Journal of Experimental Psychology:
Learning, Memory, and Cognition
1990, Vol. 16, No. 6, 1118–1133

0278-7393/90/$00.75

Simple and Complex Word Spans as Measures of Working Memory Capacity

Linda B. La Pointe and Randall W. Engle
University of South Carolina

Five experiments investigated the effects of word length in simple word span tasks and complex operation and reading span tasks and the relationship between these tasks and reading comprehension. The first 2 experiments showed word length effects using both simple and complex memory span tasks and that both simple and complex span tasks correlated with reading comprehension. In the third experiment, articulatory suppression did not eliminate word length effects. The final experiments showed that articulatory suppression eliminated the effect of word length when words were sampled with replacement from small fixed pools but not when sampled without replacement from a large pool. The word pool effects were not a result of concreteness of the words. We conclude that the reading span does not measure a working memory specific to reading. Further, in immediate memory experiments, repeating words from trial to trial may lead to a more limited coding than is used with nonrepeated words.

Most global models of reading argue that the temporary retention of recently read text and recently activated knowledge structures is important to overall comprehension (Just & Carpenter, 1980; Kintsch & van Dijk, 1978). This would seem to implicate the type of retention studied as short-term memory for the last 30 years. But digit span, the most traditional of STM tasks, does not distinguish between good and poor readers (Guyer & Friedman, 1975; Perfetti & Lesgold, 1977).

Daneman and Carpenter (1980) argued that tasks, like the digit span, measure only a storage component and that, in order to observe the impact of individual differences in temporary retention on reading, a task must measure processing as well. They developed a task that required the subject to read a set of sentences aloud and, afterward, to recall the last word of each sentence. The number of words recalled was defined as the reading span. This measure has been shown to distinguish very well between good and poor readers. Daneman and Carpenter found it to predict performance on several specific types of comprehension questions as well as more global measures like the Verbal Scholastic Aptitude Test (VSAT). The significant and sizable correlation between the reading span and reading comprehension has been found in many different laboratories and situations (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Daneman & Carpenter, 1983; Dixon, Le Fevre, & Twilley, 1989; Masson & Miller, 1983).

This research was supported by AFOSR 870069 to Randall W. Engle.

We would like to acknowledge Regina Dillingham and Heather Sarratt for their help in testing subjects and Lee Kirkpatrick for extensive statistical advice. We would also like to thank the reviewers, especially Alan Baddeley, who spotted the potential confound in Experiments 1–4 and whose comments greatly improved both the research and the manuscript. Correspondence concerning this article should be addressed to Randall W. Engle or Linda B. La Pointe, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

The reading span task is really two tasks: The background task requires the subject to read a sentence aloud, and the primary task requires the subject to keep track of and recall a set of words. Daneman and Carpenter (1980) argued that the correlation occurs only because the background task requires reading processes similar to those required in the comprehension task being predicted. However, Turner and Engle (1989) found that the background component of this complex span task did not have to be related to reading. They replaced the sentence-reading component of the span task with a requirement to verify a string of arithmetic operations. Each of the operations was followed by a word or digit and the subject had to recall the words or digits that corresponded to the set of operations. Turner and Engle found that the number of items recalled in this complex span task predicted reading comprehension as well as did the reading span. This suggests that the complex span measures something more general than just resources or capacity available during reading.

Daneman and Carpenter (1980) and Turner and Engle (1989) found that a simple word span did not predict comprehension, so presumably the simple and complex word spans do not reflect exactly the same structures, processes or strategies. The goal of the research reported here was to begin an analysis of similarities and differences between the word span task, forms of which have been studied for dozens of years, and the newer complex span tasks.

One purpose of the experiments reported here was to pursue the finding of Daneman and Carpenter (1980) and Turner and Engle (1989) that the simple word span did not predict comprehension. The innovative importance of the complex span rests totally on the fact that the simple span does not predict comprehension. However, to date, very few studies have actually used a simple word span task that is comparable to the span component of a reading or operations span task. Daneman and Carpenter used the simple word span in only one experiment (Experiment 1, 1980) and found nonsignificant correlations of .33–.37 ($n = 19$) between word span and comprehension as measured by fact and pronoun questions, and a correlation of .35 between simple word span and the VSAT. Historically, STM tasks have consisted of repeatedly presenting the same set of words on each recall trial and required precise serial recall (Baddeley, 1986). By contrast, the complex span procedures have typically used a different set of words for each trial. Even though Daneman and Carpenter used different words for each list in their simple word span condition, they presented the words orally to the subjects while the reading span task required the subjects to read the sentences, and consequently, the to-be-remembered span words, aloud. While Turner and Engle (1989) used the same presentation conditions for their simple and complex span tasks, those presentation conditions were unusual. In order to test a very large number of subjects, Turner and Engle used a group testing situation with the subjects simultaneously seeing the items presented visually on a screen, hearing the items presented via tape recorder and saying the items aloud. It is possible that this procedure leads to effects (like articulatory suppression, for example) that are unlike a typical span procedure and we do not know whether this differentially affects the predictive validity of the simple and complex span tasks.

Since the absence of a significant correlation between simple word span and comprehension is so important to our thinking about the reading span and other complex span tasks, it is crucial that additional studies be done comparing the simple and complex span tasks and whether the two tasks respond the same way to different variables. We have known for some time, for example, that articulatory coding is important in typical word span tasks. Both phonological similarity and word length have proven to be powerful variables in short-term memory tasks and it has been argued that both occur because of articulatory coding (Baddeley, 1986). Baddeley, Thomson, and Buchanan (1975) showed, for example, that subjects recall more short words than long words and that the length of time to articulate the words is more important than the number of syllables. This supported the notion of a time-based articulatory code that facilitates recall in this kind of task.

Daneman and Carpenter (1980) would argue that the simple word span and the reading span are inherently different and, while articulatory coding might be important to recall in a simple word span task, it should not be much of a factor in complex spans like the reading span. If that is true, then the word length effect should not be found with the reading or operations span tasks.

Baddeley's (1986) theory of working memory argues that two components of working memory are important to reading, the central executive and the articulatory loop. We will refer here to the articulatory loop not as if it were an inherent structural aspect of human cognition, but rather as the result of a coding strategy, one of many which may be invoked during thinking. If the reading span task predicts comprehension because it measures the articulatory component, then it should be no surprise if word length has the same effect in the reading span as it has in the simple word span task. However, one complicating factor is that the way the reading span task is typically performed requires the subject to read sentences aloud which should serve as a source of articulatory suppression. Since concurrent articulation of irrelevant speech has been shown to eliminate the word length effect, at least with visual presentation, the word length effect might be diminished or absent with complex word spans.

Thus, the absence of a word length effect would not distinguish between Daneman and Carpenter's and Baddeley's view. But the presence of a word length effect with complex span tasks would cast doubt on the idea that simple word span and complex word span tasks are different in what they measure.

Our purpose then was to compare directly the simple and complex word span tasks using procedures that could be linked more legitimately with earlier literature and to investigate whether one variable, word length, has the same effect on the two types of tasks. To give away the punchline to our story, we will report two experiments showing that word length has the same effect on complex span tasks as on simple span tasks, that the simple span significantly predicts reading comprehension, and that the correlations may be as high as those obtained with the complex spans. In our third experiment, the word length effect is not eliminated with concurrent articulatory suppression in either simple or complex tasks. The fourth experiment showed that elimination of the word length effect using articulatory suppression depends on whether the span words come from a small fixed pool or are sampled without replacement from a larger pool. A final experiment showed that the results of the previous four experiments were not a result of a potential confound between word length and concreteness.

Experiment 1

Method

Subjects. Eighty undergraduate students at the University of South Carolina, all native speakers of English, participated. Each subject was tested individually and performed a simple word span and a reading span task over the course of approximately 40 min. They gave written permission to obtain their Verbal and Quantitative SAT scores from university records.

Design. A 2 (Word Length) × 2 (Order of Span Task) × 2 (Word-Group) × 2 (Experimenter) × 2 (Span Task) design was used. Word length, order of span task, word-group, and experimenter were between subjects factors. There were two levels of word length (long and short) and two orders of span task (simple-complex and complex-simple). The word-group factor also had two levels: 1) Set A for simple span and Set B for complex, and 2) Set B for simple span and Set A for complex. There were also two experimenters. The fifth factor, span task, was a within subjects variable with two levels. All subjects performed both the simple word span task and the complex reading span task. The between subjects factors were counterbalanced resulting in 16 separate groups.

