JOURNAL OF MEMORY AND LANGUAGE 28, 127–154 (1989)

Is Working Memory Capacity Task Dependent?

MARILYN L. TURNER

Wichita State University

AND

RANDALL W. ENGLE

University of South Carolina

The complex span measure of working memory is a word/digit span measured while performing a secondary task. Two experiments investigated whether correlations between the complex span and reading comprehension depend on the nature of the secondary task and individual skill in that task. The secondary task did not have to be reading related for the span to predict reading comprehension. An arithmetic-related secondary task led to correlations with reading comprehension similar to those found when the secondary task was reading. The relationship remained significant when quantitative skills were factored out of the complex span/comprehension correlations. Simple digit and word spans (measured without a background task) did not correlate with reading comprehension and SAT scores. The second experiment showed that the complex span/comprehension correlations were a function of the difficulty of the background task. When the difficulty level of the reading-related or arithmetic-related background tasks was moderate, the span/comprehension correlations were higher in magnitude than when the background tasks were very simple, or, were very difficult.

WORKING MEMORY CAPACITY: AN INDIVIDUAL DIFFERENCES APPROACH

A hallmark of models of information processing since the early work of Shannon and Weaver (1949) has been the focus on capacity limitation. Most models attribute this limitation to some combination of short-term memory and attention (Broadbent, 1958). Presumably, the limited capacity of short-term memory affects our performance in cognitive tasks like reading or listening comprehension and general problem solving (Kintsch & van Dijk, 1978; Newell & Simon, 1972).

These studies were part of a dissertation submitted to the University of South Carolina by Marilyn L. Turner. The research was supported by AFOSR Grant 870069 to Randall W. Engle. Reprints may be obtained from Marilyn L. Turner, Psychology, Wichita State University, Wichita, KS, 67218, or Randall W. Engle, Psychology, University of South Carolina, Columbia, SC, 29208.

The digit span has been assumed to reflect output from short-term memory and is a ubiquitous component of intelligence tests (Wechsler, 1944). However, it does not correlate well with performance on such higher level tasks as reading comprehension (Perfetti & Lesgold, 1977) or even the amount of information estimated to be represented in primary or secondary memory (Martin, 1978).

Early theories, e.g., Waugh & Norman (1965), viewed short-term memory as a fixed number of slots or bins. Baddeley & Hitch (1974) argued that this focused too much on the storage functions of short-term memory and not enough on the processing functions. They preferred the name working memory and argued for the importance of both storage and processing in a functional analysis of the working memory system. Baddeley and Hitch proposed three structural components through which information is processed: (1) a central execu-

0749-596X/89 $3.00

tive, (2) an articulatory loop, and (3) a visuo-spatial scratch pad. The articulatory loop and the visuo-spatial scratch pad are maintenance systems controlled by the central executive, which is a flexible work space with limited capacity. Part of this limited capacity is used for processing incoming information with the remainder used for storage of the products resulting from that processing.

Although other models of WM have been developed (e.g., Klapp, Marshburn, & Lester, 1983; Brainerd & Kingma, 1984, 1985; Case, 1974; Kintsch & van Dijk, 1978), they all assume a limitation in the amount of information that can be kept active at any given time. Further, it has been generally assumed that this limitation affects consequent processing, i.e., that higher level processing is limited to some extent by the limitations of WM.

Daneman and Carpenter (1980, 1983) demonstrated the importance of the WM limitation in reading comprehension. They hypothesized that WM is used to represent the strategies and skills used in reading with any remaining WM capacity used to store the products of reading comprehension, e.g., facts, pronoun referents, and propositions. The two functions, the processing of the written information and the storage of intermediate products, were thought to compete for the limited resources available to WM.

They suggested that individual differences in reading comprehension could be due to variability between readers in the efficiency of their processing skills. Presumably good and poor readers have equivalent overall working memory capacities. Good readers are assumed to have efficient reading skills which demand relatively little from the gross WM resources leaving more of the WM capacity for the storage of products of the reading task. Consequently, good readers have more WM capacity available for the storage of products simply because they have more efficient reading skills. Since poor readers are assumed to have inefficient and thus capacity demanding reading skills, they have less residual WM capacity for storing the products of reading.

Accordingly, individual measures reflecting WM capacity should be tied to a specific processing task, in this case reading. To restate this idea, good readers would have more available WM capacity than poor readers while reading because of their more efficient reading skills. But this greater WM capacity would be specific to reading tasks. That is, a good reader could have LESS WM available when performing a nonreading task than a poor reader who is skilled at the nonreading task.

Daneman and Carpenter (1980) developed a measure of WM that measured capacity while the subject was performing a nominal reading task. This task, which they called the reading span task, was presumed to require both processing and storage in WM. Subjects read groups of sentences aloud while simultaneously trying to remember the last word of each sentence. The number of sentences in a group gradually increased. After each group, the subject recalled the "endwords" in any order. WM capacity was defined as the largest group of endwords the subjects recalled correctly. The task is, in effect, a word span task since the only measurement made is the number of words recalled correctly. It was hypothesized that the effect of the secondary task (reading the sentences aloud) would differ for good and poor readers. The number of words recalled, against the background of processing sentences, was considered a measure of the residual storage capacity of WM. Daneman and Carpenter (1980) found high correlations between reading span and three measures of reading comprehension; (1) answering fact questions ($r = .72$), (2) pronoun reference questions ($r = .90$), and (3) the Verbal Scholastic Aptitude Test (Exp. 1, $r = .59$, Exp. 2, $r = .49$). In addition, as they had hypothesized, a simple word span test given to the same subjects did not significantly correlate

with any of the three comprehension measures. Daneman and Carpenter (1983) replicated the significant correlations between reading span and Verbal SAT (.46 and .58) in two later studies. A similar correlation (r = .53) between WM span and the Nelson–Denny, a standardized test of reading comprehension, was found by Masson and Miller (1983). Recently, Daneman and Green (1986) found that reading span correlated with learning the meaning of novel vocabulary words in a context with sufficient cues for inferring meaning ($r(28)$ = .69). Furthermore, the correlation was still highly significant when vocabulary knowledge effects were statistically removed (r = .53).

From these findings, Daneman and colleagues argued that the reading span is an index of the WM capacity that is NOT allocated to processing (i.e., reading and comprehending) the individual sentences. Because good reading comprehenders have better or more efficient reading strategies than poor reading comprehenders, more capacity would remain for storage of to-be-remembered information. Thus, they argue that any measure reflecting the capacity of a WM that is important in reading comprehension must require the use of reading strategies. The WM span measure is dependent on the type of background task used while measuring the span, and that background task must include reading if the span measure is to predict individual variation in reading comprehension.

Another possible explanation of the Daneman and Carpenter findings, however, may be that people are good readers because they have a large WM capacity independent of the task being performed. Good readers may be remembering more words against the background of processing sentences in the reading span task because they have larger WM capacities than poor readers, NOT because good readers have more efficient reading skills than poor readers. A greater WM capacity would be independent of the type of background task used while measuring span. That is, a good reader may have more WM capacity available for processing and storage than a poor reader whether performing a reading or a non-reading task. According to this alternative theory, a measure of WM should successfully transcend task dependence in its prediction of higher level cognitive functioning. That is, the memory span task could be embedded in a concurrent processing task that is unrelated to any particular skills measure and still predict success in the higher level task.

