

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 22 (1998)
Indiana University

**Performance of Normal-Hearing Children on a
New Working Memory Span Task¹**

Janna L. Carlson,² Miranda Cleary and David B. Pisoni

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This work was supported by NIH NIDCD Research Grant DC00111 and Training Grant DC00012 to Indiana University. We would like to thank Karen I. Kirk, Luis R. Hernandez, and Sarah H. Ferguson for helpful advice and comments.

² Now at Department of Hearing and Speech, Vanderbilt University, Nashville, TN.

Performance of Normal-Hearing Children on a New Working Memory Span Task

Abstract. This paper reports memory span data from 45 monolingual, normal-hearing children ranging in age from 3;5 to 5;11 years. A newly implemented methodology (here termed the “memory span game”) required participants to reproduce sequences of lights, or sequences of lights accompanied by sounds, by pressing appropriately matched buttons on a response box. Auditory stimuli used in the light-plus-sound condition included digit-names and nonsense syllables. The effects of auditory stimulus type, and of presenting the target sequence as lights-plus-sound vs. lights-only, were examined. Digit-spans obtained using the memory span game were compared to spans from the same children obtained using the conventional digit memory span sub-test of the Wechsler Intelligence Scale for Children III, in order to assess the degree to which the processing demands of these two tasks overlap. In addition, expanding on previously reported findings in the literature concerning a relationship between working memory span and vocabulary development, we obtained measures of each child’s receptive and expressive vocabulary as well as nonword repetition skills using two standardized tests. Results show that children improve as a function of age on the new memory game task at rates very similar to that obtained for the WISC forward digit span task. We found that children younger than 3;5 were generally unable to meet the demands of the memory span game task. Differences between the group means between the three different memory span game conditions were not significant; however, correlational analyses showed clear differences between the visual-spatial/lights-only condition vs. the auditory plus visual-spatial condition in how predictive each score was of vocabulary measures. Scores obtained from the auditory plus visual-spatial condition of the memory game task accounted for approximately 12-16% of the variance in vocabulary scores with the effects of age partialled out, while the lights-only/silent condition scores only predicted approximately 2% of this variance. Results from the nonword repetition measures show that the phonetic similarity of the nonsense words to familiar words influenced the strength of this task’s relationship with vocabulary measures, but not with working memory span measures such as WISC digit-span and memory span game scores (although the previously reported small positive correlations between these measures were replicated). These results have both theoretical and clinical implications for the use of this reproductive memory span task with hearing-impaired children who may have expressive language problems or delays. The findings are also relevant to the development of phonological representations in short-term memory and spoken word recognition abilities in normal-hearing children.

Introduction

Investigators have employed many different tasks over the years to try to evaluate phonological working memory and its development in young children. The vast majority of these tasks, however, assume a “normally” developed link between the auditory perception aspect of a short-term memory task and the articulatory verbal report upon which the assessment of performance is based. That is to say, it is usually assumed that the stimulus items used in an auditory memory span task are equally discriminable to different children within the same peer group, and that the spoken report of what each child remembers is a comparable measure across children. In certain populations of children, however, this assumption is inappropriate. For example, children with severe hearing impairments often produce low-intelligibility

speech, thus compromising their verbal output responses for a traditional span task as partially a function of the perceiver administering the test. Often too, traditional tests of memory span (such as that included as an optional subtest of the popular Wechsler Intelligence Scale for Children (Wechsler, 1974; WISC-III, 1991) are administered only as spoken in a face-to-face situation, with the details of task administration (e.g., timing, voice characteristics) differing from one session to the next. A new memory span game task—a sequence reproduction task—has recently been developed in our laboratory to overcome some of these output problems. Its format is such that it can easily be used both with normal-hearing children or children with hearing-impairment. The first use of this task with a hearing-impaired population was reported in Cleary, Pisoni and Kirk (1998, this volume). The present study provides *normative* data for *normal-hearing* children between the ages of three-and-a-half to five years on the new memory span game task and compares performance on this task to *traditional* auditory memory span tasks such as auditory digit-span. The present study also examines how performance on different versions of this new sequence reproduction task relates to measures of receptive and expressive vocabulary development. We also report preliminary findings regarding how performance on two different nonword repetition tasks is related to the vocabulary and memory span scores.

Simple auditory memory span tasks requiring a subject only to immediately repeat back a sequence of words (often digit-names or letter-names), are widely used clinically as a routine check of general cognitive function (Lezak, 1995). While simple span tasks have been in use for over a century (Dempster, 1981), theoretical interest in how short-term auditory memory can be explicitly modeled has been a more recent phenomenon. One model that has enjoyed unusual longevity, perhaps due to its quite general specifications, is Baddeley and Hitch's (1974) model of short-term working memory. Baddeley and Hitch proposed that working memory consists of three sub-components: the "central executive," the "visuospatial sketchpad," and the "phonological loop." According to this account, the central executive allocates attention and controls the latter two sub-components which Baddeley and Hitch referred to as "slave systems." The function of the visuospatial sketchpad is to maintain and manipulate visual-spatial images, while the phonological loop is responsible for storing phonological (i.e., speech) information (Baddeley, 1992).

The phonological loop itself is further broken down into two hypothesized components, one having primary responsibility for the storage aspect, and the other for *rehearsing* the information, perhaps based on a sub-vocally articulated form of speech. The memory traces in the phonological loop, according to this account, are assumed to last only about two seconds unless they are maintained by sub-vocal rehearsal, which feeds the information "back through the loop." There has been a variety of evidence offered for the existence of the phonological loop, primarily centered around several reliable effects, the most important of which for this paper are the "acoustic similarity effect" and the "word length effect."

The acoustic similarity effect is the finding that immediate recall of items declines when the items are phonologically similar (Conrad, 1964). For example, people are less likely to correctly repeat a list of letters whose names are phonologically similar (e.g., "p," "t," "v") than a list of letters whose names are dissimilar-sounding. Thus, a memory span measure that uses the syllables "pah," "bah," "tah," and "dah," such as we used in this study, is predicted to be generally shorter than that obtained using a set of easily discriminable stimuli such as the names of four digits, which are acoustically quite dissimilar.

The word-length effect refers to the inverse relationship between memory span for words and the duration of the words (Baddeley, Thomson, & Buchanan, 1975). In general, people are able to recall about as many words as they can say in two seconds. Thus, memory span for words with longer durations (as measured in "word units") is shorter than memory span for words with shorter durations. For example, when digit span is measured in subjects across different languages, digit span is higher for subjects with

languages that have shorter vowel durations and fewer syllables for the digit names (Ellis & Hennessey, 1980; Naveh-Benjamin & Ayres, 1986).

The memory game task in this study uses a series of button-presses as the required response. This procedure was developed primarily as a tool for easily gathering memory span data when a verbal report of recalled items is problematic to collect. Evaluating the function of the phonological loop in children with hearing impairments is of particular importance because these children are already receiving degraded input, and a phonological loop operating with reduced function could further impede their vocabulary acquisition and language development. Children with cochlear implants have achieved various degrees of success in the acquisition of spoken language. Research shows that some children with cochlear implants (CIs) have excelled in speech perception, comprehension, word recognition, and language and speech intelligibility, whereas most children show more modest benefits (Pisoni, Svirsky, Kirk, & Miyamoto, 1997). The children who excel in language, often referred to as “Stars,” are able to use the auditory input they receive through the CI to perceive, encode and retrieve the phonological patterns of words, thereby establishing a lexicon and acquiring the grammar of their ambient language. Even when children with CIs are matched along other variables such as peripheral auditory input, there is still a great deal of unaccounted for variability in the speech perception outcome measures. In other words, some children go on to become Stars for reasons that are still not well understood. It is suggested that there may be underlying psychological mechanisms which are contributing to the Stars’ success in this area (Pisoni, 1997). Therefore, a uniform measurement of children’s working memory abilities may reveal important differences among normal-hearing, hearing-impaired, Stars, and other children with CIs.