Materials and stimuli. Stimuli for the span tasks were selected from a list compiled by Kucera and Francis (1967), in which words were ranked according to frequency of usage in the English language. A pool was created with the following constraints: Words were either one syllable long or three to four syllables long. For each one syllable word chosen, a closely ranked (within seven ranks) three to four syllable word was chosen. The next selected short word was as closely ranked to the one immediately prior to it as possible. In this manner a stimulus pool consisting of 162 short words and 162 long words was created. By alternating the assignment of the words to two groups (Set A and Set B), two sets of short and two sets of long words were created. Each set was comprised of 81 short words and 81 long words and matched as closely as possible for rank of usage in the English language. For comparison purposes prior to their use as stimuli, the words were divided into four sets: Set A short words, Set A long words, Set B short words, and Set B long words. An analysis of variance on word frequencies showed none of the sets significantly differed from each other. All of the words used in the experiments are shown in the Appendix.[1] The complex span task for this experiment was a reading span task with the to-be-remembered word presented following the punctuation at the end of the sentence. The to-be-remembered word was not part of the sentence. The words were drawn from the same Set A and Set B long and short words described above. Pairing of the word stimuli and sentences was random and in an order determined by the computer on which the stimuli were presented. A set of 162 sentences was composed which were 12 to 16 words long and randomly assigned to 2 groups of 81 sentences each. One group of sentences was used with short words and one was used with long words. Word length was a between subjects factor, thus every subject was presented with words of the same length in both the simple span and the complex reading span task. The use of word set was counterbalanced. Half of the subjects used Set A words with the simple span task and Set B words with the reading span task. Likewise, half used Set B in the simple span task and Set A in the reading span task. Some examples of sentences were: "Whenever we go to New York we visit my mother's favorite uncle.", "After being bitten last fall Ricky was afraid to pet the cat.", and "During the winter you can get a room at the beach at a low rate."

Reading span task. Sentences were presented on the monitor of an Apple IIe computer. Presentation of the sentences was controlled by the experimenter. Subjects were instructed to begin reading aloud as soon as a sentence appeared. They were to read the entire sentence and the word that followed the terminal punctuation. Immediately after the subject read the to-be-remembered word, the experimenter pressed a key that led to the presentation of the next sentence and the subject was to begin reading this sentence aloud immediately. The first trial consisted of two sentences. After presentation of the two sentences, a question mark was presented on the monitor, in the center of the screen. This signalled the end of the trial and served as a cue for the subject to begin recalling the word that followed each sentence. These two sentences constituted one trial. The number of sentences in a trial was gradually increased in the following manner. A set was completed when three trials of a given number of sentences had been presented, and the subject had attempted to recall the to-be-remembered words. When a set had been completed, the number of sentences in a trial was increased by one. The end of a trial was indicated by a question mark presented at the center of the monitor screen, whereupon subjects attempted to recall the words for that particular trial. Subjects had been advised that they could recall the words in any order they chose with the exception that the first word they wrote down should not be the last word they had read. The task continued until a set of trials consisting of five sentences had been completed. After completion of the five sentence set, the subjects stopped.

Word span task. In addition to the reading span task, each subject performed a simple word span task. The words were presented on the monitor of the computer at a rate of one per second. Subjects read each word aloud as it was presented and, when a question mark appeared on the screen, attempted to recall the words. Recall was written on a response sheet in any order the subject chose with the constraint that the last word could not be written first. The task began with a list length of two words. Three trials at a given list length constituted a set and when a set had been completed, the number of words in the list was increased by one. In the simple span task, subjects did not stop until a set of seven-word lists had been completed. Subjects who used set A words in the reading span tasks, were presented the set B words as stimuli for simple word span tasks, and subjects who used set B words in the reading span task were presented with set A words as word span stimuli. Each subject received a different random order of words.

Comprehension Test. As our measure of reading comprehension, we used the VSAT, published by Educational Testing Service. We chose this measure for several reasons: (a) the VSAT is a longer, more global test of reading comprehension and, although it includes a vocabulary component, it is clearly oriented toward reading comprehension. (b) Our previous research (Turner & Engle, 1989) had shown the VSAT to correlate as well, and frequently better, with the reading span and operations span tasks than other reading comprehension measures like the Nelson-Denny, (c) the VSAT is one of the most studied tests in the world and is given to hundreds of thousands of young Americans each year, (d) the VSAT was already available for nearly all of the subjects.

Results

Scoring. The data were scored in several ways. One method was very similar to that used by Daneman and Carpenter (1980). This score, called the strict span score, was the number of words that were recalled perfectly on two out of three trials of a given set size or list length. If a subject remembered all the words of one trial at the next higher set size perfectly, but could recall no more than one trial perfectly at that next set size, then half credit was given (0.5 points). An individual's recall for a trial was considered perfect (with this method, and all others considered here) if all words presented within a trial were remembered, regardless of serial order.

A second method of calculating scores required summing up the total number of words in each of the trials that were recalled perfectly. We called this measure the absolute span score. The third score, called the total span score, was the total of all the words correctly recalled. The difference between this score and the absolute score was that each word correctly recalled was included in the total span regardless of whether the trial led to perfect recall. Finally, we heeded Broadbent's (1971) admonition that the best way to measure STM may be to use the highest number of words individuals remember correctly in *all* trials rather than the point at which recall is perfect some portion of the time (e.g, 50%).

All four of these scoring methods were used for all of the experiments reported here but there were no major differences in conclusions drawn for the different scoring procedures. Therefore, the results of all five experiments will only describe the analyses of the absolute score. The maximum score possible was different for the simple span task and the reading span task. For this reason, scores were converted to proportions before analysis.

Analysis of variance. The primary question motivating this experiment, was whether a word length effect, similar to that observed previously with simple word span, would be

[1] It should be noted that the word *bible* was inadvertently included in the List A short word list even though it is a two syllable word. Several other words like "wire" are pronounced as two syllable words in our part of the world. The word "far" was inadvertently used in both List A and List B short word lists.

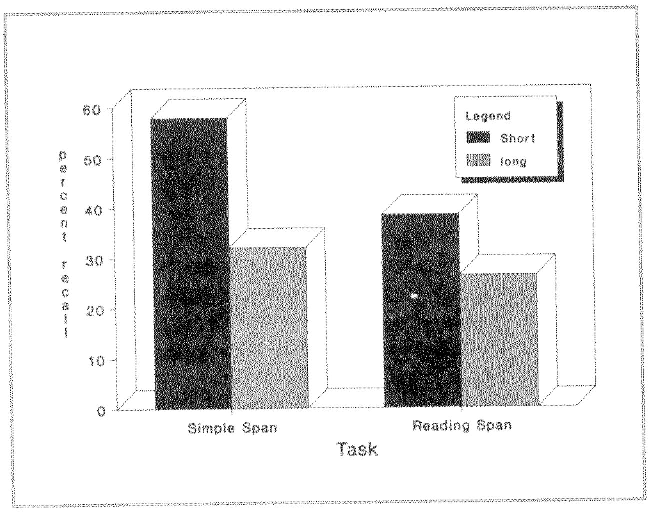

Figure 1. Experiment 1: Mean percent recall as a function of word length and task.

observed with the complex reading span task. As Figure 1 shows, subjects remembered more words in the simple span task than in the complex span task. They also remembered more short words than long words in both the simple span task and the complex reading span task.

These conclusions were based on effects that were significant at least at the $p < .05$ level. More short words were recalled than long words, $F(1, 64) = 41.86$, $MS_e = 345.6$, and more words were recalled in the simple span than reading span task, $F(1, 64) = 81.13$, $MS_e = 79.06$. There was also an interaction of span task and word length, $F(1, 64) = 23.92$, $MS_e = 79.06$, reflecting that the word length effect was larger for the simple span than it was for the reading span. Further analysis of this interaction revealed an effect of word length in the reading span, $F(1, 64) = 13.03$, $MS_e = 226.4$. So, while the effect was small, it was highly reliable.

It is not clear that the task by word length interaction has any interesting implications. The complex span tasks for all the experiments reported here are all more difficult and lead to poorer recall than the simple span task. Thus, it is likely the case that task main effects and interactions involving task occur because the complex span tasks are more constrained by performance boundaries than the simple span tasks.

The reading task is, however, clearly sensitive enough to show a word length effect and that is the important conclusion drawn from this analysis. It appears that the reading span task and simple span task are not as different as Daneman and Carpenter suggest and that they share at least some common processes.

A Word Length × Word group × Span Task interaction was also significant, $F(1, 64) = 6.37$, $MS_e = 79.06$. Further analysis of this interaction showed the word length effect to be significant, $F(1, 32) = 11.22$, $MS_e = 240.6$, in the complex task when set A words followed the sentences. However, the effect of word length was smaller and only marginally significant, $F(1, 32) = 2.91$, $p < .10$, $MS_e = 212.1$, with the complex task when set B words followed the sentences.

Reliability analysis. It has been argued (Dempster, 1985) that the reason simple span measures often do not correlate with reading comprehension and other measures of higher level cognitive functioning is that the word span measure is not reliable. For this reason we made a crude attempt to determine whether our tasks were internally consistent.