Experiment 1

A pilot study performed in our lab supports the notion that WM storage capacity is independent of the nature of the task being performed (Turner & Engle, 1986). Span measures embedded in a processing task other than reading correlated with reading comprehension just as well as did the reading span. Three span measures were used in the Turner and Engle study. One was a replication of Daneman and Carpenter's (1980) reading span (referred to here as the sentence-word span test, Sentence Word) wherein subjects read a series of unrelated sentences and recalled the last word in each sentence. In a second span measure, a to-be-remembered digit followed each sentence in the series, (sentence-digit span test, Sentence Digit) and the digits were to be recalled following the end of the series. In the third span task subjects performed simple arithmetic operations (e.g., $(3 \times 4) + 11 = \ldots$) followed by a to-be-remembered word (operation-word span test, Operations Word). Memory span was defined as the maximum number of items (digits/words) recalled. The main purpose of this pilot study was testing whether the relationship between these span measures and reading comprehension is dependent or independent of specific processing strategies required by the secondary task. That is, does the secondary task need to involve reading, as Daneman and Carpenter (1980, 1983) suggested, to predict read-

ing comprehension? Daneman and Carpenter's hypothesis would predict that the Sentence Word and the Sentence Digit span tasks would reflect differences in residual WM capacity because of individual differences in reading skills and strategies, not because of differences in WM capacity independent of task proficiency.

The alternative explanation, that the reading span measures abiding individual differences in WM capacity independent of the skills required for the processing component of the task, suggests that the complex memory span index could be embedded in any task that requires heavy processing beyond the span task and still reflect individual differences in WM capacity that are important in higher level functioning. Using a concurrent processing task involving strings of arithmetic operations followed by a to-be-remembered word (Operations Word), allowed the span task to be embedded in a processing task that requires a different set of strategies than reading comprehension. Nevertheless, the word span measured in the operation-word task predicted reading comprehension just as well as it did in the sentence-word (i.e., reading span) or sentence-digit task.

One implication of these results is that the background task in a WM measure does not need to be reading related to lead to a correlation between the span measure and a test of reading comprehension. An individual may be a better reading comprehender because of a larger WM capacity not specific to more efficient reading skills.

However, there is a possible confounding in the Turner and Engle study. Reading ability and mathematical ability tend to be highly correlated. Thus, good readers also tend to be good in math. Daneman and Carpenter's theory could still be correct by the following analysis. Good readers could have a large sentence-word span because of their efficient reading skills and, independently, they could have a large operation-word span because of their efficient mathematical skills. The dissociation of the nature of the background task and the nature of the task being predicted, in this case, reading comprehension, would only be meaningful when the two types of skills are not themselves correlated.

The purpose of the first experiment was to further address the question of whether WM storage capacity is task dependent. More importantly, the study addressed whether the correlations observed by Turner and Engle are simply due to good readers also having good quantitative skills and using both kinds of skills efficiently, or, whether they generally have larger WM capacities.

One way to approach this problem is to statistically remove the effects of quantitative skills (as reflected by Quantitative SAT scores) from the correlation between each of the WM span measures with the reading comprehension measures. If the correlation between the operation-word span and reading comprehension is simply due to good readers also having good and efficient quantitative skills, then the correlation between operation-word span and comprehension should disappear when the quantitative skills are factored out. On the other hand, if the correlation between the WM span measure and reading comprehension is independent of the particular skills involved by the background task, the partial correlation coefficient between operation-word span and reading comprehension should still be significant.

Method

The Turner and Engle experiment suggested that one of the reasons people are good readers may be that they have a large, general WM capacity that is independent of specific task strategies. The purpose of this experiment was to further test this idea while eliminating alternative hypotheses. It would help to understand the procedure of these experiments if we consider WM tasks like Daneman and Carpenter's reading span as really the combination of two tasks. The primary task is remembering

items, either words or digits. But this primary memory task is performed against the background of a secondary or processing task like reading sentences. Daneman and Carpenter argue that the significant correlation between the number of items recalled in the primary span task will only correlate with a reading comprehension task when the secondary or background task is also reading related.

Since different skills are clearly required for solving arithmetic problems than when reading sentences, the secondary or background task of the current study was either solving arithmetic operations or reading unrelated sentences. In addition, since the recall of digits likely requires different memory strategies than those used when remembering words, the to-be-remembered items in the memory component of the task were digits or words.

Therefore, the first experiment included a replication of our pilot study, and in addition, asked whether any similar correlations found between the different complex WM spans and comprehension were simply due to good readers also having good quantitative skills and using both kinds of skills efficiently, or, whether they just generally have larger WM capacities independent of these skills.

Subjects

The study tested 243 University of South Carolina psychology students who participated to satisfy a course requirement or receive extra credit. All subjects were tested in groups of at least two and no more than five and completed the tasks in two 1-h sessions.

Design

All subjects completed seven tasks: four complex WM span tasks, two simple span tasks, and the Nelson–Denny Standardized Reading Comprehension test. In addition, written permission was obtained from each subject to obtain their Verbal and Quantitative Scholastic Achievement Test (SAT) scores from the University.

The four complex memory span measures were: (1) one in which subjects read a series of sentences, verified whether each sentence made sense and then recalled the last word in each sentence (sentence-word span task, Sentence Word), (2) one in which subjects also read a series of sentences and verified whether each sentence made sense and recalled a to-be-remembered digit which followed each sentence (sentence-digit span task, Sentence Digit), (3) one in which the subjects verified the answers to strings of arithmetic operations and recalled a to-be-remembered word which followed each operation (operation-word span task, Operations Word), and (4) one in which the subjects verified the answers to strings of arithmetic operations and recalled the digit answer that was printed to the right of the equal sign regardless of whether or not it was the correct answer (operation-digit span task, Operations Digit). There were 42 stimuli in each of the four complex span tasks (Sentence Word, Sentence Digit, Operations Word, Operations Digit), presented in 12 trials. The number of stimuli in each trial was gradually increased from two to five, with three trials at each level.

The two simple span measures were: (1) a digit span task in which subjects were asked to remember increasingly larger sets of randomly generated digits in the correct serial order and (2) a word span task, in which the sets of items to-be-remembered were words rather than digits. There were 132 digits presented in 27 trials and 81 words presented in 18 trials. The number of stimuli (words/digits) in each trial was gradually increased from 2 to 9 in the digit task, and from 2 to 7 in the word task, with three trials at each level in both simple span tasks.

The Nelson–Denny Reading Comprehension Test

Form F of the Nelson–Denny standard-

ized Reading Comprehension test was administered. A total of 8 passages, each approximately 250 to 650 words in length, and 36 questions were contained in the test pamphlet. Subjects were instructed to silently read each passage and then to answer the multiple choice questions that followed each passage on the numbered answer sheets provided. They were allowed to look back at the material they had read, but were advised to leave difficult questions until they had finished answering the remaining questions. Subjects were given 20 min to complete the test. Reading rate was determined for each subject by having the experimenter call "Mark" when 30 s had elapsed. At that time subjects recorded the number of words read, which was printed to the right of the line on which each was reading. Accuracy scores were based on the number of correct answers out of the total 36 questions.

Memory Span Tasks

Word stimuli. The words to be recalled in the sentence-word, operation-word, and simple-word tasks were selected from a total pool of 243 words and then divided into three sets (i.e., Set A, Set B, and Set C) so that each subject recalled different words in each of the three span tasks. The to-be-remembered word was the last word in the sentence in the Sentence Word task, or the word that followed the operation string in the Operations Word task, or was the stimulus item in the simple word task. All three sets of words were one syllable concrete nouns, selected from the most common four to six letter words published in the Francis and Kucera (1982) frequency norms.

Sentence stimuli. Sentences for the sentence-word and sentence-digit tasks were generated to make sense with the designated last word. Each sentence was from 11 to 16 words long, was nominally unrelated to the others, and was either "correct" or "incorrect." "Correct" sentences made semantic and syntactic sense, e.g., "The grades for our finals will be posted outside the classroom door." "Incorrect" sentences were made nonsense by reversing the order of the last four to six preterminal words, e.g., "The grades for our finals will classroom the outside posted be door." The number of "correct" and "incorrect" sentences in a trial (i.e., set size) gradually increased from 2 to 5, with three trials at each set size. Approximately half of the sentences in each set size were randomly selected to be "correct," and half "incorrect," with the constraint that 21 sentences in each task were "correct" and 21 were "incorrect."