One task that may provide such a measurement across the various clinical populations is a simple immediate memory span reproduction test recently developed at our lab. This task measures a child’s ability to encode a sequence of speech-like sounds and reproduce the sequence by pressing an associated series of colored buttons on a response box. The “memory game” task bypasses the articulatory production requirement of the traditional verbal digit span tasks, thus making the test inherently more suitable for hearing-impaired children and other children with output problems. The memory game could provide a viable means for testing hearing-impaired children, such as children with cochlear implants (CIs) who have previously been exposed to degraded speech input. In addition, the memory game might also be used in the intervention process to help train children with inefficient “phonological loops” to improve their processing, storage, encoding, and reproduction abilities. Possibly, this task might be useful to examine phonological memory encoding abilities in other clinical populations such as phonologically disordered children (i.e., functional misarticulators), or children with specific language impairment (SLI). We view the memory game task as presenting a viable means of allowing a variety of populations to be tested in the same manner while perhaps minimizing the interference associated with the handicaps of particular disorders. In addition, the memory game task does not inherently have a ceiling level, so it may even be used to test children (we estimate at least three-and-a-half years old and up) as well as adults.

In a recent study, Cleary (1997) used a version of the sequence-reproduction memory game task used here to look at the effect of perceptual discriminability of vowels on memory span in adults. In that study the sequences were presented to the subject only in the auditory modality and subjects were asked to reproduce the sequence by pressing buttons learned beforehand that corresponded to each auditory stimulus item. In her study, however, Cleary did not explicitly examine the relationship of the memory game task with other more traditional verbal measures of memory span.

The present study does not employ auditory-only method of presentation as did the study above (as this is a somewhat more difficult task), but otherwise the methodology for the memory game task is very similar. The present study, however, goes beyond simply exploring what kind of effects can be replicated using this new task, and begins to examine if modality-specific relationships between short

term memory for sound sequences and developing language skills can be obtained using this method, in spite of the visual-spatial component of the basic task.

There is one kind of purported relationship that we are especially interested in. Recent work by Baddeley and colleagues (e.g., Gathercole et al., 1994; Baddeley, Gathercole, & Papagno, 1998) suggests that the phonological loop may not be used solely for short-term retention of information. Nonsense or nonword tests, such as the Children's Test of Nonword Repetition (CNRep, Gathercole, Willis, Baddeley, & Emslie, 1994), requiring immediate verbal repetition (i.e., imitation) of novel sequences of speech sounds, have provided some evidence for the phonological loop playing a role in the development of spoken language and the acquisition of new words (Baddeley et al., 1998). Specifically, these studies suggest that a child's ability to encode and briefly retain novel phonological information is closely linked to the child's vocabulary knowledge (Gathercole, Service, Hitch & Martin, 1997). Studies of normal-hearing, preschool-age children have found that children who excel on nonword repetition tests also tend to be the same children who have larger vocabularies for their age group (Gathercole, 1995; see also Bowey, 1996, Snowling, Chiat, & Hulme, 1991). Another such test, specifically Sub-test 10 of the Comprehensive Test of Phonological Processes (CTPP, Torgesen and Wagner, 1997), is currently being used to help predict reading-readiness in young children.

Possible explanations for this relationship have been hotly debated, alongside claims of a causal relationship between the variables. Computation models of how this interaction may occur have been developed to explore the issue (e.g., Gupta & MacWhinney, 1997). According to one view, the phonological loop provides temporary storage and processing of the novel sound sequences which is a prerequisite for the establishment of a more permanent phonological representation of spoken words (Baddeley, Gathercole, & Papagno, 1998). The main test of this hypothesis involves longitudinal studies of whether measures of immediate phonological memory are better able to predict measures of vocabulary development at some future time point, rather than vice versa. Within a narrow age window (specifically between the ages of four and five years), Gathercole et al. (1994) did find nonsense/nonword word repetition performance (as measured by the CNRep) to be significantly correlated with children's subsequent vocabulary growth (scores on the British Picture Vocabulary Scale, Dunn & Dunn, 1982), and reading and language comprehension improvement (scores on Reading Test A of the British Abilities Scales, Elliott, 1983) in the following year. In Gathercole et al.'s study, vocabulary at four years was not predictive of nonsense word repetition ability a year later. However, for the two years following, from age 5 to 7 years, this relation was reversed.

This reversal has led other researchers to argue that long-term lexical knowledge facilitates more efficient coding in the phonological loop rather than the reverse. The idea here is that children with larger and more diverse lexicons are more likely to be able to activate some set of similar sounding items (i.e., phonological neighbors; Luce & Pisoni, 1998; or cohorts; Marslen-Wilson, 1987) in long-term memory (i.e., the lexicon) when word-like nonsense words are heard, and thus may be able to partially retrieve the item for recall based on analogy. In other words, long-term knowledge may facilitate efficient coding of streams of segments, and thus the memory for more units at the syllable or word level. Still other studies have argued that the two systems develop lock-in-step and that it is somewhat misguided to look for one skill necessarily causing improvement in the other (Snowling, Chiat, & Hulme, 1991). We are inclined to adopt this last suggestion, though the present study does not address this issue directly. However, previous work from our lab has shown with adults that degree of vocabulary knowledge can be used to predict the speed of response in verbal processing tasks such as visual (i.e., orthographic) lexical decision and visual word naming (Lewellen, Goldinger, Pisoni, & Greene, 1993).

Of relevance to the study reported here is the general finding that measures of nonsense word repetition appear to be more strongly correlated with standardized vocabulary and reading comprehension

skills than simple verbal memory span measures. This is partly of interest because it is still unclear even after more than a century of research, if simple forward-recall memory span tasks administered in the traditional fashion predict performance on commonly-used daily skills in normal adults. Basically, the argument has been that performance on a task like the CNRep provides a better measurement of the related skills needed for vocabulary acquisition. These skills include phonological working memory, phonological analysis of unfamiliar auditory speech sequences, the articulatory output process, and access to long-term representations of lexical knowledge. At the same time, however, use of a nonsense word repetition task to assess phonological memory development in hearing-impaired populations raises a multitude of subsidiary issues that we are not yet in a position to fully address. In normal-hearing children, however, the situation regarding nonword repetition is somewhat less murky.

In the present study, we also wished to examine the relative ability of the different versions of the sequence reproduction memory game task to reflect individual differences in vocabulary achievement. Our expectation was that the memory game task using digit-names was so similar to traditional digit span tasks that the relationship of the memory game scores on this condition to vocabulary differences would be quite small (but present in the form of a positive correlation), in line with previous results using traditional digit span. The syllable condition of the memory game task might share more processes in common with nonsense word repetition (even given the delays between the syllable elements), but to rigorously test this, a differently constructed experiment would be necessary. The silent condition of the memory game, however, should have comparatively little to do with performance on vocabulary and nonword repetition measures. Unless the auditory-plus-visual/spatial version of the task does not draw sufficiently on phonological mechanisms for measurement (that is, children treat all memory game conditions alike using only visual/spatial cues), we expect to obtain evidence that a shared dependence on the auditory modality should link the vocabulary and nonword repetition scores more closely with auditory-plus-visual/spatial memory game scores than with performance on the lights-only condition of the memory game.

One last issue of concerns our choice of nonword repetition tests. Other work ongoing in our lab (e.g., Large, Frisch, & Pisoni, 1998; Vitevitch, 1998) is also concerned with the issue of how the similarity of novel words to existing known words, and a novel word's individual phonetic/phonological structure affect a listener's perception (and production) of a novel lexical item. In young children, for example, "length effects" related to the duration of the novel word can be seen even within repetitions of single word or word-like items, with articulation accuracy decreasing as the length in phonemes or syllables is increased (e.g., Gathercole, Willis, Baddeley, & Emslie, 1994). On the other hand, in the perceptual domain, it has been shown that longer real words are often perceptually more intelligible than shorter words--partly due to the tendency of shorter words to have many similar-sounding phonological "neighbors" with which they may easily be confused given minimal contextual information (Kirk, Pisoni, & Osberger, 1995). Analogous reasons appear to cause short nonwords to be judged more word-like than longer nonsense words and more likely to be misperceived as real words. The present study includes preliminary examination of how length effects interact with the particular phonemic composition of the nonsense words-to-be-repeated, by testing the children's repetition of two contrasting lists of nonsense words. One list contained many nonsense words that were judged similar to real words by adults, while the other list contained many nonsense words that were judged by adults to be less similar to real words. While Gathercole et al. (1994) showed that nonsense words that were rated as extremely "word-like" by adults were repeated more accurately by children than nonwords rated as less "word-like," they did not analyze in any great depth what phonological factors contributed to making a nonsense word particularly "word-like."

