participants to process the entire string of items (O'Brien et al., 2006).

Participants first completed the practice set, consisting of four lists (two same and two different). Participants received feedback on the practice set only because pilot-testing revealed this was necessary for participants to learn the task and to score above chance on the task as a whole. After completing the practice set, participants began with the five-item lists, and then moved on to the six- and seven-item lists. Within each list length, the order of the same and different lists was randomized. Non-word items used in this portion of the study can be found in Appendix B. All non-words used as vocabulary in the artificial foreign language were used as stimuli during this memory task. This portion of the experiment therefore served not only as a test of phonological short-term memory capacity, but also to familiarize the participants with the sounds of the vocabulary they would be learning (Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Williams & Lovatt, 2003).

Listening Span. Listening Span was used as a measure of working memory capacity. Forty sentence stimuli from Harrington and Sawyer (1992) were used and recorded by a female native English speaker. Half the sentences were grammatical and half the sentences were ungrammatical, which was done by mixing up the word order of the second half of the sentence so that it no longer made sense as a sentence in English. The sentences used in this portion of the study can be found in Appendix C. Using headphones, participants listened to these sentences, arranged in sets, with the number of sentences in each set increasing from two to six. Two sets of sentences were presented at each list length (two sets of two sentences, two sets of three sentences, etc.). Participants completed the two sets of two sentences as practice sets; the two sets each of three to six sentences were the test sets. In each set of sentences, participants heard the sentences one at a time and were asked to judge whether each sentence was grammatical (i.e., would make sense as a sentence in English). At the end of each set, participants were asked to recall the final word of each sentence. These responses were recorded on a computer using a microphone and the software Audacity.

The Artificial Language

The artificial language created for this study consisted of twenty-one nouns, ten verbs, four adjectives, and three prepositions. Each noun and verb had three forms: singular (one entity or an action being performed by one entity), dual (two entities or an action being performed by two entities), and plural (three or more entities or an action being performed by three or more entities). Plurality was marked on nouns, verbs, and adjectives by means of a prefix.

All regular nouns had no prefix for the singular form, the prefix 'zi-' for the dual form, and the prefix 'na-' for the plural form, with one vowel mutation. Nouns that began with nasal consonants, /m/ or /n/, took the prefix 'za-' for the dual form and the prefix 'no-' for the plural form. All regular verbs had no prefix for the singular form, the prefix 'ta-' for the dual form, and the prefix 'moo-' for the plural form, also with one vowel mutation. Verbs that began with nasal consonants, /m/ or /n/, took the prefix 'too-' for the dual form and the prefix 'mi-' for the plural form. Adjectives took the same prefixes as the noun they modified. The word order of the artificial language was the same as English, with one exception: adjectives followed the nouns they modify, rather than coming before them.

An example a sentence with all singular words can be found in (1):

(1) Lork cham mord kib dook

Cow big is on table

"The big cow is on the table"

As can be seen, the language is similar to English, with the exception of the noun/adjective word order and the lack of articles. Examples of sentences using plural forms can be found in (2) and (3), where DP indicates the dual prefix:

(2) Na- targ moo-dern zi- jick zi-leck

DP-fish DP- throw DP-book DP-red

"Two fish throw two red books"

(3) No-nog moo-pag za-nart

DP-man DP- catch DP-ball

"Two men catch two balls"

(2) demonstrates the regular prefixes for two entities, and (3) demonstrates their irregular forms, with the vowel mutation occurring before a nasal consonant. A complete set of words in all forms can be found in Appendix D.

The structure of this language was designed partially based on the language system used by Daneman and Case (1981). They demonstrated that language learners are able to learn the meaning of prefixes from context, without explicit instruction. The plural structure (singular, dual, and plural) was based on the structure found in Arabic (Haywood & Nahmad, 1965).

Procedure

The study was completed over the course of two approximately one-hour sessions in the computer laboratory of the English Language Institute (ELI) at the University of Michigan. A schematic representation of the study as a whole can be found in Figure 2. During the first one-hour session, participants read and signed the consent form and the researcher answered any questions regarding the study. Participants then completed all three memory tasks (Non-word

Repetition, Non-word Recognition, and Listening Span). After finishing these three tasks, the participant began learning the artificial language.

The vocabulary for the language was presented aurally through headphones while the corresponding meaning was presented visually on a computer screen, by way of an illustration. Participants were asked to repeat each word aloud, in order to reinforce its learning (Ellis & Sinclair, 1996). They then heard the translation and were asked to repeat that, as well. The singular form of each noun and verb was presented twice, once with English first and once with the artificial language first. The dual and plural forms of each word were withheld for later presentation and testing. Participants never saw a written form of the foreign language.

After presentation of the vocabulary forms, participants listened to 15 sentences in the foreign language which used all of the vocabulary they had just learned. These sentences served to reinforce the vocabulary and introduce the participants to the sentence structure in the artificial language. They were asked to repeat each sentence and their repetitions were recorded (repetition paradigm). While the participants were listening to each sentence, they also saw an illustration of that sentence and the English translation of the sentence below it. After this initial presentation, participants listened to the same 15 sentences a second time. This time, they saw the illustration of the sentence and heard the English translation. They were asked to translate each sentence into the foreign language as best they could, and to record their answer (translation paradigm). After each translation attempt, they pressed a key and heard the correct translation as feedback. These 15 sentences were presented in a random order and can be found listed in Appendix E as ‘Set 1’. This concluded the testing in the first session; at this point, participants completed a brief questionnaire that asked for their age, sex, native language(s), and foreign language learning experience.

During the second session, between two and seven days after the first session, participants began with a vocabulary review. Participants heard a word in the foreign language and saw four illustrations on the screen. They had to choose which illustration corresponded to the word that they heard. They were given feedback on whether their answer was correct or incorrect, and regardless of their accuracy, the word and its illustration were repeated to reinforce the meaning. This review continued until participants reached 85% correct on a single run-through of all 38 words. The number of errors made and the number of trials needed to reach the criterion were recorded. After reaching criterion for vocabulary knowledge, participants reviewed the 15 sentences from session one. Half the sentences were presented using the repetition paradigm and the other half of the sentences was presented using the production paradigm. The order of sentence presentation within each half was randomized for each participant.

After reviewing these sentences once, participants were presented with 30 new sentences. Presentation of this second set of sentences was similar to that of the first set. For the first half of the sentences, participants saw an illustration and the English translation below it. They heard the sentence in the foreign language and were asked to repeat it and record their repetition. For the second half of the sentences, they saw an illustration and heard the English translation of a sentence. They were asked to translate this sentence into the foreign language to the best of their ability and to record this translation. After their translation attempt they heard the correct translation as feedback. Twenty-three of the 30 sentences included at least one plurality form (not previously encountered) of a known vocabulary word. Within each half of the stimuli, sentences were presented in a random order for each participant. These sentences served to introduce the participants to the grammatical plurality patterns without explicit instruction and to provide the data from which to generalize the grammar rules. These sentences can be found in Appendix E, listed as ‘Set 2’.

Participants were given the option of a five-minute break before beginning the final phase of the study, which was the Test Productions phase. The test phase consisted of fifty sentences, again presented through headphones. Thirteen sentences were repeated from Set 1, 13 were repeated from Set 2, and the remaining 24 sentences were novel. The novel sentences each used at least one plurality form that participants had not yet encountered. For example, if participants had learned the word for one baby (‘charb’), a new sentence may have required them to produce the word for two babies (‘zi-charb’). Half the sentences were heard in English, and participants saw an illustration of each sentence. The task was to produce the foreign language translation (production scores). For the other half of the sentences, participants heard an utterance in the foreign language and had to translate it into English (comprehension scores). On these trials, participants did not see an illustration, because this would have allowed them to “translate” the sentence without necessarily having understood the foreign language. The complete set of Test Productions sentences can be found in Appendix E. The style of presentation for the language was very similar to that used successfully by Brooks et al. (1993). This similarity in presentation method helps assure that it is possible for participants to learn a foreign language in this manner.

The purpose of the generalization task was to measure how well participants had abstracted the plurality rules from Set 2 and were able to generalize them to new vocabulary. This demonstrated their overall knowledge of the grammar system. At the end of the second session, participants completed a second questionnaire that assessed their explicit knowledge of the language they learned. They were asked to describe the grammar of the language, explain the rule for forming plurals, indicate whether they had noticed any sound changes in certain contexts, describe those changes, and comment on how they had learned the language.

Scoring

All responses made by participants were recorded using a microphone and the sound-editing software Audacity (Audacity Team, 2008) and scored off-line, in order to increase the accuracy of scoring. Three individuals with transcribing experience transcribed participants' utterances from throughout the study. Inter-transcriber reliability was 89%, measured on 5% of the transcriptions.

For Non-word Repetition, scoring was done on a phoneme-by-phoneme basis. Total number of phonemes recalled (251 possible) and the maximum number of phonemes recalled on any one repetition set (maximum 22) were calculated for each participant. For Non-word Recognition, participants received one point for each correct same/different judgment. The maximum score possible was 24. For Listening Span, two scores were taken: the number of words correctly recalled in the correct relative order (LS Order, 41 possible) and the number of words correctly recalled, disregarding order (LS Item, 35 possible). The grammaticality judgments were not scored, but served as a manipulation check to ensure that the participants actually processed each sentence as a whole (Harrington & Sawyer, 1992; Turner & Engle, 1989). All participants scored at least 85% on the grammaticality judgment task.

For the repetitions and productions of all the foreign language stimuli, utterances were scored for accuracy in a number of categories: nouns, verbs, adjectives, prepositions, noun plurality, verb plurality, adjective plurality, and noun/adjective word order. Participants received 1 point for each vocabulary word that was produced correctly, .66 points for each vocabulary word that had only one incorrect phoneme, or 0 points for each vocabulary word that had two or more incorrect phonemes. For example, if the target word was “charb”, participants would earn 1 point for producing “charb”, .66 points for producing “chard”, and 0 points for producing “chob”. On Set 2 and Generalization Set stimuli, participants also received scores on their ability to repeat or produce the plural markings. They received 1 point for each correct plural marking, .25 points for each prefix that was correct in word type (noun, verb, or adjective) but incorrect in plurality (dual vs. multiple), or 0 points for a plural marking that was missing or from the wrong word type. For example, if the target word was “zi-charb”, participants would earn 1 point for producing “zi-charb”, .25 points for producing “na-charb”, and 0 points for producing “charb” or “ta-charb”. The scoring was done in this way to give participants as much credit as possible for learning even part of the grammatical structures, considering the relatively small amount of input and feedback available to them. Participants also received one point if they were able to correctly repeat or produce the verbal mutation that occurred in the prefix before a nasal consonant. Finally, participants received 1 point if they were able to produce the noun and adjective in the correct order (opposite that from English).

Participants’ foreign language learning experience was assessed by calculating the total number of years spent studying foreign languages. For example, if a participant had studied Spanish for four years and French for two, the Total Years Studied variable would be six. Responses to the second questionnaire, assessing participants’ explicit knowledge of the language system, were converted to numeric data by a point scale. Three scores were calculated in this manner. For descriptions of the overall grammatical structure of the language, participants received one point for each unique aspect of the grammar they listed (noun/adjective word order, SVO, etc.). For descriptions of the rule for forming plurals, participants received one point for each aspect of the rule that they were able to describe (prefixes, both dual and plural forms, etc.), as well as for each unique example of the plurality markings that they provided. For descriptions of the vowel mutation, participants received one point for recognizing that it existed and one additional point for each correctly identified context for mutation (e.g., in a plural prefix) and each specific example they gave.

Results

Description of Variables of Interest and Analyses

The variables of interest for the study were the demographic information about the participants (taken from the first questionnaire), their explicit knowledge of the grammar of the foreign language (taken from the second questionnaire), their scores on the memory measures, and the measures of their vocabulary and grammatical proficiency from throughout the study. In addition to the individual scores on vocabulary and grammar for each word type in each set of stimuli, composite vocabulary and grammar scores were also computed for the various sets of stimuli. Composite vocabulary scores include scores on each part of speech from the appropriate stimuli set (nouns, verbs, adjectives, and prepositions). Composite grammar scores include plural morphology scores on each part of speech from the appropriate stimuli set (nouns, verbs, and adjectives), as well as the score for correct production of noun/adjective word order.

The “Test Productions” variables are the overall scores earned during the final test phase of the study. These scores include the novel stimuli (the “Generalization Set”), as well as stimuli repeated from Set 1 and Set 2. The Test Productions scores also collapse across measures of foreign language comprehension and production. Set 1, Set 2, and Generalization Set scores are also reported individually, and refer to that specific subset of stimuli from within the Test Productions overall. As mentioned above, some data from the participant questionnaires were also included in the analyses. “Describe Grammar” refers to the number of points participants received for their general descriptions of the grammar system. “Rule for plurals?” refers to the number of points participants received specifically for their descriptions of the rule for forming plurals in the foreign language. “Irregular changes?” refers to the number of points participants received for their descriptions of the irregular vowel mutations in the language. Table 1 contains a summary of the terminology used for the data presented here, as well as full descriptions of each variable and the shorthand used for them throughout the paper. Details on the scoring of these variables can be found in the Method section.

The maximum, minimum, mean, standard deviation, and points possible for each of these variables of interest can be found in Table 2. The majority of variables were normally distributed and parametric statistics (Pearson’s correlations and least-squares multiple regressions) were used for most analyses. However, there were a few variables whose distribution was rather skewed. The scores for plural marking on verbs and adjectives were skewed in the Test Productions set overall (skewness statistics were 2.889 and 1.553, respectively), the Generalization Set (2.662 and 1.518, respectively), and production of the Generalization Set (2.662 and 1.518, respectively). Skewness was also found for comprehension of adjectives in the Generalization Set (1.454), as well as two of the three composite developmental vocabulary scores (Set 1 production, 1.483; Set 2 production, 1.541). Generally, these variables did not correlate with the memory measures, and this was likely due to their skewed distributions. To address this issue, Spearman’s rho was also used to calculate correlations for some of these scores.

The question arises as to why these scores were so skewed. Part of the problem may be the amount of exposure that participants had to these particular forms. Plural nouns occurred 167 times in the stimuli, while plural verbs only occurred 80 times and plural adjectives only 26 times. These are clearly large differences, and participants may not have been exposed to the plurality on verbs and adjectives enough times to learn them well; it is well known that frequency of presentation affects learning (Braine, Brody, Brooks, Sudhalter, Ross, Catalano, et al., 1990). This could have easily led to skewed scores, with some participants learning some of the structures but the vast majority of participants never gaining competence on them. Regarding the developmental vocabulary scores, the skewness in these data may have origin in the very fact that they are developmental scores. A few participants doing very well while most did poorly could have resulted in these skewed distributions. The nature of these specific data points requires that conclusions about them must be made cautiously.

On a different note, inspection of the first few rows of Table 3 reveals that the PSTM and WM scores did not correlate with each other. This indicates that they were indeed measuring two separate cognitive processing abilities, consistent with the previous literature (e.g., Kormos & Sáfár, 2008). This also serves as an indication that the measures were implemented effectively overall. However, Non-word Recognition scores did not correlate as highly with Non-word Repetition as might be expected from the previous literature (Gathercole & Pickering, 1999; Gathercole et al., 1999b). This concern will be addressed further in the Discussion.