Digit stimuli. The to-be-remembered digits in the sentence-digit, operation-digit and simple-digit tasks were randomly sampled with replacement from the integers, 1–9. A single to-be-remembered digit followed each sentence in the sentence-digit task, or, was the stated answer in the operation-digit task, or was the stimulus item in the simple digit task.

Operation stimuli. A total of 84 operation strings served as stimuli for the processing component of the operation-word and operation-digit complex span tasks. Each string consisted of two arithmetic operations and a stated final answer, e.g., [(9/3) − 2 = 1]. The first operation was a simple multiplication or division problem in parentheses, such as (3 × 4) or (8/2). The inter-item product or quotient was not stated and was to be solved prior to a second operation, i.e., the simple addition or subtraction of a single-digit integer. Approximately half of the operation strings in each trial listed an answer after the equal sign that was correct, and half listed an answer that was incorrect by at least 4, e.g., [(9/3) − 2 = 6]. The number of operation strings in a trial (i.e., set size) gradually increased from 2 to 5, with three trials at each set size. As with the sentence stimuli, approximately half of the strings in each set size were answered correctly and half incorrectly.

General Procedure

In all of the memory span tasks, small groups of subjects read aloud series of stimuli that they saw projected from transparencies onto a large screen while simultaneously hearing prerecorded items on a cassette tape. The length of each recording determined the presentation rate for each stimulus item. The average playback time for the four different stimulus items in the complex tasks was 423 ms for the Sentence Word task, 532 for the Sentence Digit task, 396 for the Operations Word task, and 541 for the Operations Digit task. The experimenter kept all items covered on the transparency, except the item being presented, by gradually moving a blank sheet of paper with a $6'' \times 0.5''$ cut out, creating a "window" through which the stimulus item was projected onto the center of the screen. Thus, all subjects heard the stimulus items while seeing them on the screen and were required to read the items aloud, along with the voice on the tape. Subjects performed in groups of from two to five subjects, sitting at individual desks, positioned so that the experimenter could see the mouths of all subjects. Then, if necessary subjects were reminded to read aloud, and in addition, the experimenter mouthed all items, which served as a constant reminder for subjects to continue reading aloud.

The first trial was initiated by the experimenter after reading specific instructions at the beginning of each task. In the Sentence Word task the experimenter allowed the first sentence to be projected through the "window" on the screen and played the recording of the same sentence. Subjects immediately began reading the sentence aloud paced by the recording. As soon as each subject decided whether a sentence made sense, he or she placed a check mark in the appropriate numbered blank on Side 1 of his or her answer sheet. Immediately after the first sentence was presented, the next sentence appeared on the screen and subjects read it aloud, again paced by the recording of the sentence. After a series of two sentences were read, the subjects saw a red line and heard a "recall" cue at which time they turned their answer sheets to Side 2 and wrote down the last word of each of the sentences they had read. Although they were instructed to write the words in any order, subjects typically wrote the words in the same order as presented. After a set of three trials at set size 2, the number of sentences was increased to three.

And, after each additional set of three trials, the number of sentences continued to be incremented by one until subjects were reading five sentences prior to the "recall" cue. Likewise, in the sentence Digit task, wherein each sentence was followed by a digit, subjects saw, heard, and read each series of sentences and digits aloud, along with the voice on the tape. Again, subjects verified whether the sentence made sense, and when they heard the "recall" cue they serially recalled the to-be-remembered digits. In the Operations Word task subjects also saw, heard, and read aloud series of operations strings and to-be-remembered words, again paced by the recorded stimulus presentation. Subjects immediately verified whether each stated answer following the equal sign was correct or incorrect, and finally when the "recall" cue was heard they typically wrote the to-be-remembered words in serial order, but they were told they could record them in any order. In the Operations Digit task subjects saw, heard, and read aloud the series of operation strings, paced by the recording, and as soon as possible verified whether the stated answer was correct. When the recall cue was heard they recalled the answer stated after the equal sign in the correct serial order, whether or not the stated answer was correct. In the Operations Word and Operations Digit tasks all intermediate computations were required to be silent and no aids (i.e., pencil or paper) were allowed.

In the two simple span tasks subjects

saw, heard, and read aloud words or digits, sequentially presented at a rate of one per second. The number of words presented prior to the recall cue was gradually increased from 2 to 7, and the number of digits was increased from 2 to 9. When the recall cue was heard, subjects wrote their answers on answer sheets numbered from 1 to 81 for word recall, and from 1 to 132 for digit recall. Digits were recalled in the correct serial order and words in any order.

Results Experiment 1

The data from this study consisted of 22 scores for each of the 243 subjects, they were derived from the four complex memory tasks, the two simple memory tasks, the Nelson–Denny Reading Comprehension task, and the Verbal, Quantitative, and Total SAT tests. The data from the four complex span tasks consisted of one measure of the processing component and two of the storage component of the task. The processing component measure was the number of correct verifications for sentences or operations in each task. The span data for a subject were not included in the analysis if the verification measure was below 80%. The data for two subjects were excluded from each of the complex spans. One recall measure was a Set-Size Memory Span. This was the maximum size of the set of to-be-remembered items (words or digits) in which the subject perfectly correctly recalled the memory items two out of the three times. A second recall measure was a Total Memory Span. This was the total number of correctly recalled words/digits from all trials. Within each trial, words could be correctly recalled in any order, while digits had to be recalled in the correct serial order to be included in the total memory span. These two measures were also derived for each of the two simple span tasks.

The scores derived for each subject from the Nelson–Denny Reading Comprehension test consisted of (1) the number of correct answers out of all 36 possible questions, (2) the percentage of correct answers out of the total questions completed, and (3) the number of words read per minute. The second measure was used as an attempt to express comprehension independent of reading speed. However, since this measure did not correlate with the span or any other comprehension measures used in the current study and is not the standardized measure for the Nelson-Denny, it will not be discussed further.

The two span measures led, with few exceptions, to the same conclusions so only the results of the total memory span will be reported. This measure of WM has been used by other researchers (e.g., Baddeley, Logie, Nimmo-Smith, & Brereton, 1985 and Masson & Miller, 1983) and will allow better comparison across studies. The highest possible score for this measure was 42 for the complex spans, 81 for the simple word span, and 132 for the simple digit span. Table 1 reports the descriptive statistics for the WM measures and comprehension measures.

Reliability analysis. Reliability estimates of the complex and simple span scores ranged between .89 and .93. The reliability estimates for subjects' memory span scores were derived from Cronbach's alpha formula as measures of internal consistency. Intercorrelations were computed among the total number of items correctly recalled in each of the three trials across all set sizes for each memory task. For example, in the Sentence Word task, the total number of correctly recalled words in the first trials of set sizes 2, 3, 4, and 5 was calculated as one individual Sentence Word span, a second individual Sentence Word span was the total number of words recalled in the second trials of all the set sizes, and finally, a third individual Sentence Word span was the total number of words correctly recalled in the third trials of all set sizes. The intercorrelations computed among these three trial spans ranged from .78 to .80 for the Sentence Word task, from .74 to .81 for the Operations Word task, from .74 to .83 for the

TABLE 1

Descriptive Statistics for WM total score

	Tasks					
	Sentence word	Sentence digit	Operation word	Operation digit	Simple word	Simple digit
Mean	29.1	32.5	32.6	28.4	63.0	95.8
SD	5.1	5.8	4.4	6.2	10.6	19.2
Minimum	17	18	19	15	23	24
Maximum	41	41	41	40	84	130

Descriptive Statistics for Comprehension and SAT Scores

	Nelson–Denny number correct	Reading rate	Verbal SAT score	Quantitative SAT score
Mean	26	269.8	441.5	456.9
SD	5.03	86.3	88.47	91.1
Minimum	8	36	240	270
Maximum	35	756	680	720

Sentence Digit task, from .72 to .76 for the Operations Digit task, from .71 to .82 for the word task, and from .69 to .75 for the digit task.