An analysis of the split part reliability of the absolute span scores was not appropriate because only those trials with perfect recall were included in that scoring. Thus, the reliability analysis was done on the total score (described earlier) that received a contribution from each trial. Subjects received 3 trials in each set with the set size ranging from 2–7 for the simple span and 2–5 for the complex span. Thus, 3 composite scores were calculated by randomly and exhaustively selecting one trial from each set size for each of the 3 composites. A reliability analysis using Cronbach's alpha showed the following reliabilities: Simple Span Short .74, Simple Span Long .82, Complex Span Short .68, and Complex Span Long .78. Thus, all four tasks showed relatively high reliability, and they were not much different from one another.

Correlational analysis. Table 1 shows the descriptive statistics for the absolute span score and Verbal, Quantitative, and Total SAT scores, although SAT scores were not obtained for three of our subjects. It should be noted that the sample included a sizable VSAT range (230–720) with a mean of 450. Table 2 shows that the simple span scores using both short words and long words correlated with both VSAT and QSAT, which were predicted by the reading span with short words but not by the reading span with long words. As mentioned previously, the reading span task leads to lower recall than the simple span task and the task becomes even more difficult when trying to recall 3–4 syllable words. Performance was probably so restricted in the reading span with long words that there was insufficient variability for a correlation to occur.

This experiment replicates other work in showing a sizable and significant correlation between the reading span task (at least with short words) and reading comprehension. What was different about this experiment, however, is that the simple span with both long and short words also predicted comprehension as measured by the VSAT.

Discussion

The first experiment showed that the length of the to-be-remembered words has the same effect in the reading span

Table 1
Descriptive Statistics: Experiment 1

Measure	*N*	*M*	*SD*	Min.	Max.
Verbal SAT	77	450	102	230	720
Quan SAT	77	482	89	300	700
Total SAT	77	933	175	580	1420
Simple Short	37	57.3	16	23.4	91.4
Simple Long	40	32.0	12	14.8	69.1
Complex Short	37	37.7	16	9.5	81.0
Complex Long	40	26.3	12.8	9.5	54.8

Note. SAT = Scholastic Aptitude Test.

Table 2
Correlations Between Absolute Scores and SAT

Measure	Verbal SAT	Quantitative SAT	Total SAT
Simple Short	.37*	.45**	.44**
Simple Long	.34*	.41*	.41*
Complex Short	.54**	.48**	.56**
Complex Long	.07	.07	.07

Note. SAT = Scholastic Aptitude Test.
* $p < .04$. ** $p < .006$.

task as has been repeatedly observed with the simple word span task. This calls into question any theory that the two tasks measure completely different aspects of temporary memory such a storage and processing. Whatever resources or processes the two tasks have in common may also be related to reading comprehension processes since both predict the Verbal SAT. It is our contention that while some of the processes supporting these tasks may be related to reading, they are not specific to reading. We believe that some of the processes used in performing the reading span task are general in nature, i.e., are important in tasks not directly reading-related and, that individual differences in those processes or resources will also be manifest in non-reading tasks.

Experiment 2

The purpose of Experiment 2 was to determine whether the word length effect in the complex span task would also be observed when the background component was not reading related. We used the operation span task used by Turner and Engle (1989) and manipulated the length of the words that were to be recalled in the span component.

Method

Subjects. Eighty undergraduate students at the University of South Carolina, all native speakers of English, participated for course credit. The method used in this experiment was similar to that used in Experiment 1. Subjects were seen individually and performed two tasks in approximately 40 minutes. Word length was a between subjects factor as in Experiment 1, and both a simple span and complex span task were performed by each subject. The simple word span task was performed exactly as it was in the previous experiment. However, the complex task in this experiment required the subject to verify arithmetic operations and remember words at the end of the operations.

Design. A mixed design with three between-subjects factors and one within-subjects factor was used. Word length was again a between-subjects factor as was task order. Subjects performed first either the simple task or the operations span task. The two possible levels of word-group specified which word set was used in each span task. If words from Set A were used in the simple word span task, then Set B words were used in the complex task and vice versa. The only within subjects factor was span task.

Materials and stimuli. The words that subjects attempted to recall in both the simple word span and the complex span task were drawn from the same Set A and Set B words used in Experiment 1. In Experiment 2, subjects read aloud and verified the accuracy of a series of mathematical operations, each of which was followed by a word. The operations used were generated in the following manner. Each operation consisted of two parts. The first component required the multiplication or the division of two integers; the second part of each operation required the subject to add or subtract an integer to the result of the first component. The operations were presented in the form of a question as is shown by the following examples: Is $(7 \times 2) + 3 = 17$? Is $(10/2) - 1 = 6$? These operations were generated by an Apple IIe computer, using a program which imposed the following restraints: (a) the two integers used in the multiplication/division component were integers between 1 and 10, (b) the result of the division or multiplication of these integers was an integer, (c) the number to be added to or subtracted from the first component was randomly generated and was an integer between 1 and 9, (d) determination of operations to be performed (i.e., multiplication/division and addition/subtraction) was random, and (e) answers to the operations were also presented to the subject for verification. Approximately half of the answers presented to the subject were incorrect but within two units of the correct answer. Determination of the operations presented with wrong answers was random. A pool of 162 operations was generated in the above described manner and these were divided into two groups. One group was used with short words and one group with long. Some further examples of the operations were: "IS $(8 \times 1) + 8 = 16$?", "IS $(7 \times 1) + 6 = 13$?", and "IS $(10 \times 2) + 3 = 23$?"

The operations described above were presented in the same manner as the sentences were in the complex task of Experiment 1. Each operation appeared on the monitor of an Apple IIe computer. A word from Set A or Set B was presented at the end of each operation after the answer. The pairing of operation to word was random without replacement and was determined by the program presenting the stimuli for each subject. The subject was instructed to begin reading the operation out loud as soon as it appeared, to immediately say "yes" or "no," indicating whether the operation answer was correct or not, and then to read the word aloud. Presentation was under the control of the experimenter, who also recorded the subject's responses regarding the accuracy of the operation while the next operation was read aloud. Subject and experimenter practiced prior to the task.

The operations span task began with the presentation of two operation-word pairs which constituted one trial. After this trial was presented, a question mark appeared on the screen which signalled the subject to begin recall. The subject was to write down each word that had appeared after an operation in that particular trial. As the experiment progressed, the number of these operation-word pairs was gradually increased in the following manner. Three trials at a given level (for example 2 operation-word pairs) completed a set. When a set of three trials had been performed, the number of operation-word pairs in a trial was increased by one. The number of operation-word pairs in a trial was increased in this way until subjects had read, verified and attempted to recall the final words of a set of 5 operation-word trials. The subject stopped after completion of the set of 5 operation-word pairs. As in Experiment 1, subjects were allowed to write the words in any order they chose, but were asked to refrain from writing down the last word, first.

We measured each subject's oral reading rate while reading aloud the words from which the stimulus pools were formed. However, these data showed nothing meaningful and will not be discussed further.

Results

The main goal of this experiment was to discover whether a word length effect could be demonstrated in a complex span task that did not involve reading as a background component. As with Experiment 1, our confidence in the conclusions drawn from the results of this experiment would be enhanced if we could show that the scores on which the analyses were performed were internally reliable and that the tasks were measuring something important to a real world task like reading comprehension. Therefore three types of analyses were performed and will be discussed separately.

Analysis of variance. Figure 2 shows that, once again, people remembered more words in the simple word span task than they did in the complex operation task and, as anticipated, they also remembered more short words than long words. They remembered more short words than long words in both tasks, but this effect was greater in the simple task. The word length effect, defined as the difference between recall of short words and recall of long words, was bigger in the simple span than operations span task but was still highly reliable in the operations task.

These conclusions were supported by main effects of span task, $F(1, 72) = 37.5$, $MS_e = 107.3$, and word length, $F(1, 72) = 33.95$, $MS_e = 380.3$, and the span task by word length interaction $F(1, 72) = 15.64$, $MS_e = 107.3$. Further analysis of this interaction showed that while the word length effect was smaller in the operations span than the simple span, it was significant in both with the effect for simple word span giving an $F(1, 72) = 62.78$, $MS_e = 190.3$, and the effect for operation span giving an $F(1, 72) = 8.88$, $MS_e = 297.3$.

Reliability analysis. Once again we performed reliability analyses to determine the internal reliability of the span measures used in these analyses. The scores used in the reliability analyses were derived in exactly the same manner as in Experiment 1. The reliability analysis performed using Cronbach's alpha showed the following: Simple Span Short .78, Simple Span Long .84, Complex Span Short .65 and Complex Span Long .71. Thus, as with the previous experiment, the scores were relatively high and not much different from one another.