Correlational Analyses

Memory Measures and Vocabulary Scores. As reviewed above, a common finding in the literature is a consistent relationship between PSTM and vocabulary scores (Baddeley et al., 1988; Masoura & Gathercole, 1999). In order to test whether this pattern was also found in the current study, correlations were computed between the measures of PSTM and WM and the vocabulary scores from the final test phase of the study. These correlations can be found in Table 3. NR Max (the maximum number of phonemes recalled during any one Non-word

Repetition trial) correlated with the most measures of vocabulary abilities. In the Generalization Set, it correlated with both production and comprehension of nouns and prepositions, as well as with production of adjectives and comprehension of verbs. NR Max also correlated with composite vocabulary scores from both the Generalization Set and the Set 1 stimuli. Figure 6 depicts the relationship between NR Max and the composite Generalization Set vocabulary score. These correlations remained significant when both production and comprehension were considered separately. NR Total (the total number of phonemes recalled overall) also correlated with multiple measures of vocabulary knowledge, including comprehension of verbs and prepositions, the Set 1 composite vocabulary score, and the Generalization Set composite vocabulary comprehension score. Accuracy on the Non-word Recognition task correlated with comprehension of nouns and verbs in the Generalization Set, as well as the Set 1 and Generalization Set composite vocabulary scores.

Overall, these correlations show a strong relationship between phonological memory and vocabulary knowledge during the final phase of the study. This demonstrates that they are good predictors of the final attainment of vocabulary, after a set amount of exposure to and practice with the language. This was especially true for NR Max, which correlated with 15 of the 22 vocabulary measures from the final phase of the study.

WM measures also correlated with some of the vocabulary measures from the final phase of the study. LS Item and LS Order both correlated with scores on verbs and prepositions in the Generalization Set and the Test Productions overall. They also correlated with the composite vocabulary scores from Set 2 stimuli and the Generalization Set. Figure 7 depicts the relationship between LS Order and the Generalization Set composite grammar score. However, WM scores did not correlate with any scores on nouns and adjectives, and when looking at the production and comprehension scores separately, only the correlations with vocabulary production remained significant.

The pattern of correlations between memory measures and individual vocabulary measures is interesting. While the PSTM measures correlated with all four word classes, the WM measures only correlated with scores on verbs and prepositions (one exception – a correlation with one noun score). Verbs and prepositions tend to express more abstract relationships and actions, compared to nouns, which express more concrete concepts. Since they are more abstract they may require more processing for their comprehension and production, potentially a factor in their relationship with WM measures. Nouns, and to some extent adjectives, are more easily imageable than are verbs and prepositions. They may demand less processing, and therefore rely less on the WM component of cognition. The close relationship between vocabulary scores and both PSTM and WM will be explored further in the Discussion.

In contrast to the final vocabulary scores, PSTM and WM did not show nearly as strong of a relationship with measures of vocabulary *development*, taken from the learning trials during the study (Set 1 and Set 2). The NR Total and NR Max scores correlated with the production of adjectives in Set 2 (r = .329, and r = .387, respectively, $p < .05$ for both). Listening Span Order correlated with production of verbs in the same set of stimuli (r = .328, $p < .05$). As mentioned above, many of the developmental vocabulary scores had skewed distributions, which may partly explain why so few correlations were found in this analysis. Future studies will need to address this issue to more clearly define memory capacity's role in vocabulary development.

Vocabulary Scores and Grammar Scores. Another finding reported in the literature is a strong relationship between vocabulary and grammar abilities (French & O'Brien, 2008; Bates & Goodman, 1997; Service & Kohonen, 1995). In order to determine whether this pattern of results also emerged in the present study, correlations between the composite measures of vocabulary and grammatical ability were calculated; these correlations can be found in Table 4. Overall, there were strong correlations between composite vocabulary and grammar scores, with *r* values typically ranging .60-.75. All three composite grammar scores (Test Productions overall, the Generalization Set, and Set 2) correlated with all six composite vocabulary scores: Test Productions overall, Set 1, Set 2, the Generalization Set, and both production and comprehension in the Generalization Set. Figure 8 illustrates the close relationship between Generalization Set vocabulary and grammar scores.

The composite vocabulary scores from each set of stimuli also correlated with multiple *individual* measures of grammatical ability. Test Productions vocabulary scores correlated with Test Productions noun plurality ($r = .622$, $p < .01$), verb plurality ($r = .619$, $p < .01$), and noun/adjective word order ($r = .741$, $p < .01$). Test Productions vocabulary scores also correlated with Generalization Set noun plurality ($r = .622$, $p < .01$), verb plurality ($r = .538$, $p < .01$), and noun/adjective word order ($r = .71$, $p < .01$). Similar correlations were also found when production and comprehension in the Generalization Set were considered separately.

Memory Measures and Grammar Scores. The main purpose of this study was to explore the relationship between memory measures and grammar abilities in a foreign language. To do this, correlations were computed between the PSTM and WM measures and various measures of grammatical ability from the final test phase of the study. These correlations can be found in Table 5. As with the vocabulary measures, NR Max was the PSTM measure that correlated with the highest number of grammar scores. NR Max correlated with composite grammar scores from the Test Productions, the Generalization Set, and Generalization Set production. The relationship between NR Max and Generalization Set composite grammar scores is shown in

Figure 9. NR Max also correlated with various individual measures of grammatical competence, including Test Productions noun/adjective word order and comprehension and production of noun plurality in both the Generalization Set and the Test Productions. Non-word Recognition accuracy correlated with correct use of noun/adjective word order in the Generalization Set. The NR Total score did not correlate with any measures of grammatical ability.

WM scores also correlated with grammar abilities. Listening Span Item scores correlated with four composite grammar scores: Test Productions, Set 2, Generalization Set, and Generalization Set production. LS Item also correlated with noun plurality in the Test Productions and the Generalization Set. Listening Span Order scores correlated with these same measures of grammatical competency – four composite grammar scores and noun plurality in the Generalization Set. As an example, the relationship between LS Order and Generalization Set grammar scores is illustrated in Figure 10. In addition, LS Order scores correlated with correct noun/adjective word order in both the Test Productions and the Generalization Set. As mentioned above, Spearman's rho was used to calculate correlations between the memory measures and the grammar scores with skewed distributions, which were verb and adjective plurality scores. This analysis revealed a significant relationship between LS Order and verb plurality in Test Productions ($r = .367$, $p < .05$) and the Generalization Set ($r = .336$, $p < .05$). Correlations with adjective plurality scores remained non-significant.

Inspection of Table 5 reveals that LS Order scores correlated with the most measures of final grammatical competence (ten), followed closely by NR Max scores (eight). It appears that both measures have strong relationships with participants' ability to generalize and apply grammar rules in a foreign language. However, the fact that a WM measure correlated with more grammar scores than a PSTM measure is contrary to predictions and suggests that both memory skills may be critical for grammar abilities in a foreign language. The implications of this finding will be considered further in the Discussion section.

Finally, correlations were also computed between the PSTM and WM measures and the developmental grammar scores earned during the plural training phase of the study, Set 2. These scores reflect grammar learning ability, rather than its final attainment. Two correlations were significant: both LS Item and LS Order correlated with the correct use of plurality on nouns (r = .422 and r = .410, respectively; $p < .01$ for both). These correlations suggest that WM is more important for the *development* of grammar abilities than is PSTM. This relationship is investigated further in the regression analyses, presented below.

Demographic and Questionnaire Information

Two demographic variables were considered in correlational analyses: age and total years of foreign language study. Age correlated with only two variables: Listening Span Item (r = -.391, $p < .01$) and Listening Span Order (r = -.403, $p < .001$). Working memory tasks are often included on standard measures of IQ (such as the Weschler's Adult Intelligence Scale, WAIS, Wechsler, 1944) and are assumed to measure fluid intelligence. Since fluid intelligence declines with age (Salthouse, 1992), the negative correlations found in this study most likely reflect this age-related decline in fluid intelligence. No other variables correlated with age, indicating that it did not have a large influence on any other scores. This conclusion is also supported by the regression analyses, detailed below, in which age never contributed a significant amount of variance to any of the scores.

Surprisingly, the total number of years participants had studied foreign languages had a negative correlation with four variables: verb plurality overall and in Generalization Set production (r = -.315, $p < .05$ for both), the Generalization Set composite grammar score (r = -

.327, $p < .05$), and Generalization Set grammar production ($r = -.343$, $p < .05$). This unexpected finding may be attributable to the extreme skewness found in the verb and adjective plurality scores. This interpretation is supported by the fact that the negative correlations between years of language study and the composite grammar scores were no longer significant when the skewed verb and adjective plurality scores were not considered (*p*s > .05).

Participants' scores on their explicit knowledge of the grammar were also included in the correlational analyses. Table 6 lists all the vocabulary and grammatical measures with which these data correlated. "Describe Grammar" correlated with the six composite vocabulary scores and the use of adjectives and prepositions in both the Generalization Set and the Test Productions overall. It also correlated with two grammar measures: noun/adjective word order and verb plurality in the Test Productions. "Rule for forming plurals?" correlated with the most vocabulary and grammar measures. These include comprehension and production vocabulary scores, noun and verb plurality measures in the final phase of the study, and noun/adjective word order usage (see Table 6 for details). "Irregular changes" correlated with scores of verb plurality and noun vocabulary score, both from the Generalization Set.

Regression Analyses

A series of forward-entry multiple regression analyses were conducted on measures of vocabulary and grammar abilities from throughout the study in order to determine the relative contribution of various predictor variables to the variance in the scores on these language abilities.

Vocabulary and Grammar Abilities in the Final Phase of the Study. In the first analysis, the dependent variable was grammar production in the Generalization Set. The predictor variables available to the model were NR Max, Non-word Recognition accuracy, LS Item, and

LS Order. The results of this analysis are presented in Table 7. LS Order was entered in the first step and explained a significant 13.8% of the variance. NR Max contributed an additional 9.9% of the variance when entered in the second step. This analysis supports the finding that both PSTM and WM contributed to a significant amount of variance in grammatical generalization scores.

Some researchers have found that the relationship between memory measures and the development of grammatical competency is mediated by vocabulary development (Bates & Goodman, 1997; Service & Kohonen, 1995). To test this hypothesis, further regression analyses were conducted with three grammar scores from the final phase of the study as dependent variables: Generalization Set grammar overall, Generalization Set grammar production, and comprehension of Generalization Set noun plurality. This last measure was used because comprehension of verb or adjective plurality or correct noun/adjective word order could not be measured when participants were translating into English, which does not mark these in the same way. It was therefore not possible to create a composite grammar score for comprehension of the foreign language.

For this set of analyses, the vocabulary score from the Generalization Set was forced in as the first predictor variable. This was done to control for vocabulary knowledge and to test for any unique variance contributed by PSTM or WM. The results of these analyses are presented in Table 8. The vocabulary score accounted for a significant amount of variance (30-55%) for each of the grammar variables. None of the PSTM or WM measures explained any additional variance in these grammar scores once vocabulary had been entered into the models. This finding seems to support the view that the relationship between PSTM and WM and grammar abilities is mediated by vocabulary knowledge.

Despite the finding that none of the memory measures contributed unique variance to the final grammar scores once vocabulary was controlled, there were still possible sources of variance in the data that had not yet been addressed, including Age and experience studying foreign languages. To address this issue, a final set of regression analyses was performed, with four of the grammar scores each treated separately as a dependent variable for analysis: composite Test Production grammar scores, composite Generalization Set grammar scores, and Generalization Set production grammar scores and comprehension of noun plurality.

In the first step, age and the total number of years studying foreign languages were forced into the model. This was done to control for any variance in the data that was due to these factors. After controlling for their influence, forward-selection multiple regression was used to determine which variables contributed additional variance to these grammar measures. The predictor variables included the memory measures, vocabulary scores, and participants' explicit knowledge of the language. The results of these analyses are detailed in Table 9 and summarized below.

Age and experience studying foreign languages did not account for a significant amount of variance in any of the grammar scores. After controlling for those factors, Test Productions vocabulary scores were the first predictors for all four grammar scores. These models predicted 35-60% of the variance in the grammar scores, indicating that vocabulary still contributed to a large amount of variance in the grammar scores. None of the other predictor variables (memory measures, explicit grammar knowledge) explained any unique variance.

Final Vocabulary Knowledge. A forward-selection multiple regression analysis was also performed with Test Productions composite vocabulary as the dependent variable and the PSTM and WM measures as predictor variables. This was done to determine the relative contribution of these variables to vocabulary scores, based on the importance of this vocabulary knowledge for grammar scores. Results from this analysis are presented in Table 10. Age and years of language study were forced into the model first to control for their influence, but did not account for any significant variance. When NR Max was added to the model, it predicted a marginally significant 11.1% of the variance in vocabulary scores. LS Order accounted for an additional 10% of the variance. This result supports the finding from the correlations analyses that PSTM and WM did indeed contribute to a significant amount of variance in the vocabulary abilities in this study.

Development of Vocabulary and Grammar Abilities. The previous analyses were concerned with the factors that contributed to the final attainment of vocabulary and grammar. It is also possible that PSTM and/or WM could have an effect on the development of vocabulary and grammar abilities (Williams & Lovatt, 2003). In order to further investigate this possibility, forward-selection multiple regression analyses were performed on composite vocabulary and grammar measures from throughout the learning periods of the study. The measures used as dependent variables were the vocabulary scores from Set 1, the review of Set 1, and Set 2, as well as grammar scores from Set 2. Given the few correlations found between these developmental scores and the memory measures, age, total years of foreign language study, and participants' explicit knowledge of the language system were used as additional predictor variables. Detailed results from these analyses are displayed in Table 11 and a summary follows.

Age and total years of foreign language study were forced into each model first to control for their influence, but did not account for a significant amount of variance in any of the developmental scores. For Set 1 vocabulary, the first significant predictor was participants' scores on their descriptions of the grammar of the language as a whole. After this, LS Item was the only other significant predictor. Together, these variables accounted for 30.7% of the variance in Set 1 vocabulary scores. For the review of Set 1 vocabulary, presented during the second session, participants' ability to describe the rule for forming plurals and their descriptions of the grammar in general both contributed significant amounts of variance. Their inclusion in the model explained 33.6% of the variance in Set 1 review vocabulary scores.

Looking at the Set 2 scores, only NR Max predicted a significant amount of variance in the vocabulary scores, accounting for 16.8% of the variance. For the Set 2 grammar scores, NR Total was the first predictor, followed by Non-word Recognition accuracy. Together, these variables accounted for 33.7% of the variance in the developmental grammar scores. This result is in line with Williams and Lovatt's (2003) finding that PSTM correlates more strongly with the development of grammatical abilities than their final attainment in the study. The broader implications and interpretations of these data will be explored further below.

Discussion

The goal of the present study was to explore the relationship between phonological short-term memory (PSTM) and working memory (WM) and the ability to learn grammatical structures in a foreign language. This research was motivated by an unclear relationship between PSTM, WM, vocabulary, and grammar learning, as reviewed in the introduction. While WM has been linked to language skills such as reading comprehension and verbal SAT scores (Harrington & Sawyer, 1992; Marton & Schwartz, 2003), fewer studies have shown a connection between WM and vocabulary and grammar in foreign language learning.

In this study, participants completed three memory tests: Non-word Repetition, Non-word Recognition, and Listening Span. They were presented with vocabulary from an artificial foreign language and practiced producing sentences in the language, using only singular forms.

They were exposed to plural forms in a sentence context and then tested on their production and comprehension of fifty sentences which included novel plural combinations. Measures of their vocabulary and grammar abilities were taken throughout the study and used as the dependent variables. The major correlations between the variables of interest are repeated schematically in Figure 3 and Figure 4 for convenience.