Reading comprehension. Table 2 shows the Pearson Product Moment correlation coefficients calculated between the total memory span and the comprehension measures. The major goal of Experiment 1 was to determine whether the relationship between the WM span scores and reading comprehension varies as the nature of the WM background task was varied in the complex tasks. Table 2 shows the correlational coefficients central to this question. Clearly all four complex span measures correlated significantly with the Nelson–Denny even though the magnitude of that correlation was nearly twice as high for those tasks involving the recall of words as for those involving the recall of digits. The Nelson–Denny correlated significantly with Sentence Word span, $r(241) = .37$, $p < .0001$, with Operations Word span, $r(241) = .40$, $p < .0001$, with Sentence Digit span, $r(241) = .20$, $p < .002$, and with Operations Digit span, $r(241) = .24$, $p < .002$. These correlations suggest that comprehension may be related to WM span, whether measured against a background task involving reading or performing arithmetic problems. Subjects with higher Sentence-Word spans were better reading comprehenders than subjects with lower Sentence-Word spans. More importantly, subjects with higher Operations-Word spans were also better at reading comprehension than those subjects with lower Operations Word spans. The correlations of Nelson–Denny with Sentence-Word and Operation-Word spans were not statistically different, $t(241) = 0.32$, $p > .50$.

TABLE 2
CORRELATIONS BETWEEN TOTAL SPANS AND COMPREHENSION MEASURE

Span task	Nelson–Denny accuracy scores	Verbal SAT scores	Quantitative SAT scores
Sentence Word	.37**	.28**	.26**
Operation Word	.40**	.34**	.33**
Sentence Digit	.20*	.08	.24*
Operation Digit	.24*	.11	.24*
Word	.07	.08	.09
Digit	.10	.12	.12

* $r(241)$, $p < .002$.
** $r(241)$, $p < .0005$.
$N = 243$.

The magnitude of the correlation between the two complex digit spans and comprehension was relatively smaller than it was between the two complex word spans and comprehension. The correlations

between the Sentence-Digit span and Nelson–Denny, $r(241) = .20$, $p < .0023$, and between the Operations-Digit span and Nelson–Denny, $r(241) = .24$, $p < .0002$, were not significantly different from each other, $t(241) = 0.57$, $p > .50$. Although all four complex spans correlated significantly with reading comprehension, those requiring word recall predicted reading comprehension significantly better ($r = .37$ and .40) than those tasks requiring the recall of digits ($r = .20$ and .24). Significant differences at the .05 level were found between all combinations of the Sentence Word and Operations Word pairs of spans and the Sentence Digit and Operations Digit pairs of spans; Sentence Word/ND ($r = .37$) and Sentence Digit/ND ($r = .20$), $t(241) = 2.90$, Sentence Word/ND ($r = .37$) and Operations Digit/ND ($r = .24$), $t(241) = 1.99$, Operations Word/ND ($r = .40$) and Sentence Digit/ND ($r = .20$), $t(241) = 3.23$, and Operations Word/ND ($r = .40$) and Operations Digit/ND ($r = .24$), $t(241) = 2.13$.

Verbal SAT. Differences in the predictability of reading comprehension by the two types of complex spans (digit and word) were also found in the correlations between the complex spans and VSAT scores. Table 2 shows that both complex word spans correlated significantly with VSAT, Sentence Word/VSAT, $r(208) = .28$, $p < .0001$, and Operations Word/VSAT, $r(208) = .34$, $p < .0001$. On the other hand, neither of the correlations involving complex digit spans and VSAT were significant, Sentence Digit/VSAT, $r(208) = .08$, $p > .24$, and Operations Digit/VSAT, $r(208) = .11$, $p > .10$. Thus, there was no relationship found between VSAT and complex digit spans, but complex word spans did predict VSAT.

Quantitative SAT. There was a relationship, however, between all four complex spans (Sentence Word, Operations Word, Sentence Digit, and Operations Digit) and QSAT. Table 2 shows significant correlations between Sentence Word and QSAT, $r(208) = .26$, $p < .0002$, Operations Word and QSAT, $r(208) = .33$, $p < .0001$, Sentence Digit and QSAT, $r(208) = .24$, $p < .0005$, and Operations Digit and QSAT, $r(208) = .25$, $p < .0003$. These significant and nearly equivalent correlations between the four complex spans and QSAT might be expected, considering that the different components of the QSAT would invoke both verbal and quantitative skills, i.e., names of the digits, etc., and that VSAT correlated with QSAT, $r(208) = .54$, $p < .0001$.

Simple spans. As expected, the simple word and digit spans did not predict reading comprehension. The correlation between the simple word span and Nelson–Denny was, $r(241) = .07$, $p > .261$, and between simple digit span and Nelson–Denny was, $r(241) = .10$, $p > .19$. In addition, Table 2 shows there were no significant correlations between VSAT or QSAT with word span, $r(208) = .08$ and .09 respectively, or with digit span, $r(208) = .12$ and .12, respectively. Therefore, individual differences in the number of words or digits recalled did not predict individual differences in comprehension as measured by Nelson–Denny, VSAT, or QSAT scores.

Reading rate. There was no relationship between reading rate and WM capacity. The only measures that approached significance in correlation with reading rate were the Sentence Word and Operations Word complex spans, $r(241) = .17$ and .19, respectively, $p < .06$. No other span and reading rate correlations approached significance.

Span intercorrelations. The four complex spans tended to correlate with one another, ranging between .38 and .58 (see top half of Table 3). Not surprisingly, the highest span correlations were between the two complex spans requiring word recall (Sentence Word and Operations Word), $r(241) = .53$, $p < .0001$, and, between the two complex spans requiring digit recall (Sentence Digit and Operations Digit), $r(241) = .58$, $p < .0001$. The remaining correlations were also fairly large among the complex

TABLE 3

Correlations among Complex and Simple Spans

	Tasks					
	Sentence word	Sentence digit	Operation word	Operation digit	Simple word	Simple digit
Sentence Word	—					
Sentence Digit	.50**	—				
Operation Word	.53**	.43**	—			
Operation Digit	.38**	.58**	.38**	—		
Word	.10	.03	.08	.15	—	
Digit	.24*	.27*	.35**	.18*	.46**	—

Correlations among Comprehension and SAT Scores

	Nelson–Denny accuracy scores	Verbal SAT scores	Quantitative SAT scores
Nelson–Denny	—		
VSAT	.66**	—	
QSAT	.41**	.54**	—

* $r(241)$, $p < .01$.
** $r(241)$, $p < .0001$.

spans, Sentence Word/Sentence Digit, $r(241) = .50$, Sentence Word/Operations Digit, $r(241) = .38$, Operations Word/Sentence Digit, $r(241) = .43$, and Operations Word/Operations Digit, $r(241) = .38$, at the $p < .0001$ level. In addition, Table 3 shows a large correlation between the simple word and digit spans $r(241) = .46$, $p < .0001$. Thus, the simple spans were highly correlated, and the complex spans highly correlated; however, the correlations across the different types of spans (complex and simple) were much lower, ranging between $-.03$ and .35. To the extent that the correlations among the complex spans were similar, these memory measures may be tapping the same underlying process. And, of course, the correlation between the two simple spans suggested these two measures may be, at least partially, determined by the same underlying process.