To review, the present study was designed to examine the validity of the memory game task as a useful tool in assessing the phonological working memory of normally developing preschool-age

children. One of our main concerns was just how similar this sequence reproduction task is to traditional verbal digit span tasks. We will show that age-correlated improvement on our new task parallels that demonstrated for traditional span tasks, but that the spans generated with the memory game task are typically somewhat shorter than those obtained for the WISC-III version of the digit-span task (to the extent that these spans can be directly compared, given some differences in administration and scoring). Additionally, our results indicate that while our hypothesized differences in average span as a function of phonological vocabulary (digit-names vs. syllables vs. silence) were not obtained across the children as a group, this result was partially due to a pattern of individual differences between children. These differences were examined in conjunction with other measures from the same children in a correlational analysis. Results indicated that the different memory game conditions were not equivalent in how well they predicted performance on other tasks included in this study. The correlational analysis also showed significant positive correlations between both receptive and expressive vocabulary scores and nonword repetition scores (with the effects of age statistically controlled for) when the nonsense words to be repeated were judged to be somewhat word-like in their composition. When the nonword repetition test contained a large number of words that were less word-like in their composition, a significant positive correlation was obtained only with expressive vocabulary scores, and not receptive vocabulary scores.

If the new memory game task provides a valid way to assess phonological working memory function in normally-developing children, the task may be very useful for routinely monitoring memory development in other populations as well, such as children with hearing impairment, in conjunction with more direct assessments of progress in speech and language skills.

Methods

Participants

Sixty-one children (24 three-year-olds, 18 four-year-olds, and 19 five-year-olds) from the Bloomington, Indiana community participated in the experiment. All the participants were monolingual, native English speakers who had normal speech-language development (as indicated by a parent's response on a questionnaire). A hearing screening was administered to ensure that all the children had normal hearing at the time of testing. Eight of these sixty-one children served as pilot subjects. The data from four subjects (1 three-year-old, 1 four-year-old and 2 five-year-olds) were not used in the final analysis because the children did not pass the hearing screening. Four of the children (1 three-year-old, 1 four-year-old and 2 five-year-olds) were unable to complete all of the experimental tasks because of a time constraint or lack of attention. Thus, the final sample of children whose data were analyzed included a total of 45 children (15 three-, four- and five-year-olds). There were 7 male and 8 female three-year-olds, 7 male and 8 female four-year-olds, and 8 male and 7 female five-year-olds.

Stimulus Materials

The children were tested using four different types of behavioral measures: memory span game tests, traditional WISC digit span tests, vocabulary tests, and nonword repetition tests.

Memory Span Game. The memory game response-box, modeled after the commercial product "Simon" produced by Milton Bradley, consists of a clear circular container with four flat, large colored buttons (red, yellow, blue and green) which can easily be depressed by a small child. (See Figure 1.) Each button is back-lit by a light which may be rapidly turned on and off via the attached PC computer.

 Insert Figure 1 about here.

The memory game program allows for pre-recorded auditory stimuli to be presented in concert with individual lights illuminated in sequence. In this study, the concurrent auditory stimuli were presented over a loudspeaker. In all studies using this apparatus so far, only four different auditory stimuli have been used in each condition, and a particular auditory stimulus item is always consistently associated with a single light at one of the four buttons during each condition. For example, the blue button is always illuminated and remains on while the auditory stimulus “five” is presented.

A male speaker of American English recorded the auditory stimuli for the memory game. The stimuli were recorded in a sound-attenuated, single-walled anechoic recording chamber (Industrial Acoustics Company Audiometric Testing Room, Model 402) using a head-mounted, close-talking microphone (Shure SM98A). The recordings were digitally sampled on-line at a rate of 22.05 kHz with 16-bit amplitude quantization using a Tucker-Davis Technologies (TDT) System II with an A-to-D converter (DD1). The signal was then low-pass filtered at 10.4kHz (anti-aliasing filter, FT5) which was controlled by an updated version of Dedina’s 1987 Speech Acquisition Program (Dedina, 1987; Hernandez, 1995). The stimuli were recorded into individual sound files, and the overall root mean squared (RMS) amplitudes were digitally equated to 70 dB RMS to allow uniform presentation levels for the stimulus files.

The digits one, three, five, and eight were arbitrarily chosen for use in the memory-game because they are all monosyllabic words that are neither consecutive nor a set of all odd or even numerals. The confusable consonant-vowel combinations /ba, pa, ta, da/ were selected as a set of stop consonants with the common vowel /a/.

Traditional WISC Digit Span Task. The forward and backward digit span sub-tests from the WISC-III (1991) were used as a measure of the children’s verbal memory span. In the forward digit span task, the experimenter presents a list of digits (beginning with the list length of two) via live voice. If the child successfully repeats at least one of two presentations at a given list length, the list length is increased by one digit. The child reaches ceiling when he/she fails to correctly repeat two presentations of a particular list length. In the backward digit span, a list is presented, and the child must repeat the digits in reverse order. Once again, the child must successfully repeat at least one presentation of each list length in order to progress to the next length.

Vocabulary Tests. Measurements of the children’s expressive vocabulary were obtained by using the EVT (Williams, 1997). The first two levels of the EVT require each child to verbally label a standardized set of pictures. The rest of the EVT stimuli include a picture and one possible label for the picture that the examiner says aloud. The child must then verbally produce a synonym for the presented word and picture. The PPVT-III, Form III-A (Dunn & Dunn, 1997), was used to measure the children’s receptive vocabulary. In this test, on each trial the child is presented with a group of four picture plates and must point to the picture that corresponds to the word spoken by the examiner.

Nonword Repetition. The 40 nonwords from the CNRep (Gathercole et al., 1994) were digitally recorded by a female talker (JLC) with a Midwestern dialect and sampled at a rate of 22.05 kHz. The same audio recording procedure used for the memory game auditory stimuli was followed for the nonword stimuli, except for the leveling procedure. The stimuli were again recorded into individual sound files. Overall root mean squared (RMS) amplitudes were digitally equated to 65 dB RMS to allow uniform presentation levels for the stimulus files within this part of the study.

Some of the pronunciations of the CNRep nonwords were slightly adjusted to conform to the phonology of American English. The female speaker also recorded the 28 nonwords from sub-test 10 of

the CTPP (Torgesen & Wagner, 1997). Torgesen and Wagner's nonword list consists of three practice nonwords and 25 nonwords that vary in length from one to five syllables. Transcriptions of the nonwords are given in Appendix A.

To examine the phonological characteristics of the nonsense word stimuli, and to help insure that recorded items were heard as intended, twelve monolingual adult speakers of American English each reviewed all of the nonsense word stimulus tokens. Items were played through headphones (Beyerdynamic, DT 100) at approximately 70 dB SPL. Subjects were run in groups of one to five participants. In a first pass through the stimulus set, subjects were asked to listen to each item and rate the "word-likeness" of each token by pressing one of seven buttons on a labeled response box. Subjects were not provided a formal definition of what was meant by "word-likeness" although the instructions stated that they were to listen to each nonsense word and judge how much it sounded like a typical word of American English. Button 1 was defined to subjects as representing "Could not be a word of American English," while Button 7 was described representing "This is very word-like in American English." Button 1 was labeled "Not Word-like," Button 4 as "Medium Word-likeness," and Button 7 as "Very Word-like." In a second pass through the stimuli, the same subjects were asked to listen again, and then after each item, type in as best they could, a transcription of what they heard using conventional English orthography. Presentation of the items was self-paced, but subjects were told that the ratings task would be timed. Presentation order during each pass through the items was randomized. During the rating and transcription tasks, each subject was seated within an individual three-quarters-enclosed testing carrel, in front of a computer monitor, keyboard, and a seven-button response box, each linked to a 133 MHz Pentium Gateway PC with SoundBlaster 16AWE soundboard, timer board, and specialized parallel port.

Based on the adult listeners' transcriptions, a few minor changes were made to the phonemic transcription of our stimulus items such that in some cases we decided to accept either of two or in few cases, three different phonemes as "correct" when scoring the children's responses. These cases are noted by the "/"s in some of the transcriptions included in Appendix A. Mean ratings values and median rating reaction times as measured from the offset of each word are also listed in Appendix A. Appendix A shows that, on average, items from the CTPP nonword test were judged less word-like than items from the CNRep.