Correlational analysis. The correlation analysis performed in this experiment used the absolute scores for the span tasks and SAT scores. Tables 3 and 4 show the descriptive

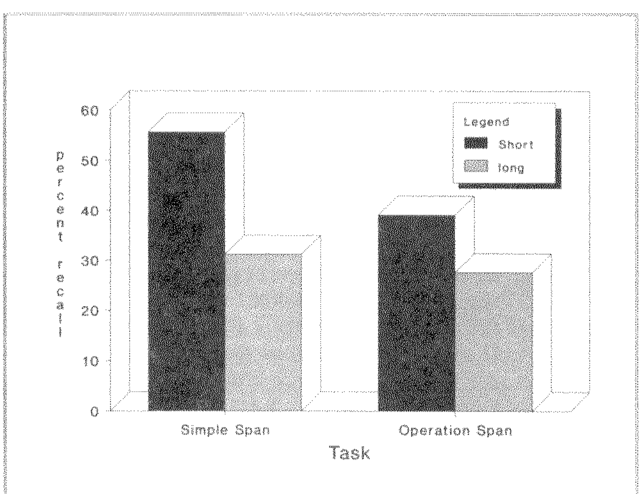

Figure 2. Experiment 2: Mean percent recall as a function of word length and task.

Table 3
Descriptive Statistics: Experiment 2

Measure	*N*	*M*	*SD*	Min.	Max.
Verbal SAT	76	440	77	290	661[a]
Quan SAT	76	476	85	330	699[a]
Total SAT	76	915	146	630	1360
Simple Short	37	55.4	14.4	19.8	91.3
Simple Long	39	31.0	13.3	11.1	75.3
Complex Short	37	39.1	18.5	14.3	100.0
Complex Long	39	27.8	16.3	0.0	78.6

[a] The scores for this subject are unusual in that they do not end in zero but are the scores reported by the Educational Testing Service. SAT = Scholastic Aptitude Test.

statistics and correlation coefficients that resulted from this analysis. Simple span measures using both long and short words significantly correlated with VSAT. The magnitude of the coefficient was somewhat greater for short words than it was for long words. As in the previous experiment, the recall from the complex span task (in this case the operations span task) significantly correlated with VSAT. This was true for operations spans with both short and long words. In addition to the demonstration of the word length effect with the operation-word span task, we have shown that performance on this task correlates with VSAT. Also, we have shown again that the simple word span task significantly predicts reading comprehension. As with the previous study, the highest correlation with VSAT of the four span tasks was the complex span with short words (0.54) but, in this study, the correlation between the simple span with short words and VSAT was nearly as high (0.49).

Discussion

Several conclusions can be drawn from these two experiments. One is that the background component of the complex span does not have to be reading for the complex span to predict comprehension. The reading span-comprehension correlation in Experiment 1 and the operation span-comprehension correlation in Experiment 2 were identical (0.54). Secondly, reading comprehension correlates with both simple and complex spans. The correlation with complex span may be somewhat larger, but not much, Thirdly, both of the complex span tasks studied here show a highly reliable effect of word length that may be somewhat smaller than that found with simple span tasks. However, the difference between the word length effect with complex and simple span tasks in both experiments is probably a result of scaling differences between the two tasks and not a theoretically interesting one. Both of the first two conclusions seriously question Daneman and Carpenter's (1980) interpretation of what is going on in the reading span task and why it correlates with comprehension.

We are not prepared at this point to discuss why both of our experiments showed a correlation between simple word span and comprehension and previous studies by Daneman and Carpenter (1980) and Turner and Engle (1989) did not find these correlations to be significant. Our lab is studying that question and the next three experiments reported here

Table 4
Correlations Between Absolute Scores and SAT

Absolute score	Verbal SAT	Quantitative SAT	Total SAT
Simple Short	.49**	.37*	.48**
Simple Long	.33*	.18	.28
Complex Short	.54**	.42*	.53**
Complex Long	.34*	.12	.25

Note. SAT = Scholastic Aptitude Test.
* $p < .05$. ** $p < .003$.

may cast some light on the issue. Suffice it to say that what, at one time, may have appeared to be trivial variations in procedures may in fact be quite important in whether the simple word span task measures something that is important in higher level cognitive tasks like reading comprehension.

The third conclusion suggests that there are some major similarities in the processes that are used in simple and complex word span tasks. The immediate reaction might be that the common process involved is articulatory coding or what Baddeley (1986) calls the articulatory loop. The problem with that interpretation is that, as mentioned above, the complex span has built-in articulatory suppression with the vocalizing of the sentences and the operations. Irrelevant vocalization has been shown repeatedly to eliminate the word length effect, at least with visual presentation (Baddeley, 1986). One possibility is that the vocalizing of the words in the simple and complex spans constitutes auditory presentation and, thus, the irrelevant vocalization in the complex tasks would not eliminate the word length effect.

As mentioned above, the work to date on the word length effect, and most other work using short list recall procedures, has used the same small sample of words on every trial. The simple and complex span procedures here used sampling without replacement from a large pool of words. It is possible that this leads to the use of a different code than that used with a small sample of words used repeatedly. Two papers by Besner and his colleagues (Besner & Davelaar, 1982; Besner, Davies, & Daniels, 1981) have argued that at least two different articulatory codes are used by subjects in immediate memory span experiments. One of these codes is used for lexical access from print and is not eliminated by articulatory suppression. The other code underlies both word length and phonological similarity effects in span tasks and is masked by irrelevant articulatory suppression.

One explanation for the results of our first two experiments is that there is yet a third articulatory or phonological code that comes into play when the same words are not used repeatedly. This code would aid retention in span experiments but would not be vulnerable to suppression. The next two experiments addressed this possibility.

Experiment 3

This experiment was designed to be much like the previous one except that subjects articulated an irrelevant letter string during the performance of both simple and operation span tasks. If the word length effect we have observed with both simple and complex tasks derives from mechanisms other than the articulatory loop, it might be immune to the effects of articulatory suppression and a word length effect like that observed in Experiments 1 and 2 would occur. If, on the other hand, subjects use a code like that described by Baddeley et al., (1975) then the word length effect should be eliminated for both tasks. Of course, another possibility is that we would see the effect eliminated with simple span and not with complex span, which would indicate that two different time-based articulatory or phonological codes are used for retention in the two tasks.

Method

Subjects. Sixty-four undergraduate students at the University of South Carolina, all native speakers of English, participated in this experiment as part of a course requirement. Subjects were tested on two operation span tasks and two simple word span tasks and their SAT scores were obtained from the school's admissions office. Each subject was seen individually and in the course of approximately 1.5 hours performed the four tasks.

Stimuli. The same sets of long and short words were used as in the previous experiments. The operations for this experiment were designed to be much simpler than those used in Experiment 2. Each required the multiplication or division of two integers between 1 and 5. The operations were presented in the form of a question as is shown by the following examples: Is $(5 \times 2) = 10$? Is $(1/1) = 0$. These operations were generated with an Apple IIe computer, using a program which imposed the following restraints: (a) the two numbers used in the multiplication or division were integers between 1 and 5, (b) determination of operation to be performed (i.e., multiplication or addition) was random, (c) answers to the operations were also presented to the subject for verification, (d) the answers which were presented were either correct or incorrect (by ±1 or ±2), and (e) choice of the answer that was presented was also random. A pool of 162 operations was generated in the above described manner and these were divided into two groups. One group was used with short word stimuli and one group with long.

Procedure. Word span task. The words appeared on the computer monitor at a rate of one per second and subjects attempted to recall them when a question mark signalling the end of a trial appeared in the center of the screen. Subjects were required to silently read the words and to continuously articulate the letter string "abcabc . . ." while the words appeared on the screen. The instructions for recall were identical to the previous experiments. Subjects performed the word span once with short words and once with long words. The wordset factor was fully crossed with task order, stimulus order and operations.

Operation span task. The pairing of operation to word was random without replacement and was determined by the program presenting the stimuli. The subject was instructed to begin silently reading the operation as soon as it appeared, to immediately indicate orally whether the operation was correct or not and then silently read the word at the end. As with the word span tasks, continuous articulation of the letter string "abcabc . . ." was required while the individual made the decision and read the operations and words silently. A verbal "yes" or "no" response regarding the accuracy of the answer presented for the operation, was required from the subject and the experimenter recorded this. Subjects were instructed to perform the operation verification as quickly as possible but to be accurate. Presentation was under the control of the experimenter. The appropriate way to perform the task was demonstrated and the subjects performed 15 practice trials. The experimenter immediately presented the next operation-word as soon as the "yes" or "no" response had been made for the previous string. The operations task began with the presentation of a series of two operation-word pairs which constituted one trial. After this trial had been presented, a question mark appeared on the monitor that signalled the subject to begin recall.

Design. A mixed design was used with four between subjects factors and two within subjects factors. Word group (2), order of stimulus presentation (2), order of task presentation (2), and operation set (2) were between subjects factors. Word length (2) and task (2) were within subjects factors.