Comprehension and SAT intercorrelations. As expected, the correlations among the comprehension and psychometric measures were all large and significant: (1) Nelson–Denny (ND) and VSAT, $r(208) = .66$, $p < .0001$, (2) ND and QSAT, $r(208) = .41$, $p < .0001$, and (3) VSAT and QSAT, $r(208) = .54$, $p < .0001$ (see bottom half of Table 3).

Summary of correlational analyses. The correlational analyses showed that comprehension had a relationship with the four complex spans and that the simple spans did not correlate with comprehension. In addition, the complex word spans related more highly with comprehension than did the complex digit spans. Span intercorrelations also were higher between the two simple spans, and the complex word spans, than those correlations between simple and complex spans. In addition, the relationship between the complex digit spans (Sentence Digit/Operations Digit) and between the complex word spans (Sentence Word/Operations Word) was greater than between complex word and digit spans (Sentence Word/Sentence Digit, Sentence Word/Operations Digit, or Operations Word/Sentence Digit, Operations Word/Operations Digit). These findings were based on tests of differences in the magnitude of correlations.

However, as pointed out by Marascuilo and Levin (1981), what determines the strength of a relationship, is the closeness

of points to a regression line. Thus, any two sets of correlations may be found not significantly different (i.e., the two different sets of points, or ellipses, may be equally close to their respective regression lines), but the corresponding slopes of the two regression lines may not be similar. A linear prediction of reading comprehension may be quite different when using one WM complex span, than when using another, even though the strengths of the two linear predictions (i.e., the two correlations between spans and comprehension) are similar.

The regressions of Sentence Word with Nelson–Denny and of Operations Word with Nelson–Denny were central to the present study. An analysis was performed to test the hypothesis that the slopes of these two regression lines were parallel. This hypothesis could not be rejected, $F(1,234) = 0.07$, $p > .25$, $MSE = 40.46$. In other words, the two complex spans requiring the recall of words were similar in their ability to predict reading comprehension as measured by the Nelson–Denny.

Linear relationship between span and comprehension. A principal-components analysis confirmed that the relationships between reading comprehension and complex spans reflecting WM capacity should be additionally explored in a regression analysis. One way to approach the problem of whether one, two, or all of the four complex spans are needed to predict comprehension, or whether the spans similarly account for the variation in reading comprehension, is through a stepwise multiple regression analysis using the forward selection technique (Marascuilo & Levin, 1981).

This procedure allowed a different complex span measure to serve as the first, most powerful predictor of comprehension in each of four models. The relative contributions of each remaining complex span in predicting comprehension in each of the different models were then compared. Table 4 shows the results of the stepwise regression analysis.

Table 4 indicates the increment in the ND comprehension variability that was accounted for by each span measure when entered into two different models. In the two models, entering Sentence Word and Operations Word first, the change in the predicted variance of ND comprehension due to the addition of a third most powerful predictor span was never significant, indicating that Sentence Word and Operations Word spans were sufficient predictors of comprehension. Entering Operations Word

TABLE 4
STEPWISE REGRESSION WITH THREE COMPREHENSION MEASURES AS CRITERION

Order of entry	Predictor variable	Proportion of variance accounted for	Model		
			R	*F*	*p*
Nelson–Denny					
1	Operation Word	.1603	.1603	45.25	.0001
2	Sentence Word	.0326	.1929	9.53	.0023
1	Sentence Word	.1342	.1342	36.74	.0001
2	Operation Word	.0587	.1929	17.17	.0001
VSAT					
1	Operation Word	.1146	.1146	26.66	.0001
2	Sentence Word	.0154	.1300	3.63	.058
3	Sentence Digit	.0162	.1462	3.87	.050
QSAT					
1	Operation Word	.1101	.1101	25.48	.0001
2	Sentence Digit	.0135	.1236	3.16	.0771

span first accounted for 16.03% of the variance, while Sentence Word added 3.26%. Entering Sentence Word span first in this fashion accounted for 13.42% of the variance, while Operations Word added 5.87%.

Table 4 also shows the stepwise regression analysis with VSAT as the criterion variable. When the four spans were allowed to compete in the stepwise procedure Operations Word span was selected first, accounting for 11.46% of the variance in VSAT, with Sentence Word span next, accounting for an additional 1.54%, and Sentence Digit span last, accounting for an additional 1.62% of the final model, $F(3,204) = 11.65$, $p < .0001$. In addition, the Operations Word span made a unique contribution to VSAT when the Sentence Word span was entered first in the stepwise regression procedure. When Sentence Word span was entered first, 7.93% of the variance was accounted for in VSAT, the Operations Word span was selected next, accounting for 5.7% of the variance in VSAT. This finding that the Operations Word task is a unique predictor of reading ability, even after the Sentence Word task is partialed out is difficult for a theory based on reading skills to explain. The Operations Word span was the only variable that accounted for a significant proportion of variance in QSAT, 11.01%, $F(1,208) = 25.48$, $p < .0001$.

The question addressed by this study was whether WM capacity is task dependent. Does a WM measure need to be "reading" related to generate a significant correlation between the span measure and a test of reading comprehension? Finding similar correlations between the four complex spans and the ND comprehension measure implies WM capacity transcends task. These results clearly show that the processing component of the WM span does NOT need to be "reading" related to produce a correlation between the span measures and reading comprehension. This suggests individuals may be good or poor reading comprehenders because of a large or small WM capacity, not because of more or less efficient reading skills. However, there is another possible explanation addressed by this study, it may be that these similar correlations are simply due to good readers also having good quantitative skills in which case the significant correlation of Operations Word with the measures of reading comprehension would be accidental in nature. The correlation between Sentence Word and comprehension could occur because, as Daneman and Carpenter argue, the Sentence Word task causes reading skills to be invoked and the residual WM capacity is reflected by the number of words recalled. The Operations Word task would lead to arithmetic skills being invoked and, since verbal and quantitative skills tend to be correlated (.54 between VSAT and QSAT in the current sample), the WM capacity reflected by the Operations Word task would tend to be similar to the WM capacity reflected by the Sentence Word task. However, WM capacity would still be task dependent. This problem was approached with a partial correlational analysis of the data.

Partial correlation analysis. The partial correlation technique was used to address the question of whether the significant correlation between Operations Word and reading comprehension measures was an artifact of the tendency for verbal and quantitative skills to correlate. In this procedure, the possibility was considered that the correlations between each of the complex spans and ND comprehension were confounded by the covariance between verbal and quantitative abilities. Therefore, the effect of quantitative abilities (QSAT) was partialled out prior to measuring the relationships between the complex spans and comprehension. Table 5 shows significant values for the following partial correlations: (1) Sentence Word and ND, $r(205) = .25$, (2) Operations Word and ND, $r(205) = .25$, (3) Sentence Word and VSAT, $r(205) = .17$, and (4) Operations Word and VSAT, $r(205) = .20$. All other partial correlations

TABLE 5
PARTIAL CORRELATIONS BETWEEN SPAN AND COMPREHENSION MEASURES

Span measures	Comprehension measures: Nelson–Denny accuracy scores	Comprehension measures: Verbal SAT scores
Sentence Word	.25***	.17*
Operation Word	.25***	.20**
Sentence Digit	.10	−.07
Operation Digit	.10	.02
Word	.06	.03
Digit	.09	.02

* $r(208)$, $p < .01$.
** $r(208)$, $p < .006$.
*** $r(208)$, $p < .0004$.

were not significant. The important point is that the relationship between Operations Word span and reading comprehension (Nelson–Denny) is still present when quantitative skills (QSAT) are factored out of the association. However, neither of the complex digit spans (Sentence Digit and Operations Digit) showed a significant relationship with comprehension after the effect of QSAT was removed. The zero order correlations for Sentence Digit/ND (r = .20) and Operations Digit/ND (r = .24) were half the magnitude of the Sentence Word/ND (r = .37) and Operations Word/ND (r = .40) zero order correlations.