Procedure

While a number of behavioral measures were gathered from each child, an entire testing session was completed in approximately 1 hour. Brief breaks were allowed between each type of measure, with the younger children usually requiring somewhat longer respites.

In all cases, a hearing screening was first administered to the subjects at 250, 500, 1,000, 2,000 and 4,000 Hz at 20dB HL, using a Maico Hearing Instruments audiometer (MA27) and TDH-39P headphones. Each ear was screened separately.

Next, the Expressive Vocabulary Test (EVT, Williams, 1997) was administered to each child while the child's responses were recorded on a Sony portable Digital Audiotape Recorder (TCD-D8) to preserve a sample of the child's speech. If a child had normally-occurring developmental errors (regular substitution of a /w/-like sound for /r/, for example), the errors were noted at this point. Children with obvious articulation errors that deviated from normal patterns (Shriberg & Kwiatkowski, 1994) were not included in the final analysis.

The Children's Test of Nonword Repetition (CNRep, Gathercole, Willis, Baddeley, & Emslie, 1994) was then used to measure the child's ability to encode, store, process, and reproduce a novel sequence of speech sounds. The participants were instructed to "repeat the silly word" they heard through

the loudspeaker. These instructions are typical of those requiring children to complete a phonological repetition task, rather than a phonetic task which involves an exact reproduction of the speaker's accent and other speech features (Wells, 1995).

The nonwords were presented to the children in a random sequence and presentation order for the stimuli was randomized for each participant. Ten stimuli occurred at each syllable length (two, three, and four syllables). An experimenter controlled the computerized presentation of the nonword stimuli via an amplified loudspeaker having an output level of 70 dB SPL, as determined by a hand-held sound pressure level meter (Triplett 370) placed at the approximate position of the participant's head. The nonword repetitions were scored in real-time by one of the experimenters using a variation of the rating system from Gathercole et al. that scored the repetitions as correct or incorrect (Gathercole, Willis, Baddeley, & Emslie, 1994). In this experiment, a correct repetition received two points. One or more articulation errors received one point. If the child failed to respond (usually due to lack of attention) or refused to respond, that item received a score of zero. While an experimenter scored the repetition for accuracy, another experimenter transcribed the response. The DAT recorder was used to preserve the children's responses for later analysis.

Next the memory game span task was administered. For each memory game condition, a sequence of colored lights with corresponding auditory stimuli was presented under computer control. If the child successfully reproduced two sequences of the same length by pressing the associated buttons, the sequence was automatically increased by one item on the next trial (a list = 1 trial). If the child made a mistake of any kind during their response, on the next trial, the list length was reduced by one item. A computer program controlled the length and presentation of the sequence and recorded the child's response. The acoustic stimuli were presented at 70 dB SPL, via the same loudspeaker that was used for the nonword test.

The children were sometimes distracted during a trial's presentation. Therefore, if the child did not press at least one button before four seconds had elapsed following the last item in the list, the sequence would be repeated up to two more additional times (with additional 4 second delays between each list presentation, if no response was made). The computer program noted when each child was allowed an extra presentation. Only a very small proportion of trials involved repetitions of lists, and the child was never explicitly informed that this repetition mechanism was in place. Additionally, because the memory box never presented the same stimulus consecutively (e.g., "five" – "five" was not allowed), the box would not respond to multiple presses on a particular button. This prohibited children from being penalized if they applied too much pressure for too long to a particular response button.

After instructing each child to copy the lights and sounds exactly as the memory box presented them, each child completed a practice session with the digit-name stimuli. To familiarize the children with the task, the initial sequence length was one digit, and the experimenter prompted the children as necessary and instructed the children as to their mistakes. After completing the practice round (which only allowed the child 6 trials and a maximum length of 3 items), the children were tested with 16 trials using the digit-name stimuli. No explicit feedback relevant to the correctness of the child's response was given during these test trials.

A short break was taken, after which the child was instructed that the sounds were no longer going to be digits, but that instead they would hear the sounds "ba, pa, ta, da." The children then completed 16 trials of the memory game task using the consonant-vowel stimuli. Again no feedback was given.

The nonwords from Sub-test 10 of the Comprehensive Test of Phonological Processes (CTPP, Torgesen & Wagner, 1997) were then administered to each child, following the same presentation and scoring procedure as Gathercole et al.'s nonwords (CNRep).

The children next completed 16 trials in the final condition of the memory game task. In this condition, the acoustic stimuli were eliminated, so the children reproduced light-only sequences presented via the response box. This condition allowed for a measurement of the children's visual-spatial memory span independent of any auditory contribution. Each light was kept on for 300 ms with the inter-stimulus interval of the same time as used in the other conditions (approximately 100 ms). This duration matches the duration of the syllable stimuli quite well, but the digit-name stimuli are naturally more variable in length—the longest (“five”) being about 450 ms in duration.

Performance on the memory game task was automatically scored according to several different criterion of varying strictness. Here we report only the results using the *least* conservative criterion—that is, the child's score in a given memory game condition is the longest list length the child got correct during the entire set of 16 trials.

In this study, the memory game conditions were not counterbalanced for order of administration. While we have since slightly changed this practice, the original rationale was as follows: In pilot work with the memory game task (e.g., Cleary, 1997), we found that significant differences could be obtained between groups on overall performance (averaging between a perceptually more discriminable and perceptually less discriminable condition) simply depending on the order in which conditions were administered—that is, the groups of adult subjects given the “harder to discriminate stimuli” condition first did significantly worse over all conditions than the group given the reverse order. This order effect seemed to overshadow any practice effect that might have improved performance over the course of the experiment. Therefore, the decision was made to administer what in pilot work seemed to be the “easier” condition first. That is, we administered the digit-names first, the confusable syllable condition second, and the “silent” condition last. We reasoned that if there were practice effects that would improve performance over the course of the different conditions, these effects would work conservatively against finding differences between the two auditory-plus-visual/spatial conditions, assuming we put the perceptually easier condition first. The silent condition was administered last, both because it appears to be a somewhat more difficult condition, and also to avoid encouraging subjects from adopting a “lights-only” strategy throughout all the other experimental conditions.

The children's forward and backward digit span scores were then obtained using the number lists, directions, and scoring procedures described in the Wechsler Intelligence Scale for Children - Third Edition (Wechsler, 1974; WISC-III, 1991). The digits were spoken by the experimenter with stimulus presentation times of slightly over one digit per second. The child received one point for each correct repetition of a list of digits. The child was said to “reach ceiling” (ending testing) when both lists at a given list length were incorrectly repeated. Next, the child's backward digit span was calculated. The child received one point for each correct backward repetition of a given list. The ceiling (or baseline, in the case of the three-year-old children who did not always understand the concept “backward”) was obtained when the child failed to correctly produce in reverse order two successive lists of a given length. The WISC-III was scored based on the children's total points (the sum of the digit forward and digit backward points). Additionally, the number of digits at the highest list length that a child was able to correctly repeat at least once (ALO) was recorded for each of the WISC span tests, both forward and backward.

Finally, the children's receptive vocabulary was measured using the instructions, scoring procedure, and stimuli from the Peabody Picture Vocabulary Test (PPVT-III, Form III-A, Dunn & Dunn,

1997). The children were presented with four picture plates at a time and were instructed to point to the plate that corresponded to the word the experimenter spoke aloud.

Children received stickers to motivate them to continue during the experiment. In the nonword repetition task, children moved a playing piece around a game board and were allowed to advance for each word repeated. (This last device proved very useful for quick and smooth completion of the nonword repetition task.) Upon completion of the session, participants received payment of five dollars and a T-shirt with the laboratory logo, as well as a certificate of participation.

Results and Discussion

Descriptive Statistics. Table 1 displays descriptive statistics across all the children for each of the tasks. The average age of the children tested was 53.44 months (approximately 4 years, 5 months of age). The group means for standardized scores on both the expressive vocabulary measure (EVT), as well as the receptive vocabulary measure were within one standard deviation of the published means of 100 (SD = 15). This shows that our sample in general scored somewhat above average compared to the sample these tests were normed on, but still within normal limits.