Results

Analysis of variance. The primary question motivating this experiment was whether a word length effect would be demonstrated under conditions of articulatory suppression specifically in the complex task. Figure 3 clearly shows that subjects remembered more short words than long words in both the simple word span task and the complex operation span task. However, subjects performed no better in the simple span task than they did in the complex operation task. The word length effect was significant with both orders but the effect of word length was reduced when short stimuli came before long.

These conclusions were confirmed by a significant main effect of word length, $F(1, 48) = 98.90$, $MS_e = 40.9$. One two-way interaction was significant, Order of Word Length × Word Length, $F(1, 48) = 15.76$, $MS_e = 40.9$. Also a three-way interaction was significant, Wordgroup × Order of Word Length × Word Length $F(3, 48) = 2.92$, $MS_e = 40.9$.

Correlations. Tables 5 and 6 show the descriptive statistics and important correlations for this study. Apparently, the requirement to suppress eliminated the correlation between simple word span and VSAT. With the operation span, the correlation was somewhat reduced but was, nevertheless, significant. It may be a coincidence, but the magnitude of these correlations, for both simple and complex tasks, was very similar to the correlations reported by Turner and Engle

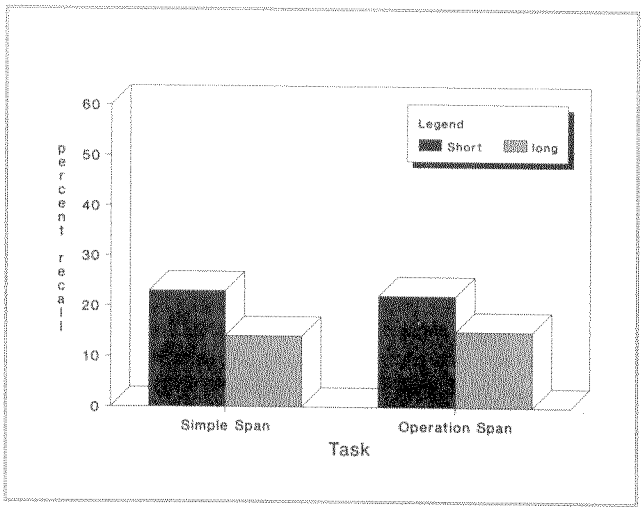

Figure 3. Experiment 3: Mean percent recall as a function of word length and task.

Table 5
Descriptive Statistics: Experiment 3

Measure	*N*	*M*	*SD*	Min.	Max.
VSAT	64	447	83	290	700
QSAT	64	485	79	300	620
TSAT	64	932	141	670	1310
Simple Short	64	22.9	9.5	6.2	59.3
Simple Long	64	14.1	6.2	4.9	45.7
Complex Short	64	22.1	13.2	4.8	66.7
Complex Long	64	15.0	9.0	0	45

Note. VSAT = Verbal Scholastic Aptitude Test; QSAT = Quantitative Scholastic Aptitude Test; TSAT = Total Scholastic Aptitude Test.

(1989). It is possible that the net effect of the procedure used by Turner and Engle resulted in articulatory suppression.

Discussion

If we compare the overall level of performance in this experiment to that in Experiment 2, we see that the requirement to articulate an irrelevant letter string drove overall performance down for both tasks. However, performance was apparently reduced more in the simple span condition than in the complex condition because they were about equal in this study. Despite this overall decrement in recall, however, the word length effect remained for both tasks, even though the words and operations were not presented auditorily nor read aloud by the subject.

The primary purpose of this study was not to examine changes in the correlations between the spans and comprehension. But we can not overlook that the significant correlations between simple word span and VSAT observed in the two previous studies disappeared when articulatory suppression was required, while those between operation word span and VSAT remained. It is difficult to know whether the disappearance of the correlation between simple span and comprehension is a psychometric artifact or is theoretically interesting. The suppression requirement substantially reduced performance in the simple span and this may have eliminated any possibility for covariation between this task and the VSAT to appear. But, if that was the case, why did the correlation still remain for the complex task which had even lower levels of performance? This question merits further study.

Numerous studies since Baddeley et al. (1975) have shown that the word length effect in a simple span task with visually presented words is eliminated by articulatory suppression (Baddeley, 1986). Why was that not the case in Experiment 3? One possibility is that with different words on every list, a deeper or different kind of code is used for the words. A direct comparison of the word length effect with fixed and unlimited pools of words was clearly in order and the next study did just that.

Experiment 4

We were interested in the presence or absence of the word length effect in eight different conditions. Subjects received both simple word span and operation span tasks. The to-be-

Table 6
Correlations Between the Simple Span and Complex Span Tasks and VSAT as a Function of Word Length

	Simple Span		Complex Span	
Test	Short	Long	Short	Long
VSAT	.13	.09	.26*	.35*

Note. $n = 64$. VSAT = Verbal Scholastic Aptitude Test.
* $p < .05$.

remembered words were short or long and were chosen from either a fixed pool or an unlimited pool. Half the subjects performed irrelevant articulation while simultaneously performing the span tasks and half the subjects did not articulate.

Method

Subjects. Sixty-four undergraduate students at the University of South Carolina, all native speakers of English, participated in this experiment as part of a course requirement. Each subject was seen individually and in the course of approximately 1.5 hours performed four tasks: simple and operation word spans with short and long words.

Stimuli. Words for the span tasks were selected from the same words used in Experiment 1. One of the variables of primary interest in this experiment was the size and familiarity of the stimulus pool. For this reason a subset of eight words from each of the Set A and Set B short and long words was randomly chosen. The fixed pool of stimuli drawn from Set A short words consisted of the following: *snake, sum, arm, talk, fact, learn, hall,* and *need.* Long words selected from Set A were *addition, company, foundation, instruction, annual, sensitive, excellent,* and *physical.* The fixed pool of Set B short words was: *seat, knee, laugh, test, act, climb, all,* and *ear.* The words chosen from Set B long words were: *theory, atmosphere, specific, emphasis, maintenance, determine, occasion,* and *employee.*

As in Experiment 3, subjects completed two types of tasks, operation span tasks and simple word span tasks, both with short and long words. The operation task required that individuals read a series of mathematical operations that were presented in the form of a question and to verify whether or not the answer presented with it was correct. These operations were the same operations used in Experiment 3.

Procedure. Word span task. Word stimuli were presented on the monitor of an Apple IIe computer. The words appeared at a rate of one per second and subjects read them silently and attempted to recall them when a question mark signalling the end of a trial appeared in the center of the screen. Subjects who performed the suppression were required to continuously articulate the letter string "abcabc . . ." while the words appeared on the screen, whereas those in the no suppression condition said nothing while the words appeared. The list length began at two words. Three trials at a given list length constituted a set and when a set had been completed the number of words in a trial was increased by one. In this task, subjects stopped when a set of seven words had been completed. Since this study used a fixed pool of items in some conditions, serial recall was required with responses written on the response sheet beginning with the first word. Subjects were instructed to place asterisks in the blanks on the response sheet when they could not recall the appropriate word. After attempting to write all the words presented in a trial in correct serial order, subjects often remembered words which had not been recalled immediately and which they had represented by asterisks. They were allowed to write these words by the appropriate asterisks. All subjects performed the word span once with short words and once with long words. The word group factor was fully crossed with task order and stimulus order.

Operation span task. The mathematical operations described in Experiment 3 were used in the operation span task. Each trial consisted of a series of operations appearing on the monitor of the computer. A word from Set A or Set B was presented at the end of each operation. The pairing of operation to word was random without replacement and was determined by the program presenting the stimuli. The subject was instructed to begin silently reading the operation as soon as it appeared, to immediately indicate orally whether the operation was correct or not and then silently read the word at the end. As with the word span tasks, subjects in the articulatory suppression condition continuously articulated the letter string "abcabc . . ." while making the decision and reading the operations and words silently. In the no-suppression conditions subjects said nothing while reading the operation and verifying it, but verbally responded to the accuracy of the presented answer with yes or no and the experimenter recorded these responses in both suppression and no suppression conditions. Subjects were instructed to perform the operation verification as quickly as possible but to be accurate. Presentation was under the control of the experimenter who had demonstrated the appropriate way to perform the task and also presented 15 practice trials to assure that the subject understood the procedure. During these practice trials, the experimenter noted whether the subject was responding accurately on the verification component and provided feedback if necessary. The experimenter immediately presented the next operation as soon as the yes or no response had been made. The operations task began with the presentation of a series of two operation-word pairs that constituted one trial. After this trial had been presented, a question mark appeared on the monitor that signalled the subject to begin recall. The subject was to write down each word that had appeared after an operation in that particular trial in serial order as described in the word span task. Three trials at a given level (for example 2 operation-word pairs) constituted a set. When a set of three trials had been completed, the number of operation-word pairs in a trial was increased by one. The number of operation-word pairs was increased in this way until subjects had read, verified and attempted to recall the final words of a set of 5 operation-word trials. The subject stopped after completion of the 5 operation-word set.