Memory Measures and Vocabulary Learning

As predicted, a positive correlation between PSTM and vocabulary was found (see Figure 3). The magnitude of these correlations, generally .30-.45, is moderately strong and is consistent with previous findings (e.g., Atkins & Baddeley, 1998; Masoura & Gathercole, 1999). Both NR Max and Non-word Recognition correlated with most of the six composite vocabulary scores, as well as a number of individual word-class vocabulary scores. WM also correlated with a number of vocabulary scores (see Figure 4); both LS Item and LS Order correlated with four of the six composite vocabulary measures. This result was unexpected because WM has not usually been associated with vocabulary learning (but see Kormos & Sáfár, 2008). Most of the research on WM has connected it with reading comprehension, fluency, and overall language abilities (Daneman & Carpenter, 1980; Harrington & Sawyer, 1992; Sunderman & Kroll, 2009; Turner & Engle, 1989), rather than vocabulary learning. The most likely reason for the correlation with vocabulary scores is the storage component of the WM task. The nature of this task requires the participant to explicitly remember and reproduce single word items. A participant's skill in remembering words for the WM task should also relate to his ability to remember individual vocabulary items.

Comparison of the correlations between vocabulary scores and PSTM and WM measures (see Figure 3) reveals that the strength of the correlation was slightly, but consistently, weaker between WM and vocabulary than between PSTM and vocabulary. This pattern is to be expected because of the nature of the stimuli in each task. Real words are used as the stimuli for WM tasks, as opposed to non-words, and research has shown that memory for real words does not correlate as well with measures of vocabulary learning as do measures of PSTM (Baddeley, 2003).

The regression analyses performed on Test Productions vocabulary scores confirmed the finding of a relationship between both PSTM and WM and vocabulary scores. After accounting for the influences of age and number of years studying foreign languages, phonological memory scores accounted for approximately 11% of the variance in vocabulary scores and WM accounted for an additional 10%. This analysis reinforced the finding that both PSTM and WM are influential for vocabulary scores. As will be argued further below, this also suggests that there is less of a distinction between these language sub-skills and their relationship with cognitive abilities than has been previously argued (Baddeley, 2003; Chomsky, 1957, 1965; van den Noort et al., 2006).

The development of vocabulary throughout the study was measured by looking at the scores from the learning periods (Set 1 and Set 2). Only a few correlations between the PSTM and WM measures and the developmental vocabulary scores were significant. Although this was not predicted, other researchers have found similar results. Masoura and Gathercole (2005) found that phonological memory did not relate the development of L2 vocabulary; instead, the beginning size of the L2 vocabulary was what accounted for the efficiency with which participants learned new L2 vocabulary. French & O'Brien (2008) also found that while phonological memory did have a significant relationship with final (Time 2) vocabulary scores, it did not relate as strongly to the actual development of vocabulary. Instead, the best predictor of vocabulary growth was Time 1 vocabulary knowledge. These results indicate that PSTM may be less of a factor in the actual process of vocabulary learning than previously thought. Since the language used in the current study was artificial and no measures of participants' English vocabulary abilities were taken, beginning vocabulary knowledge was not available for analysis. Future studies should continue to explore this issue and to clearly define the relationship between PSTM and cross-sectional vocabulary scores versus longitudinal vocabulary development.

Regression analyses were performed on the developmental data to determine which other measures may have contributed a significant amount of variance to them. Participants' explicit knowledge of the grammar of the language, measured by the detail of their descriptions of the grammar system, was the first predictor of their Set 1 developmental vocabulary scores. In contrast, NR Max was the only significant predictor of variance for the Set 2 vocabulary score. This relationship between vocabulary development and explicit grammar knowledge is unexpected and has no clear interpretation. One possibility is that it may have to do with participants' overall awareness of the language and concentration on specific aspects of it. However, it is not clear why the explicit knowledge scores would be predictors of developmental vocabulary measures and not final vocabulary or grammar scores, as well.

A few detailed aspects of the data deserve further discussion. Previous research has shown that Non-word Recognition tasks are equally as sensitive measures of PSTM as Non-word Repetition tasks (Gathercole & Pickering, 1999; Gathercole et al., 2001). However, in this study, NR Max scores correlated with 15 vocabulary measures while Non-word Recognition scores only correlated with nine. Inspection of the correlations between the various memory measures (see Table 3) reveals that Non-word Recognition only correlated with Non-word Repetition at $r = .34$. This finding raises the concern that Non-word Recognition and Non-word Repetition were not measuring the same cognitive ability and that the recognition score was not a valid measure of phonological memory in this study. It is unclear why this would be the case. The paradigm used in this study, including the stimuli, was taken from O'Brien et al. (2006), where it was used successfully to predict L2 fluency and grammar abilities. Future studies need to further address methodological issues concerning Non-word Recognition to ensure that it is a valid and consistent measure of phonological memory.

A few unexpected patterns emerged from the correlational data when the subsets of stimuli (Set 1, Set 2, and Generalization Set) were considered individually. None of the PSTM measures correlated with final vocabulary scores on stimuli originally presented during Set 2 (the plurality exposure phase). This is the only vocabulary subset that did not correlate with any of the PSTM measures and there are two statistical reasons that this is not a concern. As mentioned above, the PSTM measures *do* correlate with the overall measures of vocabulary, as well as with the other subsets of stimuli. This lack of a relationship may simply be a statistical artifact from separating out the various subsets of stimuli for the analyses. In addition, the relationship between the Set 2 vocabulary scores and the NR Max scores was only marginally non-significant ($r = .283$, $p = .076$).

A similar pattern emerged with a unique non-correlation between the Listening Span scores and the composite vocabulary scores earned on the original Set 1 stimuli. Participants' familiarity with these stimuli may have played a role, as the test phase was the fourth time participants had heard them. Because of this familiarity, participants may not have needed to process those stimuli as much to be able to produce them. Since working memory is typically assumed to measure processing abilities (Baddeley, 2003; van den Noort et al., 2006), scores on stimuli that don't need as much processing may not correlate as well with WM. This also explains why WM correlated with production measures but not comprehension measures. Comprehending an utterance in a foreign language is less difficult than producing one (Ellis & Beaton, 1993b), demanding fewer processing resources and therefore having a weaker relationship with WM. In addition, as above, the LS scores correlated with four of the five other composite vocabulary measures, indicating that there is indeed a strong relationship between the two. It is reasonable to assume that the lack of correlations in these two cases resulted simply from separating out the various subsets of stimuli; the scores are then taken from fewer trials, resulting in a less powerful measure.

Vocabulary Learning and Grammar Learning

As hypothesized, a strong relationship was found between vocabulary learning and grammar learning. All correlations between the composite measures for these two language abilities were significant and the values were in the range of .60-.75. Regression analyses were performed on the measures of grammar learning in order to determine whether or not their relationship with PSTM was mediated by vocabulary learning, as has been suggested in the literature (e.g., Service & Kohonen, 1995). The results of these analyses will be discussed below, in the context of the overall relationship between PSTM and WM measures and the grammar abilities measured throughout the study. Implications for the strong correlations between these two language skills will also be considered.

Memory Measures and Grammar Learning

The hypothesis of a relationship between the memory measures and grammar learning was supported (see Figure 5). A number of significant correlations were found between some of the measures of PSTM and grammar abilities from the end of the study, typically with strengths of r = .30-.45. Although there were not as many correlations with PSTM and grammar as there were with vocabulary, the strength of the relationships was remarkably similar. This replicates previous findings of a relationship between PSTM and grammar abilities (e.g., Kormos & Sáfár, 2008; O'Brien et al., 2006, 2007; Williams & Lovatt, 2003).

Contrary to predictions, however, the WM measures also correlated with a large number of grammar scores from the final test phase. In fact, measured by the number of correlations with grammar scores, the WM measures had a stronger relationship with final grammar scores than PSTM. This was unexpected because it is not a prevalent finding in the literature. It is PSTM which typically has a strong relationship with grammatical ability and fluency, with *r* values ranging anywhere from .30-.80 (Adams & Gathercole, 1996; 2000; Blake et al., 1994; Daneman & Case, 1981; French, 2006; Service, 1992). Fewer studies have related WM and grammatical abilities or fluency. The correlations that have been found tend to be somewhat smaller, with *r* values .20-.30 in L1 (Turner & Engle, 1989) and .30-.65 in L2 (Fortkamp, 1999; Harrington & Sawyer, 1992; Kormos & Sáfár, 2008). These results are what led to the hypothesis that PSTM would be a better predictor of grammar abilities than WM.

Despite the small amount of research on WM and grammar, the correlations between these measures are intuitively logical. Working memory, by definition, includes both storage and processing of information, with storage being necessary for memory and future use of the items. This component of WM explains its relationship with vocabulary learning – remembering individual items. However, grammar learning depends on more than just memorizing the correct morphemes. It can be viewed as the process of abstracting patterns from across language sequences presented as input (Ellis, 1996; Speidel, 1993). Although this is similar to vocabulary learning, which requires learning the sound patterns of words, vocabulary patterns are much simpler than grammatical patterns. They occur over a shorter span in the language (the individual word) and usually depend upon the word class. In contrast, grammatical patterns are much more global, applying to the utterance as a whole and not just the individual word. These more complicated and wider-ranging patterns should therefore demand more processing capacities, measured by WM but not PSTM.

Using grammatical knowledge to produce novel utterances also requires a lot of processing ability. Once learners have understood the input and parsed the utterance into individual units, they must be able to process those units and recombine them in novel ways in order to generalize the grammatical structures that they learned. While this does require some storage capacity for the individual units, it should also depend heavily on the ability to process, analyze, manipulate, and recombine these units. Since WM measures both components, it should serve as a good predictor of such abilities. This was demonstrated in the current study. When the correlations involving LS measures are inspected carefully, LS Order scores correlated with the highest number of grammar scores. Since the order scores require both remembering the correct items (storage), their recall in the correct order (processing), and processing of the distracter grammaticality judgment task (more processing), this score is the most relevant for grammatical generalization abilities. Additionally, LS Order correlated with eight out of the ten scores specifically measuring participants' ability to generalize from the grammar structures that they encountered to novel combinations. This indicates that the measure did in fact relate to grammatical generalization abilities and not just the participants' ability to remember the utterances they had previously encountered.

The manner in which participants were introduced to the grammatical structures and the way in which they got feedback on their production attempts may also have affected the relationship between WM and grammar scores. Instead of being taught the grammatical morphemes and noun/adjective word order explicitly, as would likely occur in a classroom setting, participants were simply exposed to the structures in a sentence context. This was a much more ‘naturalistic’ way of learning grammar and requires much more processing and effort on the part of the language learner. Again, because participants had to parse the utterances themselves (processing) and were not simply taught the grammatical morphemes as individual items (which would just require storage of that information), WM may have been more important for the grammar learning task. Participants also did not get explicit feedback on their productions in the foreign language. Instead, they simply heard the correct utterance as feedback. In order to process this information and to learn from it, they would have had to maintain their own utterance in memory and actively compare it to what they were hearing as the correct answer. This, again, would be highly dependent on WM, incorporating both storage and processing capabilities. To reiterate, although the numerous correlations between WM and grammar scores were contrary to predictions, the relationship is logical because of the large amount of processing involved in grammar learning and production.

Vocabulary as a Mediating Factor

Despite the overall prevalence and strength of the correlations of both PSTM and WM with grammar scores at the end of the study, it is still possible that these relationships were mediated by vocabulary knowledge. In order to address this issue, regression analyses were conducted on the grammar scores to determine which variables contributed unique variance to them. The results from these analyses indicated that once vocabulary scores were accounted for, the memory measures did not contribute unique variance to any of the Generalization Set grammar scores, but LS Order did contribute a unique 5% of variance to the final grammar scores overall (Test Productions).

These findings do seem to qualify the correlational results to a certain extent. However, one must also consider the nature of the mediating vocabulary measures. Both NR Max and LS Order contributed unique amounts of variance to these vocabulary scores, underscoring the importance of both types of memory in language learning abilities overall. The fact that vocabulary scores from the final test phase mediated the relationship of both PSTM and WM with grammar scores does not change the fact that both PSTM and WM correlated strongly with these measures and were important contributing factors to the mediating vocabulary measure. Additional analyses, including the construction of a latent structure model, need to be conducted on this data in order to determine the values of these inter-correlations. This may be of great importance for interpreting the relationship between memory measures and grammar measures. It is also important to remember that WM did account for a significant amount of unique variance in the final grammar scores when all stimuli were included. This finding lends additional support to the argument that WM is important for multiple aspects of language learning, including proficiency in grammar abilities across both familiar and unfamiliar stimuli.

Developmental Data

Finally, PSTM contributed a significant amount of variance to the *development* of grammatical abilities and this relationship was not mediated by vocabulary knowledge. This finding somewhat supports the original hypothesis of a relationship between PSTM and grammatical abilities, although the developmental aspect was not originally addressed. Williams and Lovatt (2003) also obtained a very similar result. In their Experiment 1, they observed that on a test of grammatical generalization, PSTM related more to the rate of learning than its final level of attainment. They attributed this result to learning efficiency, which depended on the quality of memory representations. However, this conclusion must be reached tentatively in the current study because of the skewness of some of these data (discussed in the Results section).

Overall Results

The overall conclusion that can be drawn from the results of this study is that PSTM and WM are almost equally important for both vocabulary and grammatical abilities in a foreign language. Contrary to the original hypotheses, PSTM did not show a markedly stronger relationship with grammar than WM. Although many (but not all) of the correlations between the memory measures and the grammar scores were mediated by vocabulary knowledge, this does not undermine their importance. This is especially true because of the influence of the memory measures on the vocabulary scores in the first place.

The finding that both PSTM and WM show comparable correlations with vocabulary and grammar is in direct contrast with much of the previous literature, which has commonly found distinctions between PSTM and WM and their influences on language learning (e.g., Baddeley, 2003). The lack of distinction between these components can be interpreted in two ways: either the variables used in the current study were not successful in measuring what they were intended to measure, or there really is only a weak distinction (if any) between vocabulary and grammar and their relationship with WM and PSTM. As many individual findings were generally consistent with the previous literature (e.g., the relationship between PSTM and vocabulary, the relationship between vocabulary and grammar scores, the lack of correlations between WM and PSTM measures), the author believes that the measures taken in this study were generally effective. This leaves the latter possibility open to further discussion.

Broader Interpretation

So far, the data have been examined in the framework of a dual-mechanism theory, which posits that grammar and vocabulary are learned separately, relying on two different processing mechanisms (Brown, 1973; Pinker, 1991, cited in Marchman & Bates, 1994). Although the view that vocabulary and grammar are distinct is the traditional approach (Chomsky, 1957, 1965; Katz & Postal, 1964; cited in Bates & Goodman, 1997), recent theories have proposed that these components may be more integrated (e.g., Construction Grammar, Goldberg, 1995). Bates (Bates & Goodman, 1997; Marchman & Bates, 1994) is a major proponent of the view that both vocabulary and grammar are processed and learned by one unitary system and that the development of grammatical abilities necessarily depends upon the development of the vocabulary it organizes. This view has also been termed the "critical mass" hypothesis because it assumes vocabulary must reach a minimum "critical mass" size before extensive grammar abilities can be supported.

The results presented in Marchman & Bates (1994) support this view. Their data showed that use of irregular verb forms was closely related to the size of the children's vocabulary. They also showed that the use of grammatical function words accelerates rapidly after children's vocabulary reaches a minimum level of approximately 300-400 words. This finding supports the critical mass hypothesis, as grammar ability depended upon the size of the vocabulary more than age or exposure to the language.

The results from the present study can also be considered within the single-mechanism framework. Although the expected correlations between PSTM and vocabulary and grammar were present, there were also unexpectedly strong correlations with the WM measures. If both grammar and vocabulary learning are mediated by the same unitary language learning mechanism, both should depend on the same set of cognitive abilities (phonological memory and working memory) to approximately the same amount. This pattern of results appears in the present study. The skewness in the verb and adjective plurality scores can also be explained by the fact that participants had not learned enough vocabulary to support the generalization of such grammatical structures. With more exposure to vocabulary, it is possible that the grammar abilities of the participants would have developed more, become more normally distributed, and correlated more consistently with the memory measures.