Table 1
Average Age and Scores Across Participants

	Mean	SD	Min	Max
Age (in months)	53.44	8.44	41	68
EVT standard scores	112.04	10.44	92	134
EVT raw scores	53.20	10.66	36	74
PPVT-III standard scores	114.69	12.48	79	136
PPVT-III raw scores	75.18	18.89	33	111
WISC-III forward scores	4.42	0.94	3	7
WISC-III backward scores	1.60	1.37	0	3
WISC-III total point scores	7.80	2.83	3	13
Memory game digits	2.96	1.17	1	5
Memory game syllables	2.98	0.97	1	5
Memory game silent	2.78	1.08	1	5
CNRep nonword scores	1.63	0.23	0.90	1.93
CTPP nonword scores	1.37	0.25	0.52	1.72

Note. N = 45. The scores for the WISC-III forward and backward digit spans were the highest number of digits correctly repeated at least once (ALO). The scores for the memory game tasks were the highest number of stimuli correctly repeated at least once (ALO). The CNRep and CTPP scoring assumes that a perfect score would be 2.0.

From Table 1, it can be observed that average scores for the memory game tasks were lower than the means obtained for the WISC-III forward digit repetition spans. The children obtained similar scores on the memory game task conditions that included auditory stimuli, as indicated by the similar means (2.96 vs. 2.98 items), and the fact that the range of scores on each of the tasks was the same (range = 1-5). Overall, the scores on the memory game silent condition were slightly lower with the mean score of 2.78. Differences between memory game condition means were not significant by a repeated-measures analysis of variance ($F(2, 88) = 1.18, p = 0.31$). These results taken alone fail to support the hypothesis that the auditory stimuli are contributing to the children's performance on the memory game task. (More detailed analyses however provide a somewhat different picture, as will be described.)

On average, the children scored a 1.63 out of 2.00 on nonwords taken from the CNRep, while on CTTTP this average score was 1.37 out of 2.00. This reflects the greater difficulty of the CTTTP nonword repetition test. Note from Appendix A, that, unlike the CNRep, the CTTTP test contains four 6-syllable and two 7-syllable items (although this is partially offset by its inclusion of four monosyllabic words). The CTTTP test also specifically avoids including items that contain syllables that sound like familiar words. This characteristic is reflected in adults' ratings of the CTTTP nonwords as less word-like than the CNRep nonwords. The difference between the mean scores demonstrates that the children are likely using their long-term phonological knowledge of real words to increase the number of novel sound sequences correctly reproduced from memory. However note, that the simple correlation between a nonsense word's average accuracy of repetition by the children and the mean adult judgement of word-likeness of that word was determined to be only $r = +0.23$. The length of the nonword and its phonemic composition strongly influence both its judged "word-likeness," and the probability that a child will be able to produce a subjectively "acceptable" token of the item, but the fit between the adult word-likeness ratings and child production accuracy measures is not very good.

Simple Bivariate Correlations. One of the goals of this study was to assess whether children's performance on the new memory game task showed the same gains as a function of age as the more traditional WISC verbal memory span tasks. Significant moderate positive correlations were obtained between the memory game tasks and age as shown in Table 2. When the WISC-III digit-spans were correlated with age, very similar moderate positive correlations were also obtained, as shown in Table 2.

Table 2 also provides the simple bivariate correlations obtained between the different memory span measures. Note that since age is a common variable in every case, to obtain a true estimate of how much these tasks have in common, it will be necessary to control for this factor.

Table 2

Simple Bivariate Correlations Showing the Effects of Age on All Simple Memory Span Tasks and Correlations between Span Tasks Before Controlling for Age (N=45)

	Age (months)	Digit-name Memory Game	Syllable Memory Game	Silent/Lights Only Memory Game	Forward WISC Digit-span	Backward WISC Digit-span
Age (months)		.51**	.55**	.60**	.57**	.69**
<i>Digit-name Memory game</i>			.69**	.59**	.43**	.59**
<i>Syllable Memory game</i>				.56**	.54**	.58*
<i>Silent/Lights Only Memory Game</i>					.47**	.61**
<i>Forward WISC Digit-span</i>						.52**
<i>Backward WISC Digit-span</i>						

Note: Memory game scores and digit-span scores reflect the highest number of correct repetitions that each child obtained at least once (ALO).

** = Correlation significant at the .01 level, but in italicized correlations, age effect has not been controlled for.

Insert Figure 2 about here.

Figure 2 allows a comparison of the scatter of individual scores as a function of age on each of the simple memory span measures. The proportion of variance accounted for by differences in age is shown to the right of each plot. From this it can be seen that children's scores on the memory game tasks improve (cross-sectionally) as a function of age at approximately the same rate as the children's scores on the WISC-III forward-recall digit-span task. On the backwards WISC task, note that many of the children were at floor on the task and were unable to correctly recall even the minimum two digits in reverse order.

Partial Correlations. A correlational analysis with the effects of age partialled out was carried out. Table 3 displays the partial correlation matrix for the measures under investigation.

Note. N = 45. * = $p < .05$. ** = $p < .01$. Note that some further adjustment of the alpha level is called for since so many correlations were computed. The scores for the WISC forward and backward digit spans were the highest number of digits correctly repeated at least once. Note also that scores on the WISC backward digit span task suffered from floor effects with a large number of children scoring zero on this task. WISC Total Points scoring combines both forward and backward performance by awarding one point for per list correctly repeated, regardless of condition (forward or backward), and summing these.

First, we briefly note that the two measures of vocabulary (the PPVT-III and the EVT) were significantly correlated with each other ($r = 0.55$, $p < .001$). Although this value is somewhat smaller than expected given previous findings suggesting that these standardized expressive and receptive vocabulary measures are correlated about +0.79 (EVT, Williams, 1997), we have determined that the origin of this difference is data contributed primarily from our four-year-old participants. Williams (1997) has reported that the correlation between the PPVT-III Form IIIA receptive measure and the EVT expressive vocabulary measure is slightly weaker (+0.66) between four and five years of age than in other age groups.

We will focus on three aspects of Table 3 that are relevant to the goals of this study. First, we address the question of how closely performance on the new memory game task resembles performance on the WISC forward digit span task. From the intersection of Row 3 with Columns 6, 7, and 8 in Table 3, we can see that while the correlations are relatively small, they are all positive in value as one would expect. Somewhat unexpectedly, the nonsense syllable condition of the memory game task is better correlated with the WISC digit-span measure than is the memory game digit-name condition.

Next, we address the issue of whether the different memory game conditions can be distinguished. Note from the intersection of Row 6 with Column 7 and the two adjoining cells, that the two auditory-plus visual conditions of the memory game task are correlated more strongly with each other than either is with the silent/lights-only version of the task. Also note that, replicating previously reported findings in the literature, WISC forward digit-span (Column 3) is a worse predictor of vocabulary scores, both expressive and receptive, than is performance on the Gathercole et al. CNRep nonword repetition task. Scores on the memory game conditions that included auditory stimuli can be seen from the intersection of Columns 6 and 7 with Rows 1 and 2, to be better predictors of standardized vocabulary scores than the WISC forward digit span measure. The “silent”/lights only memory game condition, however does not share this pattern, being a noticeably much weaker predictor of standardized vocabulary scores than the other two memory game conditions. In terms of proportion of variance accounted for, the memory game conditions that included an auditory component account for about 12-16% of the variance in vocabulary scores, compared to only about 2% of the variance accounted for by the memory game silent condition and about 5% in the case of WISC-forward scores.

As one further note, observe that the two nonword repetition tasks are correlated only +0.41 with each other and differ considerably from each other in terms of how well they predict performance on the other measures. In all but one case (glaringly, the silent/lights-only condition of the memory game), the nonword test containing more word-like items, (the CNRep, Column 9), is more strongly correlated with each of the other measures than is performance on the CTPP (Column 10). Gathercole et al. have reported correlations of about .40 between receptive vocabulary and the CNRep in 4- and 5-year olds (Gathercole et al., 1994), and about .33 in 2- and 3-year olds (Gathercole & Adams, 1993). Our obtained values replicate these findings.