All subjects were presented 15 practice trials for both the simple word span and operation span tasks. Additionally, those who participated in the fixed-pool condition were required to learn the pool before each span task began. This was accomplished by presenting six trials of the simple word span task in which all eight words of the pool were presented for serial recall. In each of these trials the order was varied and the words comprising the pool were in full view. If subjects were not able to write all eight words they were encouraged to guess.

Each subject preformed two operation span and two simple word span tasks with both short and long words for each task. The order of task presentation was counterbalanced and crossed with word group. Order of word length presentation was also counterbalanced. One set of operations was used with the first complex task performed and the other set was used with the second complex task.

Design. A mixed design was used with five between subjects factors and two within subjects factors. Suppression, fixed or unlimited word pool, word group, order of stimulus presentation, and order of task presentation were between subjects factors. Word length and task were within subjects factors.

Results

As with the previous studies, all the analyses were performed on all the scoring procedures and no differences were observed. Because there were only 16 subjects per condition in this experiment, no correlational analysis was done. Unlike Experiments 1–3, this study was simply concerned with whether there would be a difference between recall of short and long words in each of the eight conditions. Therefore, planned individual *t* tests were used to test for a word length effect in the eight conditions.

The results are shown in Figure 4 where the absolute span scores have been converted to percent correct to allow comparison across tasks. An asterisk over a group of two bars indicates a significant word length effect for that condition as indicated by planned dependent *t* tests at the 0.05 level.

For the fixed pool conditions, the word length effect occurred in both the simple span, $t_{15} = 2.44$, $SE = 5.1$, and operation span, $t_{15} = 2.14$, $SE = 3.6$, if there was no suppression. Articulatory suppression eliminated the word length effect in both the simple span, $t_{15} = -.42$, $SE = 2.9$, and operation span, $t_{15} = -.25$, $SE = 4.9$.

For the unlimited pool conditions, the word length effect was significant in all conditions. There was better recall of short than long words in the simple span with no suppression, $t_{15} = 8.18$, $SE = 3.2$, simple span with suppression, $t_{15} = 3.47$, $SE = 2.6$, operation span with no suppression, $t_{15} = 4.04$, $SE = 4.2$, and operation with suppression, $t_{15} = 2.47$, $SE = 2.7$.

Discussion

Most of what was found in this experiment simply replicated previous findings. The first two comparisons (1 and 2), reading from the left of Figure 4, show what has repeatedly been shown by Baddeley and his colleagues. There was a substantial word length effect when no articulatory suppression was required and the same words were used on each trial. When a suppression requirement was imposed, however, the word length effect disappeared if the words were chosen from a fixed pool. The next two comparisons (three and four) show that the same conclusions hold for the complex operation span when the words are chosen from a small fixed pool. The effect of word length was significant when there was no suppression but was eliminated when suppression was required.

The other four comparisons essentially replicated what we have already shown in Experiments 1–3 with an unlimited word pool. The word length effect was present with both simple and complex span tasks and was not eliminated with suppression.

The pattern of results across the four experiments reported here were quite consistent. However, these is a potential confound in these experiments. Inspection of the list of words used in all the experiments showed that more of the short words were concrete and more of the long words were abstract. Therefore, we performed an additional experiment to ensure that the conclusions reached above were not the result of selection of words for the span test.

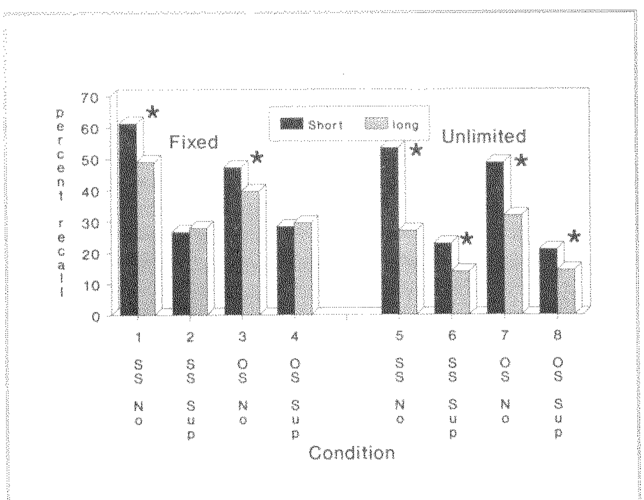

Figure 4. Experiment 4: Mean percent recall as a function of word length, task, suppression condition and nature of the word pool. (The asterisk indicates a significant effect of word length for the condition. Condition 1 is Simple Span-No Suppression, Fixed Pool; 2 is Simple Span-Suppression, Fixed Pool; 3 is Operation Span-No Suppression, Fixed Pool; 4 is Operation Span-Suppression, Fixed Pool; 5 is Simple Span-Suppression, Unlimited Pool; 6 is Simple Span-Suppression, Unlimited Pool; 7 is Operation Span-No Suppression, Unlimited Pool; and 8 is Operation Span-Suppression, Unlimited Pool.)

Experiment 5

The purpose of this experiment was to see whether the conclusions regarding word length effect, articulatory suppression and nature of the word pools remain valid when the short and long words are equated for concreteness. This opportunity was also used to see whether the conclusions were safe for both free and serial recall.

Method

Subjects. One hundred forty four undergraduate students at the University of South Carolina, all native speakers of English, participated in this experiment as part of a course requirement.

Design and procedure. The simple span task was used for this experiment. There were 12 separate conditions. The words were either sampled with replacement from a fixed pool of 8 words or sampled without replacement from a pool of 81 words. For those subjects receiving words from the fixed pool, recall was, of course, serial. For those subjects receiving words from the unlimited pool, half the subjects were instructed to recall in a strict serial order and half the subjects were instructed to recall in any order with the constraint that they were not to recall the last word first. Half the subjects in each of these three conditions performed under articulatory suppression and half performed normally, without the requirement to articulate the irrelevant items. The resulting six conditions were crossed with the length of the to-be-recalled words. Half the subjects received short words, and half the subjects received long words.

A subset of each set of 81 words was chosen to be used as the fixed pool items. The result was a list of short words and a list of long words where word frequency and concreteness were equated as well as possible. These words were used in the conditions that received items sampled from the same small set for each trial. The same procedure was used as for the simple span for the other experiments. However, all variables were manipulated between subjects.

Stimuli. As in the prior experiments, two sets of words were used as to-be-recalled items, one set of one syllable words and another of words 3 and 4 syllables long. The words were chosen from Paivio, Yuille, and Madigan (1968) to equate the two sets for concreteness. It was not possible to totally equate the two sets for both concreteness and frequency. The mean concreteness rating for the short words was 6.52 and that for the long words was 6.53. For the short words, there were 20 AA words and 15 A words and the rest of the list had a Thorndike-Lorge rating of 20.72. Among the set of long words were 20 AA and A words and the rest had a mean frequency of 13.

Results

As with Experiment 4, we were primarily interested in whether there was an effect of word length for each of the six conditions. Therefore, the critical predictions were tested with planned *t* tests at the 0.05 level. The results are graphed in Figure 5. The results demonstrate that a confounding of concreteness with long and short words was not responsible for the conclusions of the previous 4 experiments nor was the method of recall. The results were the same when concreteness was controlled and regardless of whether recall was free or serial.

Articulatory suppression eliminated the word length effect when the same words are used on each trial. However, when new words are used on each trial, articulatory suppression does not eliminate the word length effect, regardless of whether the subject recalls by free or serial recall. The only word length comparison in Figure 5 that was not significant was with articulatory suppression and lists from a fixed pool of words, $t_{22} = .79$, $SE = 4.24$. The word length effect was significant for the other five conditions: (1) fixed pool with no suppression, $t_{22} = 4.87$, $SE = 5.01$, (2) unlimited pool with serial recall and no suppression, $t_{22} = 3.52$, $SE = 3.69$, (3) unlimited pool with serial recall and suppression, $t_{22} = 2.06$, $SE = 3.07$, (4) unlimited pool with free recall and no suppression, $t_{22} = 4.12$, $SE = 3.54$, and (5) unlimited pool with free recall and suppression, $t_{22} = 3.97$, $SE = 2.27$.