The single-mechanism approach also accounts for the strong relationship found between the grammar and vocabulary measures in this study, while not discrediting the importance of phonological memory and working memory for grammar learning. The relationship between vocabulary and grammar would be expected, because in this framework both language skills are integrated and rely on the same underlying processes. Again, this is in line with the findings from the present study: that phonological memory and working memory are equally important for vocabulary and grammar development. The single-mechanism framework provided by Bates and Goodman (1997) is also similar to the view expressed by Spiedel (1993) and Ellis (1996). Generally, they argue that vocabulary is the abstraction of patterns over a set of phonemes, while grammar learning is the abstraction of patterns over a set of morphemes. It may be that vocabulary and grammar learning are more similar than different, and that both types of memory are equally important for the two.

Conclusions

The results of the current study complement existing data nicely by adding the importance of WM to the framework of knowledge about L2 vocabulary and grammar learning. PSTM and WM were found to play an almost equally important role in both vocabulary and grammar acquisition, with the grammatical abilities somewhat dependent upon vocabulary knowledge. The almost equal contributions from two traditionally distinct cognitive abilities suggests that language skills may be mutually dependent and rely on the same, or very similar, underlying learning and processing mechanism(s).

Future research should continue to explore the relative importance of PSTM and WM on the acquisition of various language skills. Given the surprisingly strong correlations with WM found in the current study, replication will be important for determining whether this is a robust finding. Reliability statistics will also be calculated for all the measures used in the current study. Paradigms that use languages of varying complexity should also be developed to allow participants to gain more experience with the language and acquire a larger vocabulary. Although an artificial language has the advantage of allowing for complete control over the stimuli and the learning process, it is also limiting in the amount of material that can be taught to participants and in the developmental analyses that can be performed. Replicating the study with more complex artificial language systems, such as Friederici's Brocanto (Friederici, Steinhauer, & Pfeifer, 2002; Opitz & Friederici, 2004), as well as true beginner students in foreign language programs, would allow for converging evidence from various language systems. Varying the way in which the language is taught and in which feedback is given is also important for determining whether the relationships found in this study are replicated in both instructed and naturalistic learning. This would also serve to test whether the critical mass hypothesis holds up for language systems of varying complexity and for different learning conditions.

Investigating the way in which participants process grammatical cues during language learning will also be important for future research. Experiments that investigate foreign-language sentence processing, and specifically the processing of grammatical morphology, could help identify which structures participants attend to and investigate variations in this behavior that may depend on individual differences in cognitive abilities. Comprehension of grammatical structures could also be tested using eye-tracking and a visual-world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Eye-movement data both in reading and spoken foreign language processing may allow for more detailed analysis of these comprehension processes (e.g., Juffs & Harrington, 1995). Research on cognitive abilities and their influences on both language acquisition and language processing has many promising directions for future development.

References

Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech and Hearing Research, 38,* 403-414.

Adams, A. M., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *Quarterly Journal of Experimental Psychology, 49A,* 216-233.

Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders, 35,* 95-116.

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research, 49,* 970-983.

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

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19,* 537-552.

Audacity Team. (2008). Audacity (Version 1.2.6) [Computer software]. Retrieved September 7, 2008 from http://audacity.sourceforge.net/

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A. D. (1993). Short-term phonological memory and long-term learning: A single case study. *European Journal of Cognitive Psychology, 5,* 129-148.

Baddeley, A. D. (1998). Working memory. *C.R. Académie des sciences Paris, Sciences de la Vie / Life Sciences, 321*, 167-173.

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

Baddeley, A. D., & Gathercole, S. E. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439-453.

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation: Vol. 8.* (pp. 47-90). New York: Academic Press.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27,* 586-595.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes, 12,* 507-584.

Berquist, B. (1997). Individual differences in working memory span and L2 proficiency: Capacity or processing efficiency? In A. Sorace, C. Heycock, & R. Shillcock (Eds.) *Proceedings of the GALA '97 Conference on Language Acquisition.* (pp. 468-473). Human Communication Research Center, University of Edinburgh.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44,* 235-249.

Bishop, D. V. M. (1989). *Test for the reception of grammar* (2nd ed.). Manchester, England: Chapel.

Blake, J., Austin, W., Canon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development, 17,* 91-107.

Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., et al. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language, 29,* 591-610.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32,* 76-95.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Cohen, N. J. (2002). Developmental language disorders. In P. Howling & O. Udwin (Eds.), *Outcomes in neurodevelopmental and genetic disorders* (pp. 26-55). New York: Cambridge University Press.

Daneman M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior, 19,* 450-466.

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology, 17,* 367-378.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422-433.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55,* 1-25.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91-126.

Ellis, N.C., & Beaton, A. (1993a). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology, 46A,* 533-558.

Ellis, N. C., & Beaton, A. (1993b). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning, 43*, 559-617.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology, 49A,* 234-250.

Fan, M. Y. (2003). Frequencey of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *The Modern Language Journal, 87*, 222-241.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307-317.

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology, 20,* 439-462.

Fortkamp, M. B. M. (1999). Working memory capacity and aspects of L2 speech production. *Communication & Cognition*, *32,* 259-296.

French, L. M. (2006). *Phonological working memory and second language acquisition: A developmental study of Francophone children learning English in Quebec.* Lewiston, NY: Edwin Mellen Press.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29,* 463-487.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 529-534.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51,* 136-158.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and Language*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25,* 84-95.

Gathercole, S. E., & Pickering, S. J. (1999). Estimating the capacity of phonological short-term memory. *International Journal of Psychology, 34,* 378-382.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology, 54A,* 1-30.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13,* 65-77.

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

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Haywood, J.A., & Nahmad, H. M. (1965). *A new Arabic grammar of the written language.* UK: Lund Humphries.

Jeffries, E., Frankish, C., & Ralph, M. A. L. (2006). Lexical and semantic influences on item and order memory in immediate serial recognition: Evidence from a novel task. *Quarterly Journal of Experimental Psychology*, *59,* 949-964.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21,* 60-99.

Juffs, A., & Harrinton, M. (1995). Parsing effects in second language processing: Subject and object asymmetries "wh"-extraction. *Studies in Second Language Acquisition, 17,* 483-516.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11,* 261-271.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* New York: Pergamon Press.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21,* 339-366.

Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory and Cognition, 21,* 176-183.

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 46,* 1138-1153.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34,* 383-388.

Masson, M. E. J., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology, 75,* 314-318.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27,* 377-402.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29,* 557-582.

Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience, 24,* 8436-8440.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30,* 331-347.

Pinker, S. (1991). Rules of language. *Science, 253,* 530-535.

Plunkett, K., & Marchman, V. (1993). From rote learning to system-building—Acquiring verbal morphology in children and connectionist models. *Cognition, 48,* 21-69.

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

Robinson, B. F., Mervis, C. B., & Robinson, B. W. (2003). The roles of verbal short-term memory and working memory in the acquisition of grammar by children with Williams syndrome. *Developmental Neuropsychology, 23,* 13-31.

Salthouse, T. A. (1992). Shifting levels of analysis in the investigation of cognitive aging. *Human Development, 35,* 321-342.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *The Modern Language Journal, 79,* 15-28.

Schneider, W., Eschman, A., & Zuccolutto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools, Inc.

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

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, *45A*, 21-50.

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155-172.

Sleve, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17,* 675-681.

Speidel, G. E. (1993). Phonological short-term memory and individual differences in learning to speak: A bilingual case study. *First Language*, *13*, 69-91.

Stark, R. E., & Tallal, P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46,* 114-122.

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30,* 79-99.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268,* 1632-1634.

Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research, 48,* 1481-1495.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

van den Noort, M. W. M. L., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist, 11*, 289-296.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1256-1271.

Watkins, R. V. (1994). Specific language impairments in children: An introduction. In R. Watkins & M. Rice (Eds.), *Specific language impairments in children* (pp. 1-17). Baltimore, MD: Paul H. Brookes.

Weatherick, N. E. (1975). The role of semantic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 471-480.

Wechsler, D. (1944). *The measurement of adult intelligence.* Baltimore: Williams & Watkins.

Wells, B. (1995). Phonological considerations in repetition tests. *Cognitive Neuropsychology, 12,* 847-855.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53,* 67-121.

Willis, C. S., & Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory, 9,* 349-363.

Appendix A

Non-word Repetition Stimuli
(taken from Gathercole et al., 2001)

barp
bon
bordge
borp
bup
charn
chim
chorg
chull
darp
derb
doob
doog
dorch
dordge
dorl
dort
gadge
garn
garm
geed
gerk
gerp
goot
gop
jarm
jerg
jert
jit
keb
kern
kig
mab
marn
mun
nam
narb
narg
narp
neeb
neeg
nerch
noog
norb
norg
nug
pab
parn
peb
ped
pem
pim
putch
tam
teeb
teeg
tep
terdge
tudge
turg
barg
cham
darch
gerb
jat
kerm
larm
meb
nart
padge
teep
bick

Appendix B

Non-word Recognition Stimuli
(taken from Gathercole et al., 2001 and also used in O’Brien et al., 2006)

barch
teeg
nup
goot
kib
dern
putch
geed
mot
chen
ped
kig
merd
garp
tam
pib
teck
barp
mitch
bon
derp
choom
kerp
lork
nug
garn
peem
targ
gerk
bock
chull
goot
barg
mern
nuck
tep
loog
jal
deech
kerm
meb

keech
jarn
mep
bick
koom
tard
modge
jup
gick
lod
tudge
jick
narb
garm
bordge
chud
neeg
dack
keb
larm
terdge
joop
leck
narg
chim
peeb
parn
mab
dordge
nog
cheem
jit
norg
boodge
tud
lig
pab
dort
tidge
mup
chan

gerb
noog
dorl
torm
pag
jeck
derb
koll
bup
korp
teeb
nool
jark
pim
gerch
padge
narp
mon
chut
goob
ged
toock
borp
lidge
jerg
dop
nart
gub
nerch
mord
cham
bool
lub
teep
charn
neeg
kom
jeel
gadge
lerb
mun

tob
gan
dorch
chool
juck
nord
pem
gell
nerg
lud
marn
meech
doob
jat
leem
chordge
jert
doog
darp
chig
nam
peb
gop
jooch
lart
terch
jarm
jick
gerb
chorg
mal
tooch
lon
norb
charb
dook
jep

Appendix C

Listening Span Stimuli
(taken from Harrington & Sawyer, 1992)

Grammaticality: G = Grammatical; U = Ungrammatical

He played baseball all day at the park and got a sore arm.	G
The clerk in the department presents the put a in bag.	U
I saw a child and her near playing the father river ball.	U
His younger brother played guitar in a rock and roll band.	G
Suddenly the taxi opened its the front in door the of bank.	U
The last thing he did was nice to a hot take bath.	U
Her best memory of England was the Tower of London bell.	G
At the very top of the tall a small sat tree bird.	U
She took a deep breath and reached into the rusty box.	G
The state of Wisconsin is famous for its butter and cheese.	G
He overslept and missed all morning the of economics class.	U
The first thing he does every golf is morning a swing club.	U
Popular foods in the watermelon and summer sweet are corn.	U
The boy was surprised to learn that milk came from a cow.	G
The only thing left in the kitchen cupboard was a broken cup.	G
The birthday party began in the morning and lasted all day.	G
The young woman and her thought boyfriend a saw they dog.	U
There was nothing left to do except leave and lock the door.	G
In order to attend the dinner needed buy to she a dress.	U
The woman screamed and slapped the old man in the face.	G

She leaned over the candle on and hair her caught fire.	U
The drinks were all gone and all that remained was the food.	G
He quickly drank some of washed the milk then the and glass.	U
He looked across the room and saw a person holding a gun.	G
The hunting knife was so sharp his it right that cut hand.	U
She soon realized that the man forgot to leave the room key.	G
The saw that he brought was not strong enough for the lock.	G
The first driver out in the morning always picks up the mail.	G
All that remained in the lunch one salted was box nut.	U
The boat engine would not run was of it out because oil.	U
The letter said to come to the market to claim the prize.	G
It was a very simple and fish boiled of meal rice.	U
They decided to take an afternoon break by the large rock.	G
He wanted to leave his bags hotel in jacket the and room.	U
There were so many people that I couldn’t find a seat.	G
He opened the bottom pulled drawer out and a shirt.	U
The skiing was so wonderful didn’t that he the mind snow.	U
They knew that it was impolite to eat spaghetti with a spoon.	G
The season that people often associate with love is spring.	G
The letter was lost because it did not have a postage stamp.	G
The people in northern Europe always like to travel by train.	G
All morning the two children and under sat a talked tree.	U
At night the prisoners hole through in the escaped a wall.	U

Appendix D

The Artificial Language

<u>Nouns</u>

English	Singular	Dual	Plural
apple	bock	zi-bock	na-bock
baby	charb	zi-charb	na-charb
egg	dack	zi-dack	na-dack
house	garp	zi-garp	na-garp
duck	jat	zi-jat	na-jat
snake	kerm	zi-kerm	na-kerm
floor	mern	za-mern	no-mern
ball	nart	za-nart	no-nart
puddle	padge	zi-padge	na-padge
road	terch	zi-terch	na-terch
clock	boodge	zi-boodge	na-boodge
turtle	chut	zi-chut	na-chut
table	dook	zi-dook	na-dook
book	jick	zi-jick	na-jick
kite	lart	zi-lart	na-lart
cow	lork	zi-lork	na-lork
lake	mup	za-mup	no-mup
bucket	nool	za-nool	no-nool
dog	peeb	zi-peeb	na-peeb
fish	targ	zi-targ	na-targ
man	nog	za-nog	no-nog

<u>Verbs</u>

English	Singular	Dual	Plural
run	barch	ta-barch	moo-barch
jump	cheem	ta-cheem	moo-cheem
throw	dern	ta-dern	moo-dern
fall	gell	ta-gell	moo-gell
eat	jep	ta-jep	moo-jep
push	kerp	ta-kerp	moo-kerp
is (to be)	mord	too-mord	mi-mord
pull	nerg	too-nerg	mi-nerg
catch	pag	ta-pag	moo-pag
swim	tidge	ta-tidge	moo-tidge

<u>Adjectives</u>

cham	big	mep	blue
gub	small	leck	red

<u>Prepositions</u>

kib	in, inside, on
nug	to
tem	along, through

Appendix E

Foreign Language Sentence Stimuli

<u>Set 1</u>

The duck pulls the red apple along the road.	Jat nerg bock leck tem terch.
The baby pulls the bucket.	Charb nerg nool.
The man throws the egg.	Nog dern dack.
The dog runs along the red road.	Peeb barch tem terch leck.
The fish swims in the lake.	Targ tidge kib mup.
The big cow is on the table.	Lork cham mord kib dook.
The baby eats the snake.	Charb jep kerm.
The turtle pushes the clock along the floor.	Chut kerp boodge tem mern.
The ball jumps in the small house.	Nart cheem kib garp gub.
The bucket catches the egg.	Nool pag dack.
The blue dog falls into the puddle.	Peeb mep gell kib padge.
The duck eats the kite.	Jat jep lart.
The apple is on the small table.	Bock mord kib dook gub.
The turtle catches the big book.	Chut pag jick cham.
The floor falls in the blue house.	Mern gell kib garp mep.

<u>Set 2</u>

<u>15 English → Artificial Language:</u>

The snake throws multiple eggs into the lake.	Kerm dern na-dack kib mup.
Multiple turtles push the big cow.	Na-chut moo-kerp lork cham.
Multiple men catch two balls.	No-nog moo-pag za-nart.