Analysis of Groups Defined by Split at Median Vocabulary Scores. The partial-correlation matrix is somewhat clumsy and unwieldy to deal with. Moreover, it is difficult to tell much from the matrix about just how differently the children with high vocabulary scores for their age were doing on the various tasks compared to children who obtained lower vocabulary scores for their age. Therefore, we split the 45 subjects into two groups according to whether they fell above or below the sample’s median standard score on the expressive vocabulary test, and in a separate calculation, according to whether they fell above or below the group median standard score on the receptive vocabulary measure. Fortunately, in each case, the mean age and spread of ages in each of these groups was nearly equal or slightly in the direction opposite than that which would cause unwanted age-related differences between the newly-defined groups. Table 4 shows scores on the memory game and nonword repetition tasks by expressive

vocabulary group. Table 5 provides the same information for the groups defined by receptive vocabulary group.

Table 4. Group Statistics Split by Expressive Vocabulary EVT Median Standard Score (= 113)

	N1=23, mean age = 52.39 months N2=22, mean age = 54.55 months	Mean	Std. Deviation
CNRep Nonword Repetition	Above Median	1.67	.19
	Below Median	1.58	.26
CTPP Nonword Repetition	Above Median	1.41	.19
	Below Median	1.33	.29
Digit-name Memory Game	Above Median	3.04	1.07
	Below Median	2.86	1.28
Syllable Memory Game	Above Median	3.09	.90
	Below Median	2.86	1.04
Silent/Lights Only Memory Game	Above Median	2.78	.95
	Below Median	2.77	1.23

*Significant at $p < .05$ level, two-tailed t-test for independent samples

**Significant at $p < .01$ level, two-tailed t-test for independent samples

Table 5. Group Statistics Split by Receptive Vocabulary PPVT Median Standard Score (= 117)

	N1 = 23, mean age = 53.04 months N2 = 22, mean age = 53.86 months	Mean	Std. Deviation
CNRep Nonword Repetition	Above Median	1.71*	0.17
	Below Median	1.55	0.26
CTPP Nonword Repetition	Above Median	1.37	0.25
	Below Median	1.36	0.25
Digit-name Memory Game	Above Median	3.39**	1.16
	Below Median	2.50	1.01
Syllable Memory Game	Above Median	3.30*	1.11
	Below Median	2.64	0.66
Silent/Lights Only Memory Game	Above Median	2.91	0.95
	Below Median	2.64	1.22

*Significant at $p < .05$ level, two-tailed t-test for independent samples

**Significant at $p < .01$ level, two-tailed t-test for independent samples

Note that the groups score differently on the memory game conditions that use an auditory component, but perform equivalently (or nearly so) on the silent/lights-only condition. Tables 4 and 5 also show that although the “above median” children are doing somewhat better on the auditory-plus visual/spatial memory game conditions than the silent condition, in the fashion we had expected, the “below median” group is not showing this consistent pattern of differences. The “below median” group is doing near equivalently on the silent memory game condition compared to the auditory-component conditions, contrary to our general expectation. We interpret this result as suggesting that children with differing levels of verbal ability as measured by standardized vocabulary tests may, at the level of the individual child, deal differently with our new memory game task across the different conditions. We suggest that the “below median” children in our sample were less able to utilize the additional auditory cues in encoding and recalling auditory-plus-light sequences than the “above median” children.

These results are not as robust as we would like, due in part, we felt, to the use of a simple integer-only unit scoring of span. We are in the process of re-scoring the memory game data in another fashion in the hopes of obtaining a more sensitive measure of individual differences by using a continuous rather than discrete span measure. In essence, this method applies the point-method of scoring of the WISC task to our memory game task. A score in a given condition is obtained by summing the proportion of trials correctly recalled at each length. This method is similar to the point system of scoring the WISC memory span task except that it avoids the problem of the memory game algorithm having a variable number of trials at each length administered depending on the subject’s performance. A portion of the data has been scored already, and this method looks more promising in terms of expanding the measurement scale.

General Discussion

The results of this experiment address several important issues. First, the results demonstrate that the memory span reproduction task can be used with children as young as three years, five months of age. Children aged 3;5 to 5;8 obtained memory spans ranging from one to five items. For practical reasons, this is an important methodological finding in terms of future use of this procedure with several clinical populations.

Second, the positive correlation (statistically controlling for age) between the memory game and the WISC-III ($r = .32, p < .05$) demonstrates that the memory game reflects processes that are similar, but not identical, to those measured by the WISC-III forward digit-span task.

Third, the differences in performance attributable to inclusion of auditory, as well as visual stimuli, in the new memory game task are subtle but clearly present and operative. Some subjects appear to be less inclined than others to use the additional redundant stimulus input to increase efficient encoding of presented sequences, and it may be partly predictable as to which children these are. This is interesting in and of itself, but raises some additional issues concerning use of the memory span task in isolation (rather than as part of a larger battery of verbal and nonverbal measures) with individual subjects, assuming that conclusions are to be drawn from an individual child’s performance (e.g., a pediatric cochlear implant user or candidate).

Finally, when using the same methods and scoring procedures as previous investigators, the current results replicated findings from Gathercole and colleagues (Gathercole, Willis, & Baddeley, 1991; Gathercole et al., 1994) regarding positive correlations between the CNRep nonword repetition test and receptive vocabulary, and smaller such correlations between the WISC simple forward digit span task and the same vocabulary measure. Our new memory span task in the auditory-plus-visual/spatial conditions is

also significantly correlated with receptive vocabulary measures approximately the same degree as the CNRep nonword repetition measure. The silent/lights-only memory game condition shows only a weak to nonexistent correlation, however, with this same vocabulary measure. The present results also demonstrate that the phonological properties and structural characteristics of the nonsense words used affect not only scores on the nonword repetition test, but also how well this measure is able to predict vocabulary achievement in individual children.

The utterances from the nonword repetition portion of the current study are now being measured and analyzed to determine whether the DAT audio recordings of each child's speech can provide information about individual differences in speech rate and speed of initiating a nonword repetition response. These measures are relevant to the role of articulatory rehearsal in determining memory span, (Schweickert, et al., 1990) and to the processing of novel or low frequency spoken words.

Future studies using the memory game task will extend our investigation to include older children. In addition, we plan to test a variety of clinical populations including children with hearing impairments, children with cochlear implants, and children with phonological disorders. Finally, modifications to the basic memory game paradigm will examine the effects of strictly auditory-only presentation conditions using lists comprised of color-names. For example, a child will hear the lists "red, blue, yellow" without the corresponding buttons illuminating, but will make his/her response by pressing the appropriately matched button sequence. We are also introducing a learning condition in which an initial pattern is repeated, and then increased by one stimulus item after every correct repetition, in order to explore how young children learn and encode patterned sequences under these conditions.

In summary, this study of normal-hearing three- to five-year old children has provided additional new data supporting the validity of a new reproductive memory span procedure for assessing working memory in a variety of populations, potentially including populations with limited verbal output capabilities. Data showing how this new memory game task compares to more conventional measures of verbal memory span have been provided, and the influence of using auditory input in conjunction with visual/spatial input, as opposed to only visual/spatial input, has been assessed. Finally we have shown that subdividing our sample of children according to standardized vocabulary scores results in a different pattern of performance across the memory game conditions in the two groups. These results provide useful data for further investigation of the development of phonological representations in working memory using this new task.

References

- Baddeley, A.D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 18, 302-309.
- Baddeley, A.D. (1992). Working memory. *Science*, 255, 556-559.
- Baddeley, A.D. (1990). *Human Memory: Theory and Practice*. Boston: Allyn and Bacon.
- Baddeley, A.D., Gathercole, S., & 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.H. Bower (Ed.), *The Psychology of Learning and Motivation Volume 8* (pp. 47-90). New York: Academic Press.