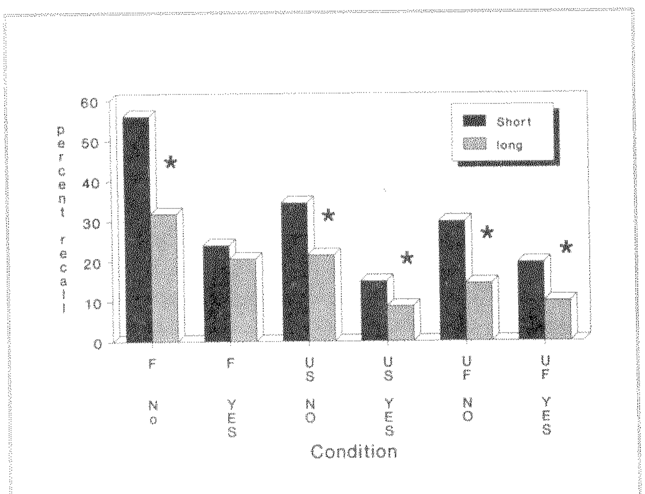

Figure 5. Experiment 5: Mean percent recall as a function of word length, word pool and recall condition and suppression condition. (The asterisk indicates a significant effect of word length for the condition. The condition labels reflect the following: [a] F-NO, Fixed Pool, No suppression; [b] F-YES, Fixed Pool, Suppression; [c] US-NO, Unlimited Pool, Serial Recall, No suppression; [d] US-YES, Unlimited Pool, Serial Recall, Suppression; [e] UF-NO, Unlimited Pool, Free Recall, No Suppression; and [f] UF-YES, Unlimited Pool, Free Recall, Suppression.)

Discussion

In both Experiments 4 and 5, we have shown that the word length effect was not eliminated with articulatory suppression. Another point should be noted from Figures 4 and 5. Baddeley and his colleagues (Baddeley, 1986) have repeatedly shown that articulatory suppression eliminates the word length effect. We have observed the same phenomenon in our studies if a fixed pool of items is used. One potential criticism of these studies is that the absence of the word length effect with suppression occurs as a consequence of floor effects. In other words, it is possible that suppression caused both long and short words to be recalled so poorly that no effect could be observed. However, in both Experiments 4 and 5, superior recall of short words to long words was observed with even lower levels of recall *if* the words were selected from an unlimited pool. In fact, in Experiment 5, the effect was observed with recall at about 10% in the serial recall of words from an unlimited pool and with articulatory suppression. This speaks to the robustness of: (a) the absence of the effect when suppression occurs with lists from fixed pools, and (b) the presence of the effect with lists from unlimited pools.

General Discussion

Our purpose in this research was not, at least initially, to study the word length effect per se. Our purpose was, however, to understand better the relationship between the simple word spans, complex word span and comprehension. This question was examined in two ways. One strategy was to see whether simple spans correlated with reading comprehension under conditions in which the presentation modality was more similar to that used with the reading span task and also under conditions more like those typically used in short-term memory experiments with words. A second strategy was to see whether both types of span tasks were affected by word length in the same manner.

Let us briefly summarize the findings of the five experiments reported here. The first experiment found a word length effect in the complex reading span task and also that the simple word span task significantly correlated with reading comprehension. The second experiment showed that the word length effect was obtained in a complex operation span task and, again showed that the simple span task correlated with reading comprehension. The third experiment showed that suppression reduced the correlation between the simple span and comprehension to non-significant levels, but the correlation between the complex span and comprehension remained significant. It was further observed that the word length effect was not eliminated by articulatory suppression in either the simple span or operation span task. The fourth experiment found that, with an unlimited pool of words, the word length effect was *not* eliminated by suppression, but with fixed pools of words it was eliminated by suppression. This generalization held for both simple and complex span tasks. The fifth experiment eliminated the hypothesis that our results were due to a confounding with concreteness of the words and the method of recall.

One interpretation that seems to leap out from these results is that the reading span task does not have the special qualities attributed to it by Daneman and Carpenter (1980). We found that reading comprehension was predicted as well by a complex span task involving arithmetic as by one involving reading. More importantly, the simple word span task also significantly predicted comprehension and, in some cases, did so as well as did the complex span task. Our lab has conducted numerous studies since Turner and Engle (1989) in which the simple span task has been compared to several forms of the complex span task. The simple span has significantly predicted reading comprehension in nearly all of them. Typically, the magnitude of the correlation between word span and VSAT is slightly smaller than that between the complex span and VSAT and, over studies, it is more variable, but it is nearly always significant. We do not yet know the boundary conditions under which the simple span-comprehension correlation will appear and when it will not appear. Our third experiment may give a hint to that question. As mentioned earlier, it may just be a coincidence, but the correlations for both simple and operation span in Experiment 3 were virtually identical to those obtained by Turner and Engle (1989). It is not clear why suppression would eliminate the correlation for simple span but not complex span and whether this has any theoretical implications.

Additional evidence for the argument that the reading span task does not have a special relationship with reading comprehension was the finding that the complex span task acted just like the simple span task as word length was manipulated in our experiments. In conditions where the word length effect appeared with the simple span task it also appeared with both of the complex span tasks. In conditions where the word length effect disappeared with the simple span, it also disappeared with the complex span. Clearly, this one variable has similar effects on the simple word span, reading span, and operation span tasks.

The complex and simple span tasks may not be greatly different in what they measure, but it is unclear what they do measure that is important to higher level cognition. One factor that both span tasks and reading comprehension tasks have in common is a reliance on verbal knowledge. However, we have shown in a recent study (Engle, Nations, & Cantor, in press) that word knowledge per se does not account for the relationship between word span and comprehension performance. It is still possible that this relationship is peculiar to verbal tasks, however, and some form of articulatory coding may be the critical factor.

It was not our intention to study the nature of the word length effect when this series of studies began but some rethinking of this variable seems to be in order. It has been argued that the articulatory loop is the process underlying the word length effect (Baddeley et al., 1975). But what is meant by the term "articulatory loop"? It is often referred to in the literature as if it were an inherent structural aspect of human cognition. But perhaps it would be more beneficial to consider it a coding strategy, one of many, which may be invoked during thinking.

This view was proposed by Reisberg, Rappaport, and O'Shaughnessy (1984). They pointed out that, although working memory is often thought to be comprised of a central executive and slave systems that act as information stores, actually "these slave systems are in fact activity-based strategic control processes and not memories in a strict sense" (p. 203). They showed that people could increase recall by learning a new coding strategy. Subjects were taught to represent each of the numbers between 1 and 10 by one of their fingers, thus each number was in essence recoded as a finger. They were also taught to use movement of the fingers as a rehearsal process. When presented a series of numbers to remember, subjects moved their fingers as if to type the numbers in the series on an imaginary typewriter. By continually repeating the motor sequence necessary to type the numbers, they could maintain them until recall, at which time they simply decoded the series of finger movements. Reisberg, et al. noted, that although it is widely accepted that phonological coding occurs in working memory, it is not necessarily a fixed attribute of working memory. Perhaps it appears that way because so much of what we do can be accomplished economically using articulatory-type strategies or processes which, because they are used so frequently, become very well-practiced and possibly automatic.

Reisberg et al. further suggested that people code information in working memory in whatever configuration that is useful to accomplish the task at hand, limited only by individual capabilities and personal history. For example, some researchers now believe that very young children rely on visual coding whereas older children appear to use articulatory or phonological coding more readily (Hitch, Halliday, Scaafstal, & Schraggen, 1988; Hitch, Woodin, & Baker, 1989).

The articulatory loop then may be thought of as the combination of a code that decays over time and a time-limited rehearsal process or strategy which, when invoked, refreshes that code. Thus, it is a simple race between the process that refreshes the code and decay of the trace. Each time the code is refreshed it is, in essence, recoded. If we accept this view of the articulatory loop, we can further consider the possibility of a whole continuum of codes, of which phonological and articulatory are but two examples. It is also important to note that the representation of information in a specific code would not preclude its representation in other coded forms as well. This variety of the levels of processing approach is not a new idea (Craik & Lockhart, 1972) and it offers a less static way of thinking about coding in working memory. Information may be extracted and recoded in a variety of ways.

In fact, it is conceivable that there are many different types and levels of speech-based codes. Besner and his colleagues have offered evidence for two different phonological codes in immediate memory tasks. To the extent that the word length effect reflects time-based phonological codes, our research suggests that additional codes come into play when the words for recall are chosen from an unlimited pool. It remains to be seen just how important the nature of the word pool is to the relationship between the span tasks and reading comprehension.

If simple span and complex spans are not greatly different, then what is the relationship between what is now called working memory and what used to be called short-term memory? We would like to argue that a distinction needs to be made between short-term memory and working memory; a distinction made by Bower (1975), Anderson (1983) and others. It is possible to define short-term memory as those knowledge structures currently active above some threshold which could be considered as the boundary of consciousness. Traditional span tasks probably overestimate short-term memory capacity which is more likely to be on the order of 3–4 items rather than 7 ± 2 (Broadbent, 1971). Processes like rehearsal and grouping allow subjects to keep active and recall approximately 2–4 additional items, but these same processes are rarely used in higher level cognitive tasks like reading or listening comprehension. Thus, individual differences in rehearsal, grouping, and associational type strategies just obscure any relationship between short-term memory capacity and higher level cognitive tasks.