Two ducks jump to the clock.	Zi-jat ta-cheem nug boodge.
Two fish swim in the puddle.	Zi-targ ta-tidge kib padge.
Two babies eat multiple kites.	Zi-charb ta-jep na-lart.
Two clocks pull the book.	Zi-boodge too-nerg jick.
The man catches the small dog.	Nog pag peeb gub.
The red apple falls to the floor.	Bock leck gell nug mern.
The table is in the road.	Dook mord kib terch.
The cow runs to the kite.	Lork barch nug lart.
The house falls to the blue road.	Garp gell nug terch mep.
Two babies jump along the floor.	Zi-charb ta-cheem tem mern.
Multiple men throw multiple turtles.	No-nog moo-dern na-chut.
Two ducks pull multiple kites.	Zi-jat too-nerg na-lart.

<u>15 Artificial Language → English:</u>

The duck pushes the apple.	Jat kerp bock.
Multiple eggs fall into two lakes.	Na-dack moo-gell kib za-mup.
The turtle runs through multiple blue puddles.	Chut barch tem na-padge no-mep.
Multiple books fall in the house.	Na-jick moo-gell kib garp.
Two cows eat the table.	Zi-lork ta-jep dook.
The man throws two clocks.	Nog dern zi-boodge.
The turtle jumps to two red balls.	Chut cheem nug za-nart zi-leck.
Multiple books are in the bucket.	Na-jick mi-mord kib nool.
The dog is in the lake.	Peeb mord kib mup.
Two fish swim in two buckets.	Zi-targ ta-tidge kib za-nool.

Two snakes swim in the puddle.	Zi-kerm ta-tidge kib padge.
Multiple dogs run to the ball.	Na-peeb moo-barch nug nart.
Two big cows jump into two buckets.	Zi-lork zi-cham ta-cheem kib za-nool.
Multiple floors fall in the house.	No-mern moo-gell kib garp.
Multiple blue tables are in the road.	Na-dook no-mep mi-mord kib terch.

Generalization Test Set

New Sentences:

Two apples fall to the floor.	Zi-bock ta-gell kib mern.
The cow eats two small kites.	Lork jep zi-lart zi-gub.
Multiple snakes swim in the small lake.	Na-kerm moo-tidge kib mup gub.
The dog catches multiple balls in the house.	Peeb pag no-nart kib garp.
Two red turtles run through two puddles.	Zi-chut zi-leck ta-barch tem zi-padge.
Multiple turtles pull the clock.	Na-chut mi-nerg boodge.
Two dogs run along the road.	Zi-peeb ta-barch tem terch.
Multiple big ducks jump to the baby.	Na-jat na-cham moo-cheem nug charb.
Multiple snakes swim in multiple lakes.	Na-kerm moo-tidge kib no-mup.
The fish throws the apple.	Targ dern bock.
Two babies eat many apples.	Zi-charb ta-jep na-bock.
Two eggs fall on two cows.	Zi-dack ta-gell kib zi-lork.
Two big eggs are in two buckets.	Zi-dack zi-cham too-mord kib za-nool.
The man runs to two houses.	Nog barch nug zi-garp.
Multiple babies fall into the puddle.	Na-charb moo-gell kib padge.
Multiple cows jump on two blue floors.	Na-lork moo-cheem kib za-mern za-mep.

Multiple men pull two tables along two roads.	No-nog mi-nerg zi-dook tem zi-terch.
Multiple snakes eat multiple clocks.	Na-kerm moo-jep na-boodge.
Two men push the ball along the floor.	Za-nog ta-kerp nart tem mern.
Multiple fish throw two red books.	Na-targ moo-dern zi-jick zi-leck.
The kite pulls multiple buckets.	Lart nerg no-nool.
The bucket jumps to the table.	Nool cheem nug dook.
The dog eats multiple books.	Peeb jep na-jick.
Multiple small snakes swim in the lake.	Na-kerm na-gub moo-tidge kib mup.

<u>Sentences from Set 1:</u>

The duck pulls the red apple along the road.	Jat nerg bock leck tem terch.
The man throws the egg.	Nog dern dack.
The dog runs along the red road.	Peeb barch tem terch leck.
The fish swims in the lake.	Targ tidge kib mup.
The big cow is on the table.	Lork cham mord kib dook.
The baby eats the snake.	Charb jep kerm.
The turtle pushes the clock along the floor.	Chut kerp boodge tem mern.
The bucket catches the egg.	Nool pag dack.
The blue dog falls into the puddle.	Peeb mep gell kib padge.
The duck eats the kite.	Jat jep lart.
The apple is on the small table.	Bock mord kib dook gub.
The turtle catches the big book.	Chut pag jick cham.
The floor falls in the blue house.	Mern gell kib garp mep.

<u>Sentences from Set 2:</u>

The snake throws multiple eggs into the lake.	Kerm dern na-dack kib mup.
Multiple turtles push the big cow.	Na-chut moo-kerp lork cham.
Two ducks jump to the clock.	Zi-jat ta-cheem nug boodge.
Two fish swim in the puddle.	Zi-targ ta-tidge kib padge.
Two babies eat multiple kites.	Zi-charb ta-jep na-lart.
Multiple men throw multiple turtles.	No-nog moo-dern na-chut.
Multiple eggs fall into two lakes.	Na-dack moo-gell kib za-mup.
The turtle runs through multiple blue puddles.	Chut barch tem na-padge no-mep.
Multiple books fall in the house.	Na-jick moo-gell kib garp.
The dog is in the lake.	Peeb mord kib mup.
Two fish swim in two buckets.	Zi-targ ta-tidge kib za-nool.
Two snakes swim in the puddle.	Zi-kerm ta-tidge kib padge.
Two big cows jump into two buckets.	Zi-lork zi-cham ta-cheem kib za-nool.

Author Note

Katherine I. Martin, Department of Psychology, University of Michigan.

I would like to thank my advisor, Dr. Nick C. Ellis, for his countless hours of help and suggestions for this study. This project developed out of an idea originally formulated during his Second Language Acquisition class, and never would have come to be without his constant guidance. I especially want to thank him for the many times that he made it subtly clear that it wasn’t quite good enough yet, and that there was more work to be done.

I also want to thank the post-doc, Heather Taylor, who has worked with Dr. Ellis and me for the past year. She was also a constant source of encouragement and a miracle worker in the last-minute editing help and feedback she provided.

Andrew Mueller and Andrew Lorelli have earned my eternal gratitude for assisting in the transcribing and scoring of participant utterances. Without their help, I would still be transcribing a year from now. Elizabeth Bogner and Ann Cassel also deserve my thanks for their help in creating the illustrations for all the vocabulary words and sentence stimuli.

Finally, I would like to thank my family and friends, especially Sean Strassenburg, for putting up with my constant obsession with this project for over a year. Know that I can now return to sanity and something resembling a normal life.

Table 1

Description of Variables of Interest and Shorthand

Variable	Description of Variable	Shorthand
Non-word Repetition Total	Total number of phonemes correctly recalled during the Non-word Repetition task overall	NR Total
Non-word Repetition Maximum	Maximum number of phonemes correctly recalled during any one Non-word Repetition trial	NR Max
Nonword Recognition Accuracy	Number of same/different judgments correctly made during the Non-word Recognition trials	NW Recognition/ NW Rec
Listening Span Item	Total number of items correctly recalled during Listening Span trials, disregarding order	LS Item
Listening Span Order	Total number of items correctly recalled in the correct order during Listening Span trials	LS Order
Test Productions variables	Scores obtained for various measures during the final, test phase of the study; includes repeated stimuli from Set 1, Set 2, and novel Generalization stimuli combined	TP
Generalization Set variables	A subset of the Test Productions scores; those scores earned on stimuli that were novel to the final phase of the study	GS
Production variables	A subset of variables from any stimuli set; only those scores earned on stimuli that were translated into (therefore producing) the foreign language	Prd

Variable	Description of Variable	Shorthand
Comprehension Variables	A subset of variables from any stimuli set; only those scores earned on stimuli that were heard in the foreign language (therefore comprehended) and translated into English	Cmp
Set 1	Scores earned on stimuli originally presented during the first phase of the study	None used
Set 2	Scores earned on stimuli originally presented during the second phase of the study, while participants were introduced to the plural forms	None used
Noun/Noun Plural	Scores earned on nouns or noun plural markings in a particular set of stimuli	N/N Plural
Verb/Verb Plural	Scores earned on verbs or verb plural markings in a particular set of stimuli	V/V Plural
Adjective/Adjective Plural	Scores earned on adjectives or adjective plural markings in a particular set of stimuli	Adj/Adj Plural or A/A Plural
Preposition	Scores earned on prepositions in a particular set of stimuli	Prep
Noun/adjective word order	Scores earned on the correct ordering of nouns and adjectives in a particular set of stimuli	N/A Word Order or N/A WO
Composite variables	Variables that combine separate scores from within a number of a specific stimuli set.	Cpst

Variable	Description of Variable	Shorthand
	Composite vocabulary scores include the noun, verb, adjective, and preposition scores from that specific set of stimuli. Composite grammar scores include the noun, verb, and adjective plural markings, as well as the noun/adjective word order scores from that specific set of stimuli	
Tell me about the grammar.	The number of points participants earned from the descriptions of the overall grammatical structure of the foreign language; taken from the second post-experimental questionnaire	Describe grammar
What was the rule for forming plurals?	The number of points participants earned from the descriptions of the rule for forming plurals in the foreign language; taken from the second post-experimental questionnaire	Rule for plurals?
Were there any irregular changes or mutations?	The number of points participants earned from their ability to identify and describe the irregular vowel mutations that occurred in the language; taken from the second post-experimental questionnaire	Irregular changes?
Total Years of Study	The total number of years participants had studied foreign languages	Total years/Total Years Studied

Table 2

Summary Data for Variables of Interest

Variable	N	Mean	SD	Min	Max	Possible
NR Total	47	156.62	22.70	105	202	251
NR Max	47	13.74	1.91	11	19	22
NW Recognition	47	14.91	2.86	8	20	24
LS Item	47	36.13	4.27	23	42	42
LS Order	47	25.49	4.10	13	31	31
Age	47	22.43	7.19	18	55	N/A
Total Years Studied	47	6.76	3.74	1	16.5	N/A
Tell me about the grammar.	40	1.20	.91	0	4	N/A
Rule for forming plurals?	40	2.35	1.61	0	7	N/A
Irregular changes/mutations?	40	1.40	1.48	0	5	N/A
TP Noun	40	56.32	18.87	19.96	91.28	106
TP Noun Plural	40	25.53	12.30	0	45	49
TP Verb	40	16.74	9.64	1	42	50
TP Verb Plural	40	1.16	2.49	0	12	13
TP Adjective	40	5.68	3.93	0	17	18
TP Adjective Plural	40	.78	1.08	0	4	6
TP Preposition	40	10.75	7.16	0	29	34
TP N/A Word Order	40	2.43	2.15	0	7	7
GS Noun	40	26.66	9.36	11.62	44.96	51
GS Noun Plural	40	15.89	7.82	.50	30	32

Variable	N	Mean	SD	Min	Max	Possible
GS Verb	40	7.29	4.34	1	17.98	24
GS Verb Plural	40	.59	1.28	0	6	7
GS Adjective	40	2.51	1.86	0	8	8
GS Adjective Plural	40	.56	.89	0	3	5
GS Preposition	40	4.41	3.21	0	14	16
GS N/A Word Order	40	1.85	1.61	0	5	5
GS Production Noun	40	14	4.43	5.98	22.98	25
GS Production Noun Plural	40	8.41	3.99	0	14	14
GS Production Verb	40	4.44	2.98	0	12	12
GS Production Verb Plural	40	.59	1.28	0	6	7
GS Production Adjective	40	1.99	1.39	0	5	5
GS Production Adj Plural	40	.56	.89	0	3	5
GS Production Preposition	40	1.01	1.99	0	7	7
GS Production N/A Word Order	40	1.85	1.61	0	5	5
GS Comprehension Noun	40	12.66	5.62	0	22	26
GS Comprehension Noun Plural	40	7.49	4.26	0	16	18
GS Comprehension Verb	40	2.85	2.17	0	8	12
GS Comprehension Adjective	40	.53	.72	0	3	3

Variable	N	Mean	SD	Min	Max	Possible
GS Comprehension						
Preposition	40	3.40	1.63	0	7	9
TP Grammar Composite	40	29.90	15.23	1.50	63.50	75
GS Grammar Composite	40	18.90	9.67	1.50	40.25	49
GS Production Grammar						
Composite	40	11.41	6	1	28	31
Set 2 Grammar Composite	40	10.72	5.62	0	22.25	25
TP Vocabulary Composite	40	89.48	36.19	31.24	175.56	208
Set 1 Vocabulary Composite	40	25.79	10.72	7.98	54.30	56
Set 2 Vocabulary Composite	40	22.82	9.36	8.64	43.98	53
GS Vocabulary Composite	40	40.87	16.98	14.62	78.64	99
GS Production Vocabulary						
Composite	40	21.44	9.32	9.30	45.64	49
GS Comprehension Vocab						
Composite	40	19.44	8.70	0	37	50
Set 1 Vocab Production[a]	47	14.12	12.42	0	47.30	64
Set 1 Vocab Review[a]	40	14.33	7.45	1	32.64	36
Set 2 Vocab Production[a]	40	38.29	12.16	15.96	87.62	58
Set 2 Grammar Composite[a]	40	9.60	5.37	0	22.50	27

Note. [a]These scores are taken not from the final phase of the study but rather from the first presentation of these stimuli sets, during the learning phases of the study. They are used as measures of the development of vocabulary and grammar, as opposed to their final attainment.

Table 3

Correlations between Memory Measures and Final Test Phase Vocabulary Scores

Variable	NR Total	NR Max	NW Rec	LS Item	LS Order
NR Total	--				
NR Max	.847**	--			
NW Rec	.346*	.347*	--		
LS Item	.057	.100	.079	--	
LS Order	-.020	.055	.080	.971**	--
TP Noun	.241	.371*	.285	.293	.313*
TP Verb	.171	.284	.401*	.334*	.371*
TP Adj	.112	.258	.278	.198	.216
TP Prep	.395*	.476**	.251	.302	.295
GS Noun	.233	.378*	.325*	.264	.288
GS Verb	.182	.288	.403**	.349*	.376*
GS Adj	.166	.271	.279	.164	.186
GS Prep	.383*	.445**	.275	.327*	.320*
GS Noun Prd	.234	.354*	.139	.252	.260
GS Verb Prd	-.050	.120	.253	.323*	.383*
GS Adj Prd	.239	.352*	.261	.185	.195
GS Prep Prd	.240	.334*	.283	.299	.325*
GS Noun Cmp	.204	.351*	.432**	.241	.274
GS Verb Cmp	.433**	.413**	.474**	.256	.227
GS Adj Cmp	-.033	.020	.218	.066	.105

Variable	NR Total	NR Max	NW Rec	LS Item	LS Order
GS Prep Cmp	.462**	.469**	.195	.279	.234
TP Composite	.262	.391*	.335*	.323*	.344*
Set 1 Composite	.339*	.446**	.345*	.278	.277
Set 2 Composite	.142	.283	.239	.359*	.405*
GS Composite	.266	.396*	.365*	.314*	.336*
GS Prd Composite	.182	.330*	.243	.314*	.344*
GS Cmp Composite	.323*	.419**	.451**	.277	.286

Note. See Table 2 for explanation of shorthand variable labels. All scores (including Set 1 and Set 2) were taken from the presentations of those stimuli during the final test phase of the study. N = 40.