In summary, the significant partial correlations between the Operations Word and Sentence Word spans and comprehension suggests the complex span measures of WM capacity may be independent of the particular skills involved in the processing component of the span task. The correlation between Operations Word span and reading comprehension was not simply due to good readers also having good quantitative skills.

EXPERIMENT 2

The purpose of this experiment was to study the relationship between the complex WM span and comprehension measures while manipulating the difficulty of the processing component of the span tasks. The findings of Experiment 1 showed that the complex spans reflecting WM correlated with reading comprehension. More importantly, the correlations found between complex spans and comprehension did not depend on the use of specific reading skills in the processing component of the span task. The complex memory span related to reading comprehension even if it was embedded in a concurrent processing task that was unrelated to reading skills. However, there was no relationship found between comprehension and the traditional span measures in Experiment 1. Why was reading comprehension predicted by the complex WM spans, but not by the simple word or digit span measures? One reason for this may be that individual differences in the simple spans are a result of differences in the use of memory strategies such as chunking, rote rehearsal, phonetic recoding, and elaboration. It is unlikely that these same strategies, with the possible exception of phonetic recoding, would be very important to reading comprehension. The complex span measures may more closely reflect the number of "items" that can be represented in the working memory without the aid of rehearsal. The complex span measures may correlate with reading comprehension because the processing component of the task (i.e., reading unrelated sentences or solving operation strings) inhibits the use of these memory strategies. This theory suggests that any method of eliminating the use of memory strategies while measuring WM should lead to a more accurate measure of this limited capacity memory.

One way to test whether the processing component of the complex span task is simply preventing memory strategy use, is to vary the difficulty of the processing component. Theoretically, the span measured while processing very simple sentences, e.g., "See the dog," should be similar to the simple span measure. The processing of these simple sentences would not be so demanding that subjects could not make some

use of memory strategies. On the other hand, reading sentences, such as, "She saw the small dog run behind the truck," would be more demanding and would minimize strategy use. Therefore, spans measured against an easy background task would not likely predict reading comprehension as well as spans measured against a more difficult background task. This suggests that as the difficulty of the processing component of the task increases, strategy use would be less likely, and the resulting relationship between span and reading comprehension would increase. However, when trying to remember items while processing very difficult sentences, such as, "The young lady with the old man saw the small dog run behind the brand new 1987 Ford truck," the correlation between span and comprehension may break down. If the background becomes too difficult several things could happen that would diminish the relationship between the complex span and reading comprehension. One is that some subjects would just give up on the background task and allocate all their resources to the span component. So, to some extent, for these subjects the task would be more like the simple span.

Another result of an exceedingly difficult background task could be that span performance would be compressed to the extent that the range of scores would be restricted. And, to the extent that the range of the span scores is restricted, the resulting correlation between comprehension and span would be decreased. Therefore, since spans measured against easy background tasks may not inhibit the use of memory strategies, resulting in a span similar to the traditional simple spans, and since spans measured against very difficult background tasks would be restricted in range, a comparison of correlations between comprehension and easy, moderate, and difficult complex spans might result in an inverted U-shaped function.

It is also important to note that if the correlation between sentence-word spans and reading comprehension is considered to depend on the use of reading skills in the processing component of the span task, as Daneman and Carpenter argued, then, as the background task is made more difficult the resulting Sentence Word span and comprehension correlations would behave in the manner described above. The Sentence Word span/comprehension correlations should increase in magnitude, up to that point where the difficulty level restricts the range of the Sentence Word span measure. However, if, as Daneman and Carpenter argue, the correlation between comprehension and complex span is determined by specific task skills used in the background task, then when the difficulty of a background task consisting of operation strings is manipulated, the resulting pattern of span and reading comprehension correlations should not be the same as the pattern of correlations between Sentence Word span and reading comprehension. Solving operation strings clearly does not require reading skills, yet the Operations Word span measured against the background task of verifying operation strings correlated with reading comprehension ($r = .40$ in Experiment 1). The Operations Word span/comprehension correlation suggested specific task skills did not determine the complex span measures reflecting WM capacity. If a comparison of the behavior of Operations Word span/comprehension correlations with Sentence Word span/comprehension correlations should result in similar shaped functions, it would appear that a task independent explanation would be further supported.

Method

In this experiment the difficulty of the processing component of the complex Sentence Word and Operations Word span tasks was manipulated. Three levels of difficulty were defined for the processing components of the Sentence Word and Operations Word tasks. Unrelated sentences used in the verification component of the

Sentence Word task were simple, moderately difficult, or difficult, and, the operation strings used in the verification component of the Operations Word task were simple, moderate, or difficult. Thus, there were three Sentence Word and three Operations Word span tasks. Reading comprehension was tested with Form F of the Nelson–Denny Reading Comprehension Test. The stimuli and administrative procedures for this standardized test were identical to those described in Experiment 1.

Subjects

There were 52 undergraduate students enrolled in psychology courses at the University of South Carolina who were tested individually in two 1-h sessions. Two of the subjects were unable to complete the difficult Operations Word verification task, and therefore, their data were not used. The remaining 50 subjects completed six complex WM span tasks and one comprehension task. The order in which the seven tasks were completed was randomized across individuals and sessions to balance practice and boredom effects.

Memory Span Tasks

The word stimuli used in the memory component of the Sentence Word and Operations Word complex tasks were identical to those used in Experiment 1. Again, three sets of words were used so that subjects would not recall the same words in any one session.

Stimuli for sentence word task. As described in Experiment 1, the unrelated sentences were generated backwards using words from the three sets of stimulus words as the last word in each sentence. Sentence difficulty was varied in two ways: (1) linguistic difficulty and (2) the length of the sentences. The linguistic difficulty of sentences has been tested by many researchers. For example, Miller and McKean (1964) found the reaction time required to transform kernel active-affirmative sentences was a function of the complexity of the transformation required. These findings suggested that processing is more difficult for passive and negative than active and affirmative sentences. Therefore, the unrelated sentences in this experiment were active-affirmative (simple), passive-affirmative (moderate), or passive-negative (difficult). The length of the sentences was varied by adding one phrase (for the moderate sentences) or two phrases (for the difficult sentences) to the kernel sentences used in the simple sentences.

Sentence stimuli. There were three sets of sentences totalling 180 simple sentences (60 in each of three sets), 126 moderately difficult and 126 difficult sentences (42 in each of three sets). The simple sentences were active and affirmative, consisting of four to five words, e.g., "People gave their time." The moderately difficult sentences were passive and affirmative, consisting of 8 to 11 words, e.g., "Money is given by people at Christmas time." The difficult sentences were passive and negative, consisting of from 10 to 15 words. and were one phrase longer than the moderately difficult sentences, e.g., "Money was not given by people in that state at Christmas time."

Half of these sentences made sense and half were made "nonsense" by reversing the order of the middle two words in the simple sentences, e.g., "People their gave time," or the last four or five preterminal words in the moderate and difficult sentences, e.g., "Money is given Christmas at people by time" or "Money was not given by people in Christmas at state that time."