- Baddeley, A.D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 36A, 233-252.
- 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.
- Bishop, D.V.M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioral marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry*, 37, 391-403.
- Bowey, J.A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology*, 63, 44-78.
- Cleary, M. (1997). Measures of phonological memory span for sounds differing in discriminability: some preliminary findings. In *Research on Spoken Language Processing Progress Report No. 21*. Pp. 93-140. Bloomington, IN: Speech Research Laboratory, Indiana University.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, 55, 75-84.
- Dedina, M.J. (1987). SAP: A speech acquisition program for the SRL-VAX. In *Research on Speech Perception Progress Report No. 13*, Pp. 331-337. Bloomington, IN: Speech Research Laboratory, Indiana University.
- Dempster, F.N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Review*, 89, 63-100.
- Dollaghan, C. & Campbell, T.F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1136-1146.
- Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test – III*. Circle Pines, MN: American Guidance Service.
- Dunn, L.M. & Dunn, L.M. (1982). *British Picture Vocabulary Scale*. Windsor: NFER-Nelson Publishing Company Ltd.
- Elliot, C.D. (1983). *British Abilities Scales*. Windsor: NFER-Nelson Publishing Company Ltd.
- Ellis, N.C. & Hennelly, R.A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology*, 71, 43-51.
- Gathercole, S.E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 23, 83-94.
- Gathercole, S.E. & Adams, A. (1993). Phonological memory in very young children. *Developmental Psychology*, 29, 770-778.
- Gathercole, S.E., Hitch, G.J., Service, E., & Martin, A.J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, 33, 966-979.

- Gathercole, S.E., Willis, C.S., & Baddeley, A.D. (1991). The influences of number of syllables and word-likeness on children's repetitions of nonwords. *Applied Psycholinguistics*, *12*, 349-367.
- Gathercole, S.E., Willis, C.S., Baddeley, A.D., & Emslie, H. (1994). The children's test of phonological working memory. *Memory*, *2*, 103-127.
- Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*, 267-333.
- Hernandez, L (1995). Current computer facilities in the Speech Research Laboratory. In *Research on Spoken Language Processing Progress Report No. 20*. Bloomington, IN: Speech Research Laboratory, Indiana University, 389-393.
- Kirk, K.I., Pisoni, D. B., & Osberger, M.J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing*, *16*, 470-481.
- Kirk, K.I., Pisoni, D.B., & Robbins, A.M. (1997). Assessing speech perception in children. In Mendel, L.L. and Danhauer, J.L. (Ed.), *Audiologic Evaluation and Management and Speech Perception Assessment*. San Diego: Singular Publishing Group Inc.
- Large, N., Frisch, S.A., & Pisoni, D.B. (1998). Perception in wordlikeness: Effects of segment probability and length on subjective ratings and processing of non-word sound patterns. In *Research on Spoken Language Processing Progress Report No. 22*. Pp. 101-132. Bloomington, IN: Speech Research Laboratory, Indiana University.
- Lewellen, M.J., Goldinger, S.D., Pisoni, D.B., & Greene, B.G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, *122*, 316-330.
- Lezak, M.D. (1995). *Neuropsychological Assessment, 3rd Edition*. New York: Oxford University Press.
- Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*, *19*, 1-39.
- Marslen-Wilson, W.D. (1987). Function and process in spoken word recognition. *Cognition*, *8*, 1-71.
- Naveh-Benjamin, M. & Ayres, T.J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology*, *38A*, 739-751.
- Pisoni, D.B. (1997). Cognitive factors and cochlear implants: An overview of the role of perception, attention, learning, and memory in speech perception. In *Research on Spoken Language Processing Progress Report No. 21*. Pp. 335-366. Bloomington, IN: Speech Research Laboratory, Indiana University.
- Pisoni, D.B., Svirsky, M.A., Kirk, K.I., & Miyamoto, R.T. (1997). Looking at "stars": A first report on the intercorrelations among measures of speech perception, intelligibility and language in pediatric cochlear implant users. Paper presented at the Vth International Cochlear Implant Conference, May 1-3, 1997, New York City, NY.

- Salame, P. & Baddeley, A.D. (1982) Disruption of short-term memory by unattended speech. Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21,150-164.
- Schweickert, R., Guentert, L, & Hersberger, L. (1990). Phonological similarity, pronunciation rate, and memory span. *Psychological Science*, 1, 74-77.
- Shriberg, L.D. & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research*, 37, 1100-1126.
- Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, 12, 369-373.
- Torgeson, J. & Wagner, R. (1997). *Comprehensive Test of Phonological Processes (CTPP)*. Austin, TX: Pro-Ed.
- Vitevitch, M. (1998). All neighborhoods are not created equal: The phonological P-metric and spoken word recognition. In *Research on Spoken Language Processing Progress Report No. 22*. Pp. 317-328. Bloomington, IN: Speech Research Laboratory, Indiana University.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children: Revised*. New York: Psychological Corporation.
- Wells, B., (1995). Phonological considerations in repetition tests. *Cognitive Neuropsychology*, 12, 847-855.

Appendix A

	Gathercole CNRep: (10 words each of length 2, 3, 4 & 5 syllables)	# o f s y l l	Measured Acoustic Durations (ms)	Transcription of recorded stimuli	Mean Success in Reproduction Score (Subjective), 2 = "correct reproduction, 1 = one or more errors, 0 = no attempt	Adults' mean ratings of word- likeness 6 = very word-like, 0 = not word-like	Adults' Median Reaction Time to make judgement of word- likeness
1.	altupatory	5	1046	Θλ :τυ π↔, το ρι	1.48	3.33	853
2.	balop	2	547	:βΘ λ↔π	1.82	4	1464
3.	bannifer	3	779	:βΘ υI/↔ φ↔↓	1.70	3.75	1339
4.	bannow	2	738	:βΘ ναΥ	1.76	3	1122
5.	barrizen	3	887	:βE ρ↔ ζI/↔ν	1.62	3.75	747
6.	blonterstaping	4	1280	:βλAv τ↔↓, στε πIN	1.64	1.83	979
7.	brasterer	3	881	:βρΘ στ↔↓, ε ↓	1.30	2.5	1113
8.	commerine	3	804	:κA μ↔↓ (ρ)Iv	1.88	4.08	1468
9.	commesatate	4	1104	κ↔: μι σ↔/I, τετ	1.74	3.5	1321
10.	confrantually	5	1318	κ↔ν: φρΘν τΣυ ↔ λι	1.58	4.08	963
11.	contrampanist	4	1248	κ↔ν: τρΘμ π↔ νIστ	1.54	4	1359
12.	defermication	5	1202	, δι φ↔↓ μI/↔: κε Σ(↔)ν	1.52	4.75	831
13.	detratapilic	5	1163	, δι τρΘ P/τ↔: πI λIκ	1.52	3.25	1396
14.	diller	2	481	:δI λ↔↓	1.84	4.5	1501
15.	dopalate	3	942	:δA π↔ λετ	1.78	4.17	1290
16.	emplifervent	4	1186	, Εμ : πλI φ↔↓ ϖEντ	1.44	2.83	900
17.	fennerizer	4	1060	φEν ↔ ↓, ραI ζ↔↓	1.66	3.33	868
18.	freskavent	3	1010	:φρE σκ↔/I ϖEντ	1.36	3.67	1247
19.	glistering	3	925	:γλI στ↔↓ IN	1.68	4.75	1036
20.	glistow	2	811	:γλIσ ταΥ	1.66	2.67	924
21.	hampent	2	846	:ηΘμ πEντ	1.94	4.25	1280
22.	loddenappish	4	1171	, λA δ↔/I : νΘ π↔/IΣ	1.42	1.42	1255
23.	pennel	2	529	:πEν λ ↔↔λ	1.86	3.50	1331
24.	penneriful	4	988	π↔: νE ρ↔, φ↔λ	1.72	3.50	1050
25.	perplisteronk	4	1172	π↔↓: πλI στ↔↓, ρA Nκ	1.22	2.92	1378
26.	prindle	2	635	:πρIv δλ ↔↔λ	1.86	4.25	1319
27.	pristeractional	5	1237	, πρI στ↔↓: (ρ)Θκ Σ ↔↔ν λ	1.34	3.08	1250
28.	reutterpation	5	1089	, ρι Θ/ ρ P/τ↔↓: πε Σ ν	1.54	3.50	1066
29.	rubid	2	676	:ρυ βIδ	1.70	4.58	1457
30.	sepretenial	5	1138	, σE πρ↔: τE νι λ ↔ ↔↔λ	1.46	5.08	1313
31.	skiticult	3	1153	:σκI P/τI, κ ρλτ	1.64	3.17	926
32.	sladding	2	855	:σλΘ δIN	1.72	3.83	1346
33.	stopagratc	4	1195	, στA π↔: γρΘ P/τIκ	1.62	3.83	984
34.	tafflist	2	901	:τΘ φλ↔/Iστ	1.66	4.08	1533
35.	thickery	3	773	:TI κ↔↓ ι	1.74	4.83	1174