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

Table 4

Correlations between Grammar and Composite Vocabulary Measures, Final Test Phase

Variable	1	2	3	4	5	6	7	8	9	10
1. TP Vocab	--									
2. Set 1 Vocab	.969**	--								
3. Set 2 Vocab	.968**	.913**	--							
4. GS Vocab	.986**	.930**	.935**	--						
5. GS Prd Vocab	.934**	.870**	.901**	.946**	--					
6. GS Cmp Vocab	.924**	.882**	.860**	.938**	.775**	--				
7. TS Grammar	.722**	.663**	.751**	.708**	.670**	.663**	--			
8. Set 2 Grammar	.683**	.633**	.727**	.655**	.612**	.621**	.970**	--		
9. GS Grammar	.718**	.655**	.737**	.710**	.675**	.663**	.990**	.927**	--	
10. GS Prd Grammar	.756**	.694**	.771**	.749**	.738**	.670**	.956**	.900**	.960**	--

Note. See Table 2 for explanation of shorthand variable labels. All scores (including Set 1 and Set 2) were taken from the presentations of those stimuli during the final test phase of the study and are composite variables (see Table 2 for explanation). N = 40.

** $p < .01$.

Table 5

Correlations between Memory Measures and Grammar Scores

Variable	NR Total	NR Max	NW Rec	LS Item	LS Order
TP N Plural	.192	.373*	.189	.405**	.445**
TP V Plural	.112	.122	.278	.233	.276[a]
TP A Plural	.029	.010	.008	-.037	.069
TP N/A WO	.204	.332*	.311	.307	.354*
GS N Plural	.239	.420**	.221	.346*	.377*
GS V Plural	.079	.115	.185	.242	.276[b]
GS A Plural	.034	.037	.079	-.052	.075
GS N/A WO	.134	.251	.314*	.275	.332*
GS N Plural Prd	.191	.363*	.187	.341*	.361*
GS N Plural Cmp	.260	.431**	.230	.316*	.354*
TP Grammar Cpst	.204	.368*	.235	.406**	.460**
Set 2 Grammar Cpst	.152	.294	.170	.472**	.528**
GS Grammar Cpst	.229	.400*	.263	.353*	.404**
GS Prd Grammar Cpst	.185	.339*	.260	.345*	.400*

Note. See Table 2 for explanation of shorthand variable labels. All scores (including Set 2) were taken from the presentations of those stimuli during the final test phase of the study. N = 40.

[a]This non-significant correlation was found using Pearson's correlations. When Spearman's rho was used, the correlation was significant: $r = .367$, $p < .05$. [b]This non-significant correlation was found using Pearson's correlations. When Spearman's rho was used, the correlation was significant: $r = .336$, $p < .05$.

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

Table 6

Questionnaire Data Correlations

Variable	Describe grammar[a]	Rule for plurals[b]	Irregular changes[c]
LS Order	-.017	.334*	.130
TP Noun	.295	.599**	.238
TP N Plural	.167	.560**	.229
TP Verb	.317*	.531**	.139
TP V Plural	.392*	.466**	.291
TP Adjective	.367*	.375*	.174
TP Preposition	.375*	.546**	.269
TP N/A Word Order	.322*	.557**	.163
GS Noun	.291	.558**	.224
GS N Plural	.170	.535**	.273
GS Verb	.289	.559**	.126
GS V Plural	.257	.399*	.313*
GS Adjective	.351*	.460**	.153
GS Preposition	.442**	.505**	.194
GS N/A Word Order	.301	.555**	.166
GS Noun Prd	.262	.484**	.325*
GS N Plural Prd	.171	.462**	.268
GS Verb Prd	.240	.440**	.152
GS V Plural Prd	.257	.399*	.313*
GS Adjective Prd	.311	.390*	.285

Variable	Describe grammar[a]	Rule for plurals[b]	Irregular changes[c]
GS Preposition Prd	.404**	.524**	.204
GS N/A WO Prd	.301	.555**	.166
GS Noun Cmp	.279	.547**	.117
GS N Plural Cmp	.153	.550**	.250
GS Verb Cmp	.249	.515**	.043
GS Adjective Cmp	.307	.437**	-.155
GS Preposition Cmp	.376*	.356*	.134
TP Grammar Cpst	.251	.610**	.267
Set 2 Grammar Cpst	.272	.629**	.226
GS Grammar Cpst	.229	.581**	.287
GS Prd Grammar Cpst	.261	.546**	.286
TP Vocab Cpst	.352*	.603**	.234
Set 1 Vocab Cpst	.330*	.543**	.221
Set 2 Vocab Cpst	.338*	.627**	.270
GS Vocab Cpst	.356*	.596**	.209
GS Vocab Prd Cpst	.334*	.541**	.289
GS Vocab Cmp Cpst	.338*	.584**	.098

Note. See Table 2 for explanation of shorthand variable labels. All scores (including Set 2) were taken from the presentations of those stimuli during the final test phase of the study. N = 40.

[a]Tell me about the grammar. [b]What was the rule for forming plurals?. [c]Were there any irregular changes or mutations in any context? (See Table 2 for full description of these variables).

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

Table 7

Regression Results from Generalization Set Grammar Production

	Predictors	β	p	Adj. R^2	F	p
Step 1	LS Order	.400	.011	.138	7.219	.011
Step 2	LS Order	.402	.007			
	NR Max	.342	.019	.237	7.074	.003

Note. β = Standardized regression coefficient, beta. Adj. R^2 = Adjusted R^2 for the model as a whole.

Table 8

Regression Results from Three Final Grammar Scores, Including Vocabulary

Dependent Variable	Predictors	β	p	Adj. R^2	F	p
Generalization Set						
Comp. Grammar[a]	Gen. Set Vocab[d]	.710	.000	.491	38.675	.000
Generalization Set						
Grammar Prod.[b]	Gen. Set Vocab	.749	.000	.560	48.428	.000
Generalization Set						
N Plural Cpst.[c]	Gen. Set Vocab	.558	.000	.293	17.156	.000

Note. β = Standardized regression coefficient, beta. Adj. R^2 = Adjusted R^2 for the model as a whole.

[a]Generalization Set composite grammar score. [b]Generalization Set grammar production score. [c]Generalization Set noun plural comprehension score. [d]Generalization Set composite vocabulary score.

Table 9

Regression Results from Four Final Grammar Scores, Including Age and Years of Foreign Language Study

Dependent Variable	Step	Predictors	β	p	Adj. R^2	F	p
Test Productions							
Grammar Cpst. [a]	1	Total Years[e]	-.285	.079			
		Age	-.034	.832	.032	1.650	.206
	2	Total Years	-.177	.125			
		Age	-.032	.773			
		Test Productions Vocab	.695	.000	.516	14.876	.000
	3	Total Years	-.178	.106			
		Age	.023	.836			
		Test Productions Vocab	.610	.000			
		LS Order	.247	.042	.559	13.354	.000
Generalization Set							
Grammar Cpst. [b]	1	Total Years	-.327	.042			
		Age	-.011	.945	.059	2.218	.123
	2	Total Years	-.222	.054			
		Age	-.009	.933			
		Test Productions Vocab	.684	.000	.527	15.494	.000
Generalization Set							
Grammar Prd. [c]	1	Total Years	-.344	.032			
		Age	-.042	.785	.072	2.517	.094
	2	Total Years	-.233	.030			
		Age	-.041	.690			
		Test Productions Vocab	.720	.000	.595	20.107	.000

Dependent Variable	Step	Predictors	β	p	Adj. R^2	F	p
Generalization Set							
N Plural Cpst. [d]	1	Total Years	-.258	.113			
		Age	.035	.824	.018	1.352	.271
	2	Total Years	-.175	.206			
		Age	-.037	.787			
		Test Productions Vocab	.538	.000	.296	6.467	.001

Note. β = Standardized regression coefficient, beta. Adj. R^2 = Adjusted R^2 for the model as a whole.

[a]Test Productions composite grammar score. [b]Generalization Set composite grammar score. [c]Generalization Set grammar production score. [d]Generalization Set noun plural comprehension score. [e]Total years of foreign language study.

Table 10

Regression Results from Test Productions Vocabulary Scores

	Predictors	β	p	Adj. R^2	F	p
Step 1	Total Years[a]	-.154	.348			
	Age	-.002	.990	-.029	.452	.640
Step 2	Total Years	-.151	.324			
	Age	-.060	.698			
	SR Max	.398	.013	.111	2.618	.066
Step 3	Total Years	-.134	.352			
	Age	.019	.898			
	SR Max	.389	.010			
	LS Order	.345	.024	.211	3.605	.015

Note. β = Standardized regression coefficient, beta. Adj. R^2 = Adjusted R^2 for the model as a whole.

[a]Total years of foreign language study.

Table 11

Regression Results from Developmental Vocabulary and Grammar Scores

Dependent Variable	Step	Predictors	β	p	Adj. R^2	F	p
Set 1 Vocabulary	1	Age	-.075	.641			
		Total Years[a]	-.243	.135	.014	1.267	.294
	2	Age	-.064	.657			
		Total Years	-.347	.023			
		Describe grammar[b]	.467	.003	.211	4.472	.009
	3	Age	.012	.932			
		Total Years	-.360	.013			
		Describe grammar	.489	.001			
		LS Item	.336	.020	.307	5.317	.002
Set 1 Vocabulary Review	1	Age	.065	.691			
		Total Years	-.161	.327	-.022	.579	.565
	2	Age	.046	.745			
		Total Years	-.034	.816			
		Rule for plurals?	.521	.001	.226	4.801	.006
	3	Age	.058	.660			
		Total Years	-.137	.335			
		Rule for plurals?[c]	.431	.004			
		Describe grammar	.364	.012	.336	5.933	.001
Set 2 Vocabulary	1	Age	-.029	.855			
		Total Years	.283	.085	.030	1.594	.217
	2	Age	-.087	.566			
		Total Years	.285	.063			
		NR Max	.395	.012	.168	3.558	.024

Dependent Variable	Step	Predictors	β	p	Adj. R^2	F	p
Set 2 Grammar	1	Age	-.083	.620			
		Total Years	.124	.456	-.031	.422	.659
	2	Age	-.245	.107			
		Total Years	.049	.734			
		NR Total	.555	.001	.240	5.002	.005
	3	Age	-.246	.084			
		Total Years	.093	.494			
		NR Total	.435	.006			
		Non-Word Recognition	.349	.019	.337	5.823	.001

Note. β = Standardized regression coefficient, beta. Adj. R^2 = Adjusted R^2 for the model as a whole.

[a]Total years of foreign language study. [b]Points scored for description of the grammar system overall. [c]Points scored for description of the rule for forming plurals.

Figure Captions

Figure 1. Baddeley's (2003) model of working memory.

Figure 2. The procedure of the current study.

Figure 3. Major correlations between phonological memory measures and final vocabulary scores.

Figure 4. Major correlations between working memory measures and final vocabulary scores.

Figure 5. Major correlations between phonological and working memory measures and final grammar scores.

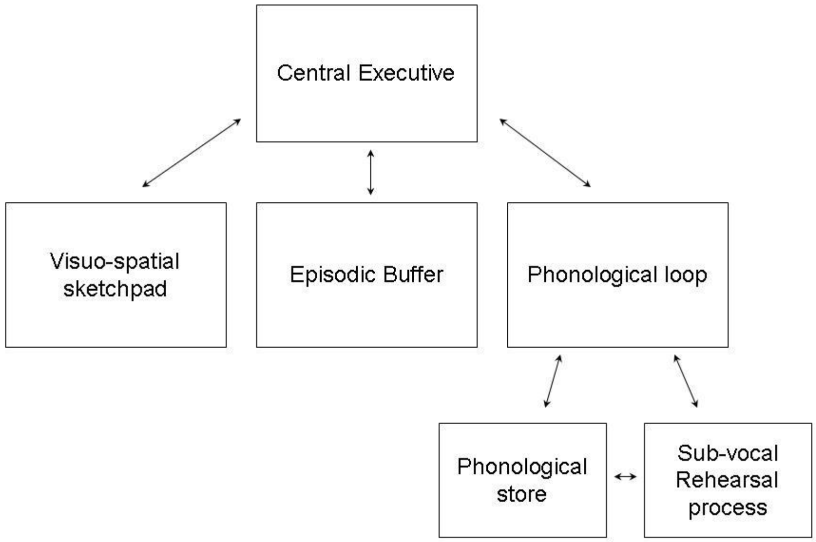
Figure 6. NR Max and Generalization Set vocabulary scores.

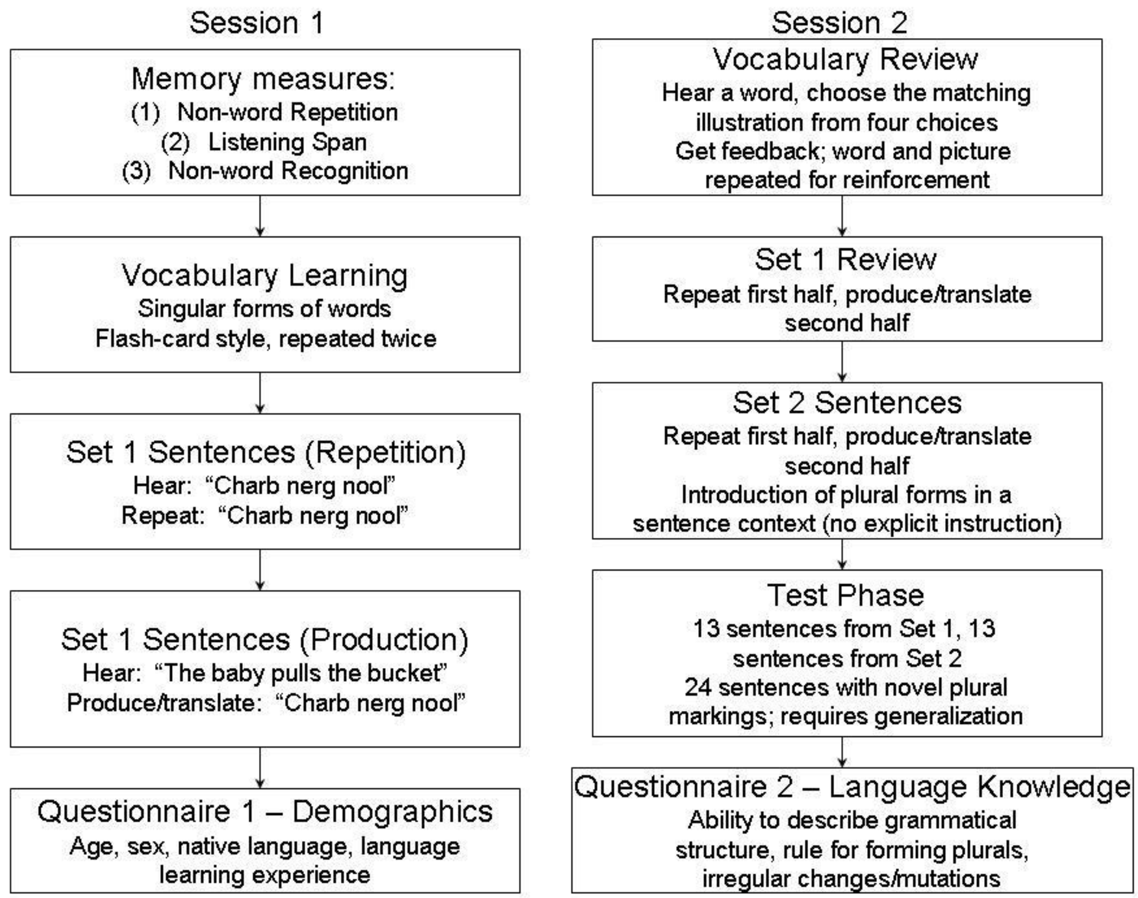
Figure 7. LS Order and Generalization Set vocabulary scores.