Manipulation check on sentence difficulty. A study was conducted to test whether response time to verify the three different types of sentences conformed to expectations. All simple, moderate, and difficult sentences used in Experiment 2 were presented on a computer monitor, one at a time, to 24 subjects whose task was to verify whether each of the sentences made sense. Subjects were given reaction time and accuracy feedback after each verifica-

tion trial and were instructed to maintain performance at 90% accuracy. Each verification error produced a "beep" from the computer. When there were more than four beeps in a row the experimenter repeated the instructions, stressing that the subject take longer to respond in order to be more accurate. Subjects were most accurate when verifying simple sentences ($\bar{x}$ = 98.7% correct), somewhat less accurate when verifying moderate sentences ($\bar{x}$ = 93.7% correct), and even less accurate when verifying difficult sentences ($\bar{x}$ = 86.6% correct). Subjects were fastest at verifying simple sentences ($\bar{x}$ = 1955 ms), slower at verifying moderate sentences ($\bar{x}$ = 2943 ms), and slowest at verifying difficult sentences ($\bar{x}$ = 3947 ms). Thus, the performance function resulting from the pilot study confirmed that the sentences used as stimuli in Experiment 2 are differentially difficult as described.

Stimuli for operations word task. The difficulty of the operation strings in the Operations Word task was varied in two ways: (1) the number of operations in the string and (2) whether the fraction, 0.5, was used in the string. Each of the simple operation strings (60 in each of three sets) consisted of the addition or subtraction of two single-digit integers followed by an equal sign followed by an answer. All of the addends were randomly selected with replacement from the digits 1–9, with the constraints that the same digit not be used for both addends, and, that the correct sum also be a digit between 1 and 9 (e.g., 3 + 1 = 4; 8 − 5 = 3). Half of the stated sums were correct and half were incorrect by 4.

Each of the 126 moderately difficult operations (42 in each of three sets) consisted of two operations with a stated final answer, (e.g., (9/3) + 4 = 7). The first operation in the string was a multiplication or a division problem in parentheses, such as (6/3) or (2 × 3), followed by the addition or subtraction of a number between 1 and 9. The integers were all randomly sampled with replacement from the numbers 1–9 with the constraints that the unstated interitem product, and the stated final answer, be whole numbers between 1 and 9. Half of the stated final answers were correct and half were incorrect by 4.

Similar to the moderate operations, each of the 126 difficult operation strings (42 in each of three sets) consisted of a multiplication/division operation, followed by an addition/subtraction of a number, with a stated final answer. However, the digits in the problem were constrained so that the stated answer or the unstated interitem product included the fraction 0.5, e.g., [(2.5 × 2) + 1.5 = 6.5], or, [(7/2) − 2.5 = 1]. The final answers were also constrained to be between 0.5 and 9.5, with half correct and half incorrect by 1 or 2.

Manipulation check on operation difficulty. Similar to the sentence verification study, an operation verification study was conducted to test whether subjects differentially processed the different types of operation strings. All operation strings were randomly presented on a computer monitor to 24 subjects whose task was to verify whether each of the stated answers were correct. Subjects were given reaction time and accuracy feedback after each trial and were instructed to maintain 90% accuracy and urged to slow down when making too many errors. Subjects were most accurate when verifying simple operations ($\bar{x}$ = 92.78% correct), less accurate when verifying moderate operations ($\bar{x}$ = 86.2% correct), and least accurate when verifying difficult operations ($\bar{x}$ = 65.97%). Subjects were fastest at verifying simple answers ($\bar{x}$ = 1977 ms), slower at verifying moderate answers ($\bar{x}$ = 2563 ms) and slowest at verifying difficult answers ($\bar{x}$ = 3459 ms).

General procedure

The procedure for administering the six complex WM span tasks was similar to that described in Experiment 1, except subjects were tested one at a time with the series of stimuli presented on an IBM microcomputer. The number of sentences (in the

three Sentence Word tasks) or operations (in the three Operations Word tasks) within each 3-trial series was gradually increased from 2 to 5 for the moderate and difficult stimuli levels, and from 2 to 6 for the simple stimuli levels. Specifically, each subject read the visually presented unrelated sentences (Sentence Word tasks) aloud and verified whether the sentence made sense on their answer sheet. Immediately after a subject finished reading a sentence aloud the experimenter pressed the return key causing the next sentence to appear in the series. The elapsed time each subject required to read and verify each sentence was recorded. When a series of sentences was completed (i.e., one trial) a question mark appeared at which time the subject wrote, in any order, the last words of the sentences just read. Similarly, in the three Operations Word tasks, each subject read aloud a series of operations and to-be-remembered words, verifying whether each stated answer was correct. Immediately after a subject finished reading and verifying an operation string the experimenter pressed the return key and the next string appeared. When a series of operation strings was completed the question mark appeared which cued the subject to recall the words in any order. No limit was placed on the time allowed for the computations, and the time taken to read and verify the problem was recorded. Subjects were warned that the number of stimuli in each trial would increase as they progressed through the series of each task.

Results Experiment 2

The data for this experiment consisted of scores obtained for each subject from 36 measures of WM span, reading comprehension, and the Verbal and Quantitative components of the Scholastic Achievement Test. Table 6 shows the mean, standard deviation, minimum, and maximum for the span measures and the reading comprehension measures. In each of the six complex tasks there were two measures of the processing component of the task: (1) mean response time to complete reading and verifying the item and (2) accuracy stated in proportion of correct verifications. Two dependent measures of span were also recorded for each task: (1) an absolute memory span consisting of the total number of correctly recalled words and (2) the memory span consisting of the highest set size in which memory items were correctly recalled from two of the three memory trials

TABLE 6

Descriptive Statistics for WM Total Score

	Sentence			Operation		
	Simple	moderate	Difficult	Simple	moderate	Difficult
Mean	42.1	29.9	26.8	46.5	32.3	29.0
% recall	70	71	64	78	77	69
SD	5.40	3.78	5.04	4.80	4.20	5.04
Mean span	3.2	2.9	2.4	3.8	3.1	2.5
Minimum	30	20	17	31	18	18
Maximum	55	40	37	56	39	39

Descriptive Statistics for Comprehension and SAT Scores

	Nelson–Denny			
	Number correct	Reading rate	Verbal SAT score	Quantitative SAT score
Mean	26.7	289.5	443.4	479.6
SD	5.45	85.6	406.0	107.7
Minimum	12	170	250	270
Maximum	35	544	680	690

given at each set size. As in Experiment 1, the analyses of total WM spans are reported.

Verification statistics. The level of accuracy in the verification component of the complex span tasks was generally high with not much difference across the difficulty conditions. The sentence verification accuracies were 95, 97, and 94% for simple, moderate, and difficult sentences, respectively. The operation verification accuracies were 98, 97, and 87% for simple, moderate, and difficult operations, respectively. On the other hand, response time did vary considerably across the difficulty condition. The mean response times were 2.80, 4.28, and 5.28 s per sentence for simple, moderate, and difficult sentences. For the operations, response times were 3.45, 6.01, and 11.14 s for simple, moderate, and difficult operations.

Reading comprehension. This study was directed at the question of why reading comprehension was predicted by the complex WM spans (consisting of simultaneous processing and span tasks), but not by the simple word or digit spans (consisting of only the span task). The processing component of the complex WM span may simply prevent the use of memory strategies and thereby generate a "purer" measure of WM capacity, which then better predicts comprehension than the simple span. If the use of memory strategies are inhibited by using a verification task while measuring span, then as the difficulty of the sentence (Sentence Word) and operation (Operations Word) verification tasks is increased, the correlation between WM span and comprehension ought to increase up to a point. Table 7 shows the Pearson Product Moment correlations calculated between the simple, moderate, and difficult complex span tasks and comprehension measures that are central to this question.