36.	trumpetine	3	930	:τρ ρ μ π ↔ ,τιν	1.84	4.25	808
37.	underbrantuang	5	1418	,↔ν δ ↔ ↓ βρΘν τΣυ Ανδ	1.38	3.33	1061
38.	versatrationist	5	1425	,π ↔ ↓ σ ↔ :τρε Σ ↔ ν Ιστ	1.52	4.08	744
39.	voltularity	5	1267	,π/βαλ τΣυ :λε ρΙ τι	1.40	3.33	879
40.	woogalamic	4	1021	,ω/βυ γ ↔ :λα μκκ	1.48	1.33	1115
				Sum	64.58	143.88	
				Mean	1.61	3.60	

	Torgeson & Wagner (CTPP Subtest 10): (4 words each of length 1, 2, 3, 4 & 6 syllables; 3 5-syllable words, 2 7-syllable words)	# o f s y l l e s	Measured Acoustic Durations (ms)	Transcription of recorded stimuli	Mean Success in Reproduction Score (Subjective), 2 = "correct reproduction, 1 = one or more errors, 0 = no attempt	Adults' mean ratings of word-likeness 6 = very word-like, 0 = not word- like	Adults' Median Reaction Time to make judgement of word- likeness
1.	berlugahgendaplo	6	1362	β↔↘↓, λυ γ↔ :δZEν δ↔ , πλο	1.20	0.92	872
2.	byleedouge	3	1114	,βAI λι :δοδZ	1.59	1.00	1146
3.	chassydolid	4	1228	:τΣΘ σι ,δυ λιδ	1.57	1.75	1433
4.	dookershatapytazom	7	1726	δυ κ↔↘↓ ΣΘ τ↔ :πAI τ↔ ζAμ	0.88	0.17	226
5.	gekeezysikad	5	1475	γE κι :ζAI σε κΘδ	0.96	0.50	519
6.	jupe	1	479	:δZυπ	1.59	4.00	1460
7.	kablook	2	707	κ↔ :βλυκ	1.80	3.08	1260
8.	lissashrull	3	994	:λI σ↔/I Σρ↔λ	1.45	1.25	1330
9.	mawgeebooshernew- shike	6	1615	:μA γι :βυ Σ↔↘↓ :νυ ΣAIκ	0.88	0.17	493
10.	meb	1	725	μΘβ	1.84	2.92	1481
11.	myjaxsem	3	1175	μAI :δZΘκ σμ ↗↔μ	1.63	1.75	935
12.	naigon(g)	2	928	νAI :γA/□ν	1.80	1.75	1047
13.	nyesheengitherpoidge	5	1880	,νAI Σιν :γI T↔↘↓ ποIδZ	1.41	0.25	989
14.	pate	1	621	πετ	1.63	4.42	1344
15.	runyeteeayusip	6	1383	ρυ ,νAI τι ↔↘↓ :υ σ ρ/Eπ	1.08	0.50	1277
16.	shaberyehuvoimush	6	1750	:ΣΘ β↔↘↓ (ρ)AI :η ρ ω/βo I μ↔Σ	0.88	0.33	670
17.	slekizoybreshatitadge	7	1941	,σλ ρ κι ζ/ωI βρI :ΣΘ τι τεδZ	0.80	0.08	698
18.	soynewrayherdast	5	1779	,σoI νυ :ρε η↔↘↓ δΘστ	0.98	0.50	950
19.	teebutteshahalt	4	1213	τι :β ρ P/τAI Σ□/Aλτ	1.31	0.58	1062
20.	viversoomoudge	4	1371	,β/ωI ω↔↘↓ :συ μΘλδZ	0.98	0.58	756
21.	vozitoove	3	1065	,ωo ζI :τυω	1.47	1.58	1036
22.	wivoo	2	678	ωI :ωυ	1.49	1.25	1233
23.	wudoip	2	739	ω/ωυ :δοIπ	1.61	1.08	1022
24.	wulaniwup	4	1013	ωυ :λA ν↔/I ωYπ	1.39	0.83	710
25.	zid	1	554	ζIδ	1.76	2.83	1240
				Sum	33.98	34.07	
				Mean	1.36	1.36	

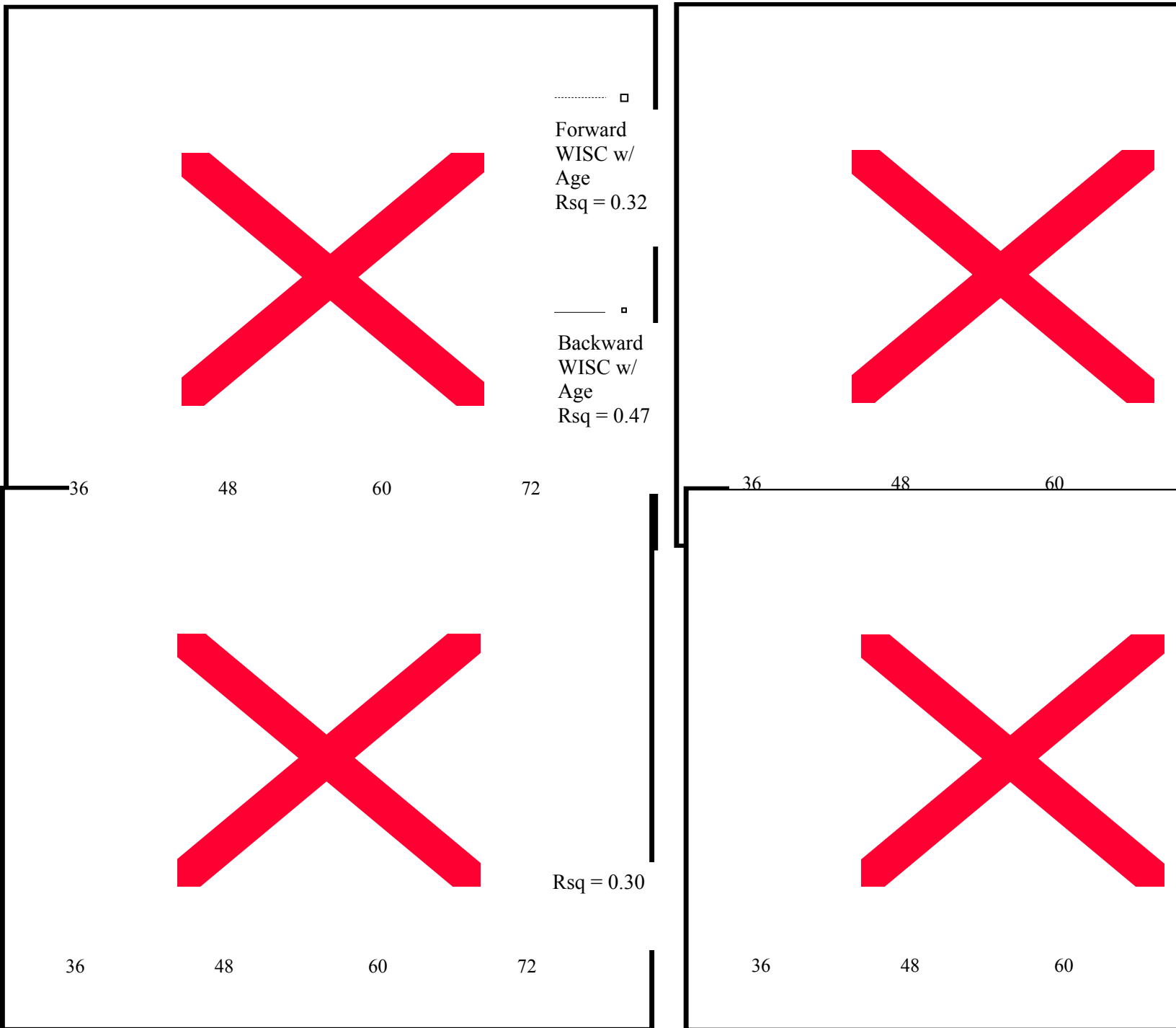


Figure 2. Effects of age compared across different memory span measures and memory game conditions.