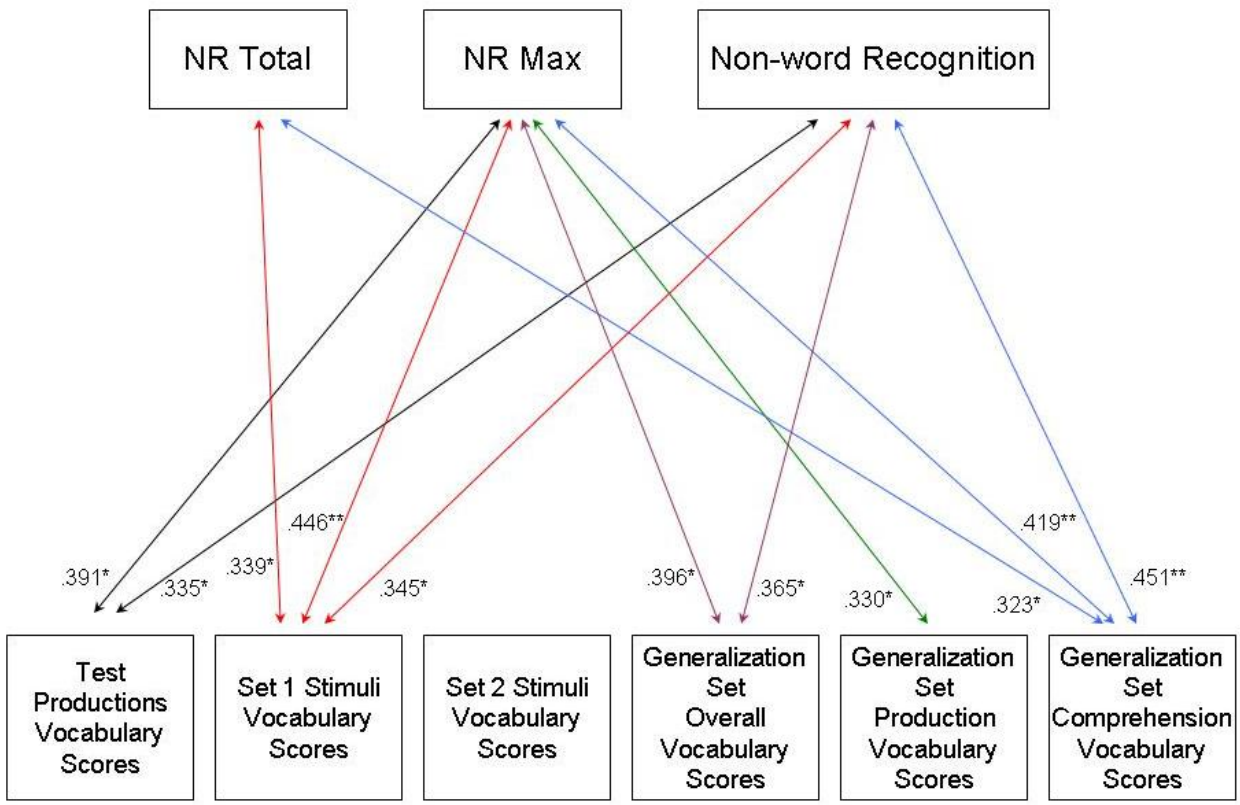
Figure 8. Generalization Set vocabulary and grammar scores.

Figure 9. NR Max and Generalization Set grammar scores.

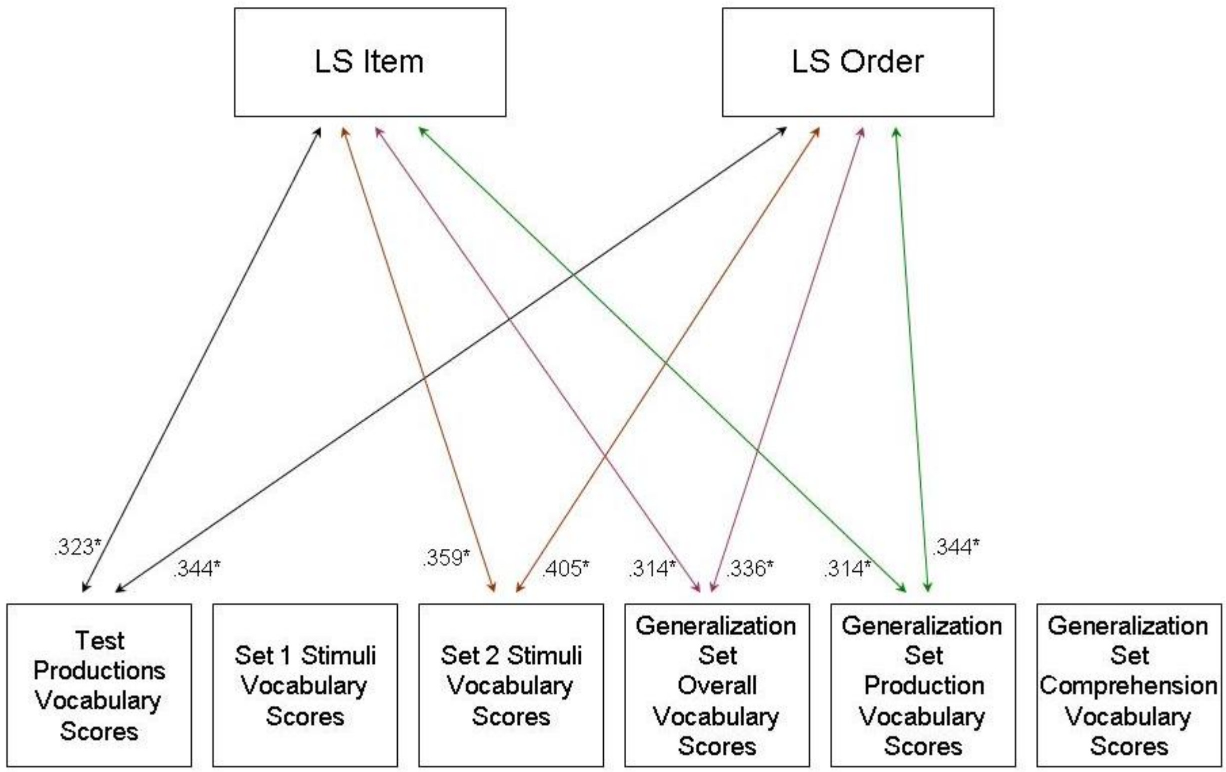
Figure 10. LS Order and Generalization Set grammar scores.


Central Executive
Visuo-spatial sketchpad
Episodic Buffer
Phonological loop
Phonological store
Sub-vocal Rehearsal process

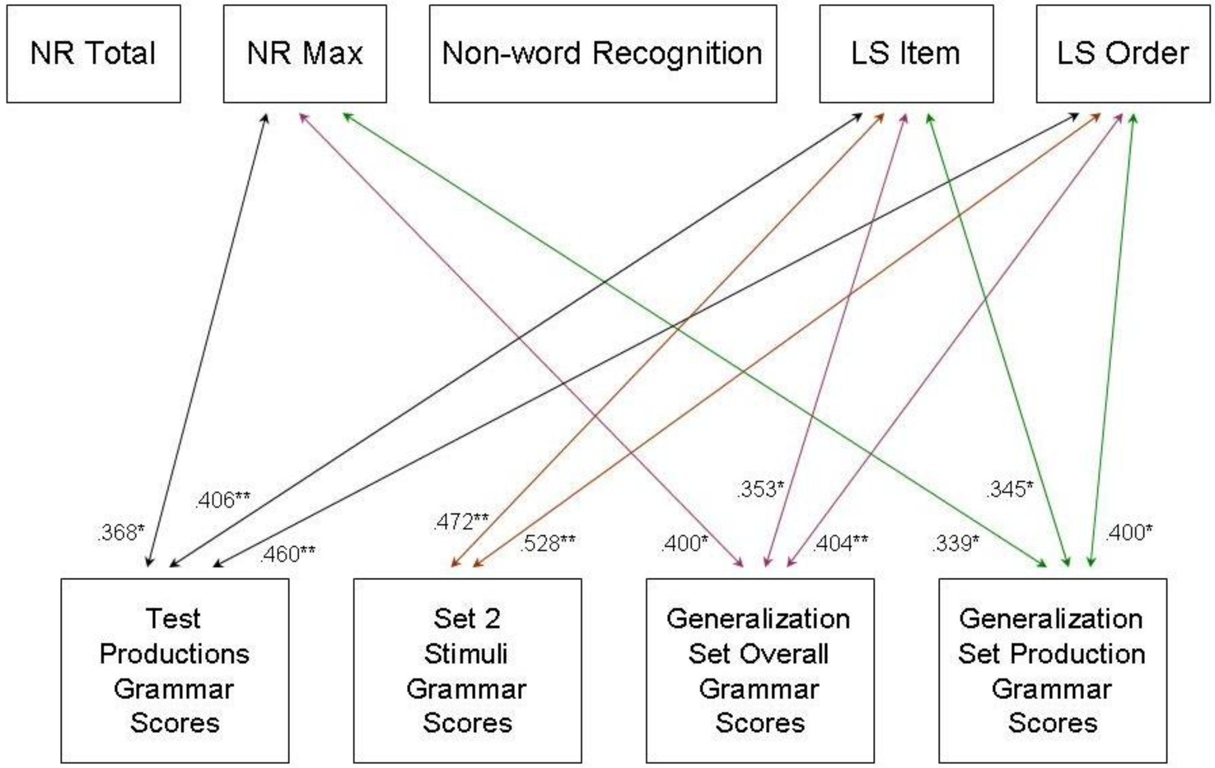

Session 1
Memory measures:
(1) Non-word Repetition
(2) Listening Span
(3) Non-word Recognition
Vocabulary Learning
Singular forms of words
Flash-card style, repeated twice
Set 1 Sentences (Repetition)
Hear: "Charb nerg nool"
Repeat: "Charb nerg nool"
Set 1 Sentences (Production)
Hear: "The baby pulls the bucket"
Produce/translate: "Charb nerg nool"
Questionnaire 1 – Demographics
Age, sex, native language, language learning experience
Session 2
Vocabulary Review
Hear a word, choose the matching illustration from four choices
Get feedback; word and picture repeated for reinforcement
Set 1 Review
Repeat first half, produce/translate second half
Set 2 Sentences
Repeat first half, produce/translate second half
Introduction of plural forms in a sentence context (no explicit instruction)
Test Phase
13 sentences from Set 1, 13 sentences from Set 2
24 sentences with novel plural markings; requires generalization
Questionnaire 2 – Language Knowledge
Ability to describe grammatical structure, rule for forming plurals, irregular changes/mutations


NR Total
NR Max
Non-word Recognition
.391*
.335*
.339*
.446**
.345*
.396*
.365*
.330*
.323*
.419**
.451**
Test Productions Vocabulary Scores
Set 1 Stimuli Vocabulary Scores
Set 2 Stimuli Vocabulary Scores
Generalization Set Overall Vocabulary Scores
Generalization Set Production Vocabulary Scores
Generalization Set Comprehension Vocabulary Scores