Working memory is probably better thought of as those long-term memory knowledge structures and the connections between them that, while recently activated are, nevertheless, outside the window of consciousness. It is the working memory that allows us to retain the gist of the conversation in which we are engaged, to quickly retrieve the name of a new neighbor we met recently and to know roughly where we left our car keys last night. Our working memories provide the texture and context to our cognitive life at any given moment.

How does the complex span fit with this distinction between short-term memory and working memory? We would like to argue that the complex memory span task is really a measure of short-term memory, not the working memory described above. The complex span is probably a better measure of the true capacity of short-term memory because the background task minimizes the impact of rehearsal and grouping on recall. However, even the simple word span might be useful in this regard because chunking, grouping, and simple rehearsal may be harder and less likely to occur when the list does not use the same words on each trial. This, of course, suggests a much more general capacity notion than one based on phonological or verbal codes.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Baddeley, A. D. (1986). *Working Memory.* Oxford psychology series No. 11 Oxford: Clarendon Press.

Baddeley, A. D., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24,* 119–131.

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,* 585–589.

Bower, G. H. (1975). Cognitive psychology: An introduction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes (Vol 1).* Hillsdale, NJ: Erlbaum.

Besner, D., & Davelaar, E. (1982). Basic processes in reading: Two phonological codes. *Canadian Journal of Psychology, 36,* 701–711.

Besner, D., Davies, J., & Daniels, S. (1981). Phonological processes in reading: The effects of concurrent articulation. *Quarterly Journal of Experimental Psychology, 33,* 415–438.

Broadbent, D. E. (1971). The magic number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory.* New York: Wiley.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Carpenter, P. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 561–583.

Dempster, F. N. (1985). Short-term memory development in childhood and adolescence. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development: Progress in cognitive development research.* New York: Springer Verlag.

Dixon, P., LeFevre, J., & Twilley, L. C. (1989). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80,* 465–472.

Engle, R. W., Nations, J. K., & Cantor, J. (in press). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology.*

Guyer, B. L., & Friedman, M. P. (1975). Hemispheric processing and cognitive styles in learning-disabled and normal children. *Child Development, 46,* 658–668.

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory and Cognition, 16,* 120–132.

Hitch, G. J., Woodin, M. E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory and Cognition, 17,* 175–185.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329–354.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

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.

Paivio, A., Yuille, J. C., and Madigan, S. A. (1968). Concreteness, imagery and meaningfulness. *Journal of Experimental Psychology, 76,* 1–25.

Perfetti, C. A., & Lesgold, A. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension.* New York: Wiley.

Reisberg, D., Rappaport, I., & O'Shaughnessy, M. (1984). Limits of working memory: The digit digit span. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 203–221.

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

Appendix

List A: Short Words

add
aid
arm
back
beach
beat
bible
bird
blue
bomb
brain
break
buy
camp
care
cause
close
curve
cut
dance
desk
door
dream
dress
dust
east
fact
far
fight
file
form
forth
gas
green
guest
guy
hall
hard
head
heat
help
jump
key
king
knife
league
lock
meal
miss
moon
mouth
near
need
nose
out
own
pair
rain
rock
roll
scene
score
sea
send
set
shall
skill
snake
stay
sum
talk
tall
taste
tool
town
trade
tree
truck
type
wait
wire

List A: Long Words

addition
American
annual
appearance
arrangement
assignment
assistance
attitude
audience
authority
automobile
available
average
business
capacity
century
community
company
competition
completion
condition
confidence
consider
construction
decision
democratic
difference
discussion
distinction
distinguish
education
electric
emotional
encourage
especially
establish
example
excellent
existence
foundation
generally
impossible
instruction
liberal
location
medical
medium
memory
minister
national
newspaper
observation
obviously
orchestra
particle
particular
philosophy
physical
poetry
popular
position
preparation
probably
properly
radiation
rapidly
reality
realize
reference
resolution
scientific
sensitive
significance
similar
society
soviet
telephone
typical
understanding
universe
variety

List B: Short Words

act
all
bad
base
bay
birth
block
bond
box
brown
build
ear
end
fair
far
farm
fine
fit
five
fly
foam
force
list
meet
mix
no
nod
none
paint
pale
park
plan
post
spot
test
theme
train
tube
vote
week
why
wife
win
work

Appendix *continued*

chance
check
claim
climb
count
cross
crowd
cry
deep
doubt
draw
dry
hang
hat
home
jazz
judge
knee
lack
laugh
lean
lie
lift
lip
rise
roof
run
seat
share
shoe
shore
site
soul
south
speak
sport
worth

List B: Long Words

adequate
advantage
anything
assembly
atmosphere
avenue
capital
character
chemical
collection
communist
component
conclusion
continue
corporation
deliver
description
detective
determine
dictionary
discovery
economic
element
eliminate
emphasis
employee
equipment
estimate
everyone
everything
exactly
examine
expenditure
expression
family
formula
government
headquarters
historical
hospital
important
independent
information
initial
instrument
interest
interesting
literature
maintenance
measurement
musician
necessary
objective
occasion
opinion
otherwise
period
phenomenon
policy
political
practical
president
previously
professional
provision
quality
reasonable
religion
remember
satisfy
secretary
security
solution
somebody
specific
suddenly
sufficient
temperature
theory
tragedy
yesterday

Short Word List—Experiment 5

Unlimited Pool

bard
code
brute
dell
air
square
bloom
dawn
chief
breeze
gift
golf
sky
noose
stub
lump
spray
stain
cash
boss
slave
lice
queen
steam
crag
storm
beast
wife
camp
fowl
lad
bronze
gem
coast
church
rod
street
brain
cell
slush
dirt
dust
jail
lime
fur
vest
star
suds
nun
sauce
gold
sea
seat
inn
peach
tool
blood
lark
bar
girl
keg
tank
ink
moss
board
child
soil
claw
dove
oats
bowl
coin
corn
pipe
string
cane
toast
cord
limb
pole
boy

Appendix *continued*

chance
check
claim
climb
count
cross
crowd
cry
deep
doubt
draw
dry
hang
hat
home
jazz
judge
knee
lack
laugh
lean
lie
lift
lip
rise
roof
run
seat
share
shoe
shore
site
soul
south
speak
sport
worth

List B: Long Words

adequate
advantage
anything
assembly
atmosphere
avenue
capital
character
chemical
collection
communist
component
conclusion
continue
corporation
deliver
description
detective
determine
dictionary
discovery
economic
element
eliminate
emphasis
employee
equipment
estimate
everyone
everything
exactly
examine
expenditure
expression
family
formula
government
headquarters
historical
hospital
important
independent
information
initial
instrument
interest
interesting
literature
maintenance
measurement
musician
necessary
objective
occasion
opinion
otherwise
period
phenomenon
policy
political
practical
president
previously
professional
provision
quality
reasonable
religion
remember
satisfy
secretary
security
solution
somebody
specific
suddenly
sufficient
temperature
theory
tragedy
yesterday

Short Word List—Experiment 5

Unlimited Pool

bard
code
brute
dell
air
square
bloom
dawn
chief
breeze
gift
golf
sky
noose
stub
lump
spray
stain
cash
boss
slave
lice
queen
steam
crag
storm
beast
wife
camp
fowl
lad
bronze
gem
coast
church
rod
street
brain
cell
slush
dirt
dust
jail
lime
fur
vest
star
suds
nun
sauce
gold
sea
seat
inn
peach
tool
blood
lark
bar
girl
keg
tank
ink
moss
board
child
soil
claw
dove
oats
bowl
coin
corn
pipe
string
cane
toast
cord
limb
pole
boy

Appendix *continued*

Fixed Pool

beast	star	inn	bronze
wife	lump	dirt	golf

Long Word List—Experiment 5

Unlimited Pool

colony	nursery	restaurant	malaria
retailer	prisoner	furniture	decoration
procession	avenue	kerosene	vaccination
headquarters	gallery	alcohol	prosecutor
beverage	hurricane	tobacco	evangelist
property	professor	factory	inhabitant
performer	musician	library	material
wholesaler	pianist	tablespoon	ambassador
instrument	orchestra	lemonade	bacteria
islander	photograph	butterfly	infirmary
passageway	newspaper	diamond	proprietor
officer	cobblestone	mosquito	amplifier
admiral	utensil	revolver	vegetable
committee	physician	daffodil	alligator
acrobat	grandmother	ambulance	caterpillar
gentlemen	policeman	microscope	accordion
opium	medallion	umbrella	macaroni
appliance	animal	strawberry	automobile
instructor	hospital	elephant	university
vehicle	magazine	potato	refrigerator
	volcano		

Fixed Pool

musician	orchestra	officer	alcohol
gallery	property	mosquito	amplifier

Received June 7, 1989
Revision received March 26, 1990
Accepted March 27, 1990 ■