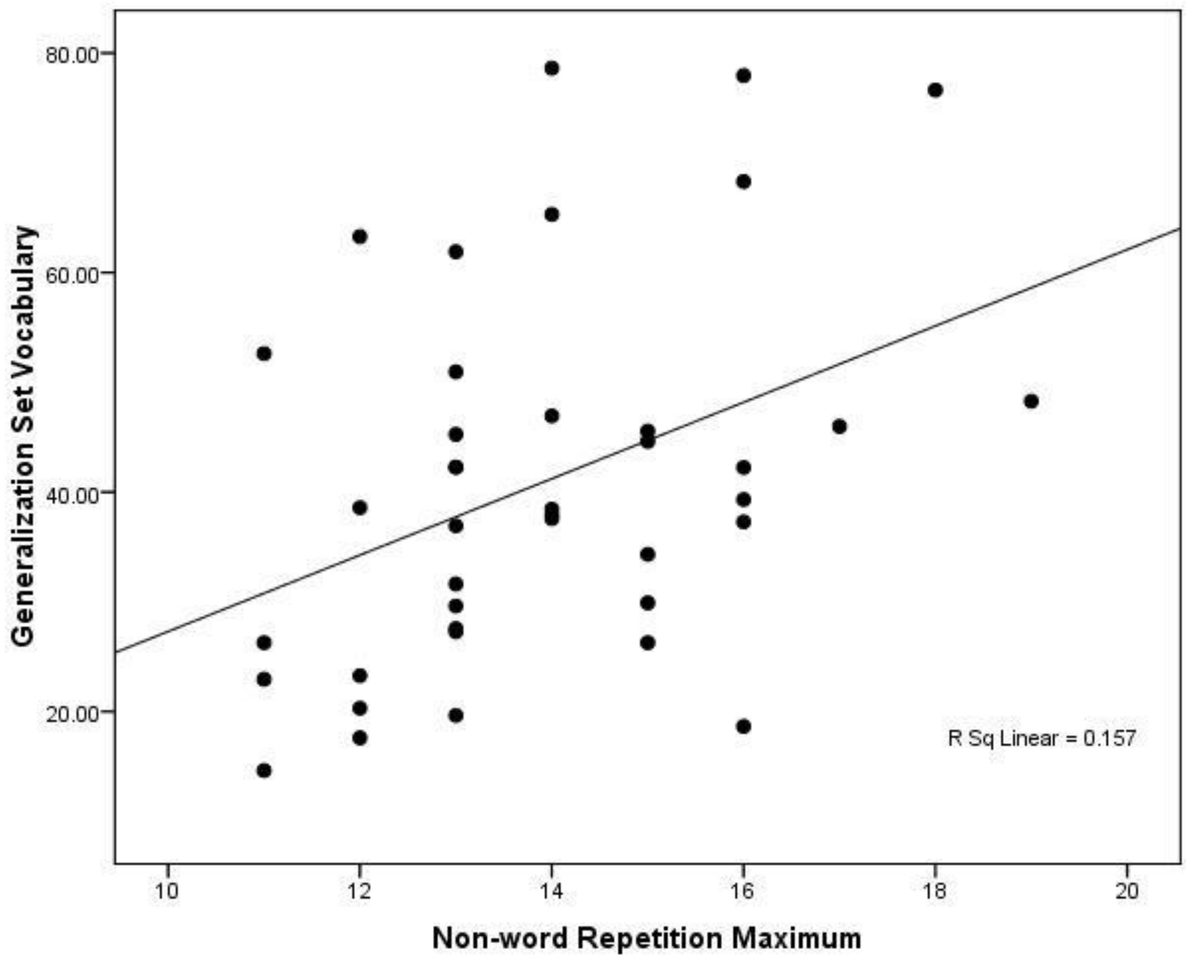
*p < .05; **p < .01

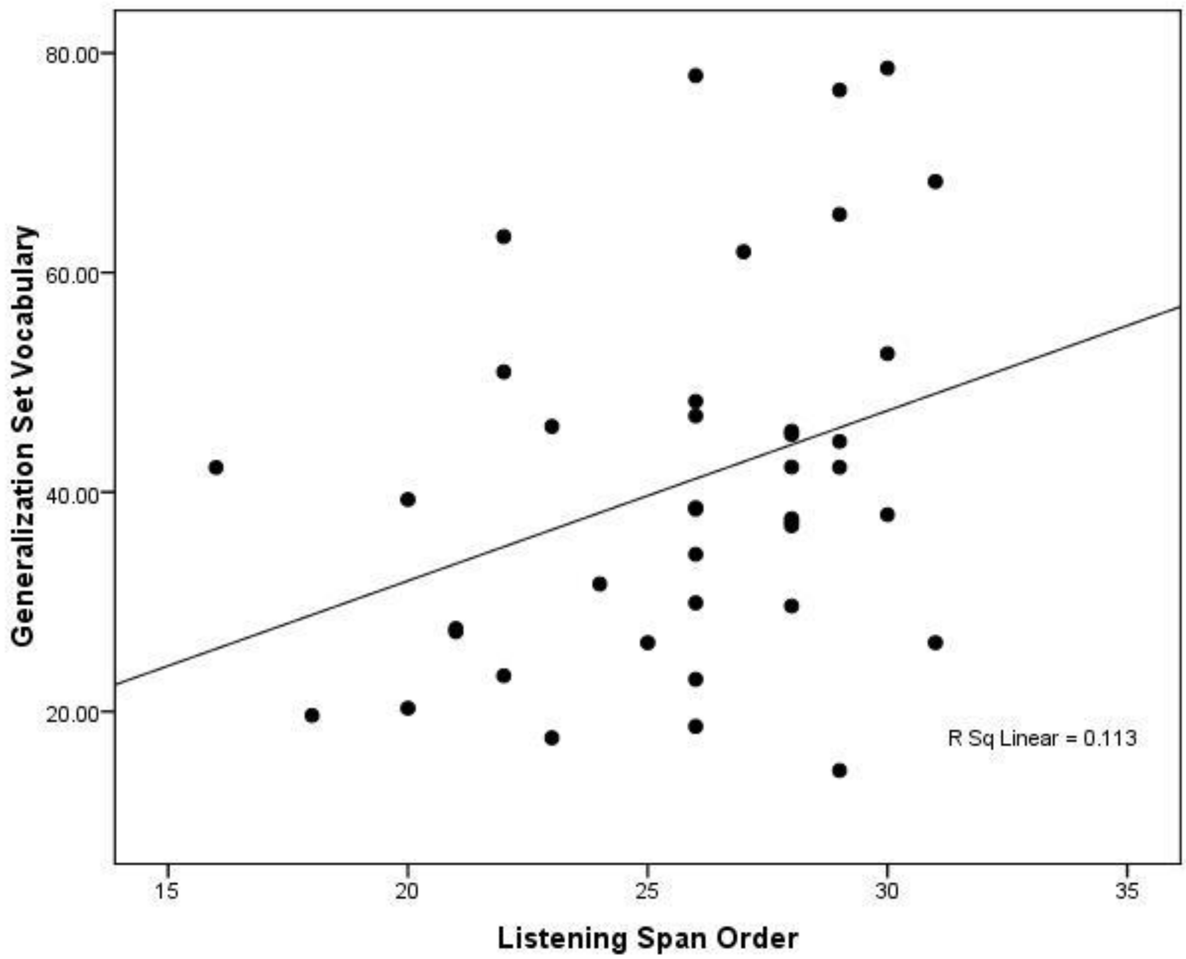

LS Item
LS Order
.323*
.344*
.359*
.405*
.314*
.336*
.314*
.344*
Test Productions Vocabulary Scores
Set 1 Stimuli Vocabulary Scores
Set 2 Stimuli Vocabulary Scores
Generalization Set Overall Vocabulary Scores
Generalization Set Production Vocabulary Scores
Generalization Set Comprehension Vocabulary Scores

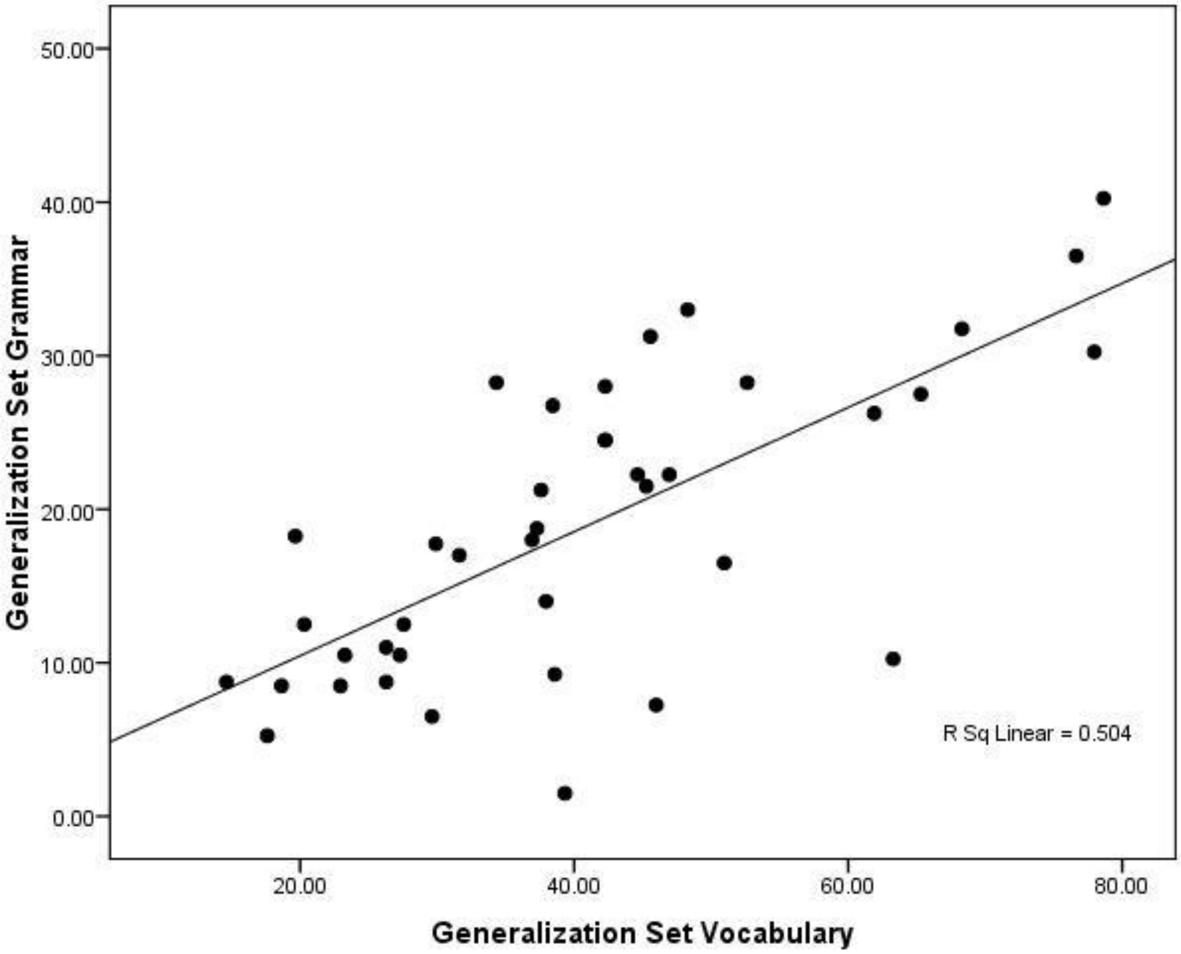
*p < .05; **p < .01

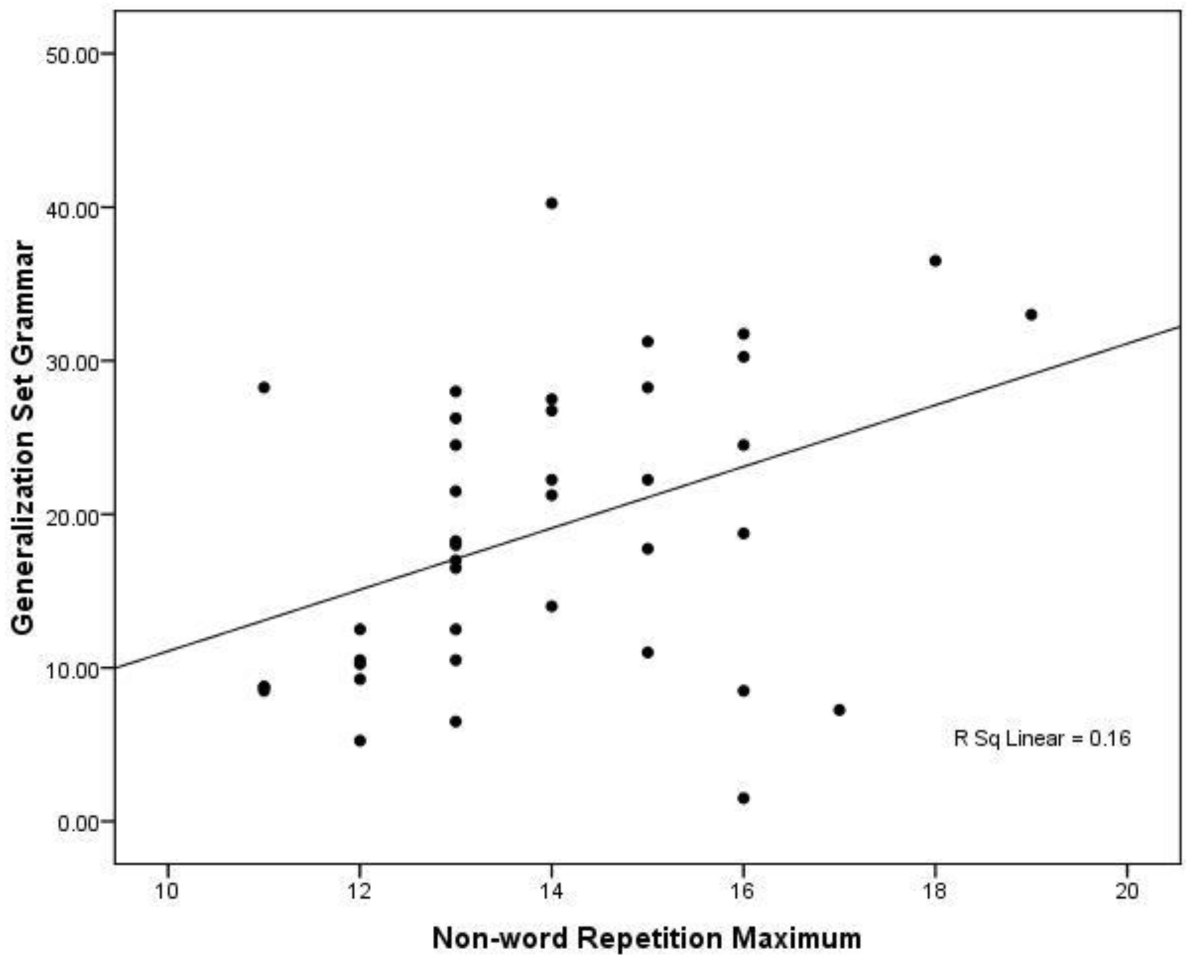

NR Total
NR Max
Non-word Recognition
LS Item
LS Order
.368*
.406**
.460**
.472**
.528**
.400*
.353*
.404**
.339*
.345*
.400*
Test Productions Grammar Scores
Set 2 Stimuli Grammar Scores
Generalization Set Overall Grammar Scores
Generalization Set Production Grammar Scores

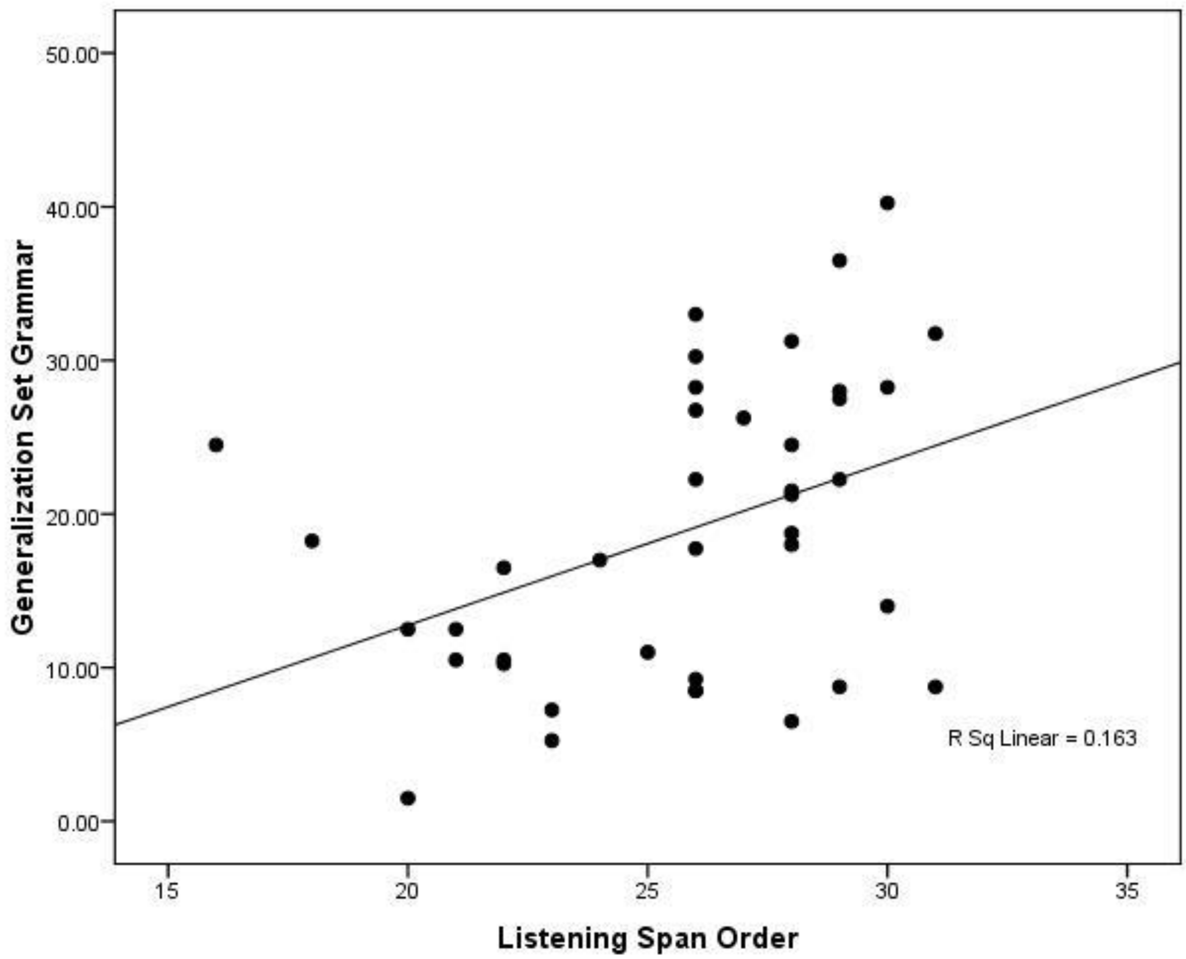
*p < .05; **p < .01


80.00
60.00
40.00
20.00
Generalization Set Vocabulary
R Sq Linear = 0.157
10
12
14
16
18
20
Non-word Repetition Maximum


80.00
60.00
40.00
20.00
Generalization Set Vocabulary
15
20
25
30
35
Listening Span Order
R Sq Linear = 0.113


50.00
40.00
30.00
20.00
10.00
0.00
Generalization Set Grammar
R Sq Linear = 0.504
20.00
40.00
60.00
80.00
Generalization Set Vocabulary


50.00
40.00
30.00
20.00
10.00
0.00
Generalization Set Grammar
R Sq Linear = 0.16
10
12
14
16
18
20
Non-word Repetition Maximum


50.00
40.00
30.00
20.00
10.00
0.00
Generalization Set Grammar
15
20
25
30
35
Listening Span Order
R Sq Linear = 0.163