The level of difficulty influenced the correlation between WM span and reading comprehension, whether the span was measured with the Sentence Word or the Operations Word complex span task. The moderate Sentence Word/ND correlation, $r(48) = .46$, $p < .001$, was larger in magnitude than the simple Sentence Word/ND correlation, $r(48) = .36$, $p < .01$ This difference between moderate and simple correlations (.46 and .36) was significant, $t(48) = 1.99$, $p < .05$. In addition, the moderate Sentence Word/ND correlation (.46) was larger than the difficult Sentence Word/ND correlation, $r(48) = .26$, $p > .07$ ($t(48) = 3.16$, $p < .01$). Similarly, the moderate Operations Word/ND correlation, $r(48) = .32$, $p < .05$, tended to be larger in magnitude than the simple Operations Word/ND correlation, $r(48) = .28$, $p < .03$ ($t(48) = .88$, $p > .10$), and was larger than the difficult Operations Word/ND correlation, $r(48) = .23$, $p > .11$ ($t(48) = 2.06$, $p < .025$).

It is important to note that the correlations between comprehension and Sentence Word span and those between comprehension and Operations Word span behaved similarly across levels of difficulty. One would not expect these correlations to vary in the same systematic manner, if the complex spans reflecting WM are consid-

TABLE 7
CORRELATIONS BETWEEN SPANS AND COMPREHENSION MEASURES

Comprehension measures	Sentence			Operation		
	Simple	moderate	Difficult	Simple	moderate	Difficult
ND	.37*	.47**	.26	.28*	.32*	.23
VSAT	.25	.30*	.27	.18	.11	.07
QSAT	.04	.08	.08	.10	.18	.10

* $r(49)$, $p < .05$ for span and ND correlations.
** $r(47)$, $p < .001$ for span and VSAT correlations.

ered to depend on the use of reading skills in the processing component of the span task. That is, if reading skills determine the Sentence Word span, then as the sentence verification task using those skills is made more difficult, the resulting Sentence Word span/comprehension correlations should increase in magnitude, as they did, up to that point where the difficulty level was very demanding. However, the Operations Word/ND correlations show a similar pattern across levels of verification task difficulty. And, since the operation verification task (i.e., the processing portion of the Operations Word span task) clearly did not require reading skills, a reading skill or task-dependent explanation simply cannot account for the Operations Word/ND correlations varying in the same manner as the Sentence Word/ND correlations. The pattern of Sentence Word and VSAT correlations across different difficulty levels of span was somewhat similar to the inverted U-shaped patterns for the Sentence Word/ND and Operations Word/ND correlations. The moderate Sentence Word/VSAT correlation, $r(48) = .33$, $p < .05$, tended to be larger in magnitude than the simple Sentence Word/VSAT correlation, $r(48) = .29$, $p < .10$, and also tended to be larger than the difficult Sentence Word/VSAT correlation, $r(48) = .27$, $p < .10$. However, the differences between moderate and simple Sentence Word/VSAT correlations were not significant, $t(48) = 0.69$, $p > .10$, and differences between moderate and difficult Sentence Word/VSAT correlations only approached significance ($t(48) = 1.34$, $p > .07$). There were no other significant correlations between Sentence Word and Operations Word with VSAT and QSAT scores. This included the relationship between Operations Word and VSAT which was significant in Experiment 1. We can only point out that all the correlations were somewhat lower in this study and that all those correlations involving QSAT disappeared in this study.

Analysis of data from the verification task. The top part of Table 8 shows the correlations between the accuracy of the verification task and Nelson–Denny, VSAT and QSAT as a function of the difficulty of the verification task. The bottom part of Table 8 shows these correlations for the verification response time. In general, both sets of correlations suggest a rather small relationship between verification accuracy and/or speed with the global ability measures. There were only 4 significant correlations with accuracy and 7 with response time. However, 7 of the 11 significant correlations were for the moderate level of difficulty (4 for sentence verification and 3 for operation verification).

It is notable that the same general in-

TABLE 8

Test	Sentence			Operation		
	Simple	Moderate	Difficult	Simple	Moderate	Difficult
Correlations between Accuracy of Verification and Ability Measures						
ND	.15	.28*	.08	.05	.16	−.03
VSAT	.10	.30*	.12	−.01	.26	.19
QSAT	.08	.24	.10	.06	.32*	.31
Correlations between Verification Response Time and Ability Measures						
ND	−.21	−.29*	−.21	−.22	−.25	−.17
VSAT	−.30*	−.33*	−.26	−.34	−.43**	−.25
QSAT	−.27	−.18	−.21	−.21	−.37*	−.29

* $r(49)$, $p < .05$.
** $r(49)$, $p < .002$.

verted U-shaped pattern of verification/comprehension correlations occurred across the three levels of difficulty that had occurred with the span/comprehension correlations. This pattern was found for 11 of the 12 sets of correlations.

An analysis was also performed on the relationship between the verification performance and the number of words recalled in the complex span tasks as a function of difficulty. Only 7 of 72 possible correlations were significant and these fit no noticeable pattern.

Span-verification trade-off. The moderate difficulty level resulted in Sentence Word and Operations Word span distributions that were similar to those of earlier studies. Therefore, based on their total recall at the moderate difficulty level of the Sentence Word task, subjects' spans were categorized into high, medium, and low Sentence Word span groups. Also, high, medium, and low Operations Word span groups were similarly formed, categorized according to their recall in the moderate Operations Word task. High span groups consisted of those subjects whose spans were one standard deviation or greater above the mean of each sample distribution, and low span groups contained subjects whose spans were one standard deviation or greater below the mean of each sample distribution. Since the moderate Sentence Word task had a sample distribution mean of 29.2 and standard deviation of 3.78 (see Table 6), this criterion produced a high Sentence Word span group with 14 subjects who had spans above 33.68, a low Sentence Word span group with 10 subjects having spans below 26.12, and a medium Sentence Word span group of 26 subjects having spans between 26.12 and 33.68.

Since the moderate Operations Word task had a sample mean of 32.3 and standard deviation of 4.2, 13 subjects were categorized into the high Operations Word span group whose spans were above 36.5, 12 subjects into the low Operations Word span group whose spans were below 28.1, and 25 subjects into the middle operations Word span group whose spans were between 28.1 and 36.5. Within each of these groups (high, medium, and low Sentence Word and Operations Word span groups) recall for each set size was then examined at each level of difficulty.

The top part of Fig. 1 shows the mean number of words recalled by subjects in the high, medium, and low groups at each set size as a function of the difficulty of the Sentence Word verification task. The bottom part of Fig. 1 shows these words for the Operations Word verification task.

The finding from this analysis is clear. Span performance of low span subjects was not affected by the difficulty manipulation, but high span subjects showed a significant decrease in recall from the moderate to the difficult tasks when they were at the limit of their span capacity. Figure 1 shows that the recall of subjects in the high Sentence Word span group decreased from 35 in the moderate task to 25 in the difficult task at set size 5, $F(1,26) = 7.75$, $p < .01$; and decreased from 25 in the moderate task to 18 in the difficult task at set size 4, $F(1,26) = 4.38$, $p < .05$. Recall of subjects in the high Operations Word span group decreased from 37 in the moderate task to 26 in the difficult task at set size 5, $F(1,24) = 6.68$, $p < .03$; and from 25 in the moderate task to 19 in the difficult task at set size 4, $F(1,24) = 4.30$, $p < .05$. On the other hand, low span subjects did not show any significant differences in Sentence Word or Operations Word recall as a function of the difficulty level of the tasks at any set size. However, at set size 5, well above these subjects' spans, there is a trend toward a decrease from 29 in the easy to 24 in the moderate Sentence Word tasks, $F(1,18) = 4.01$, $p < .07$, and from 28 in the easy to 24 in the moderate Operations Word tasks, $F(1,22) = 3.12$, $p < .11$. These results are similar to those reported by Carpenter and Just (1988) with mixtures of easy and difficult sentences.

Finding a difference in the recall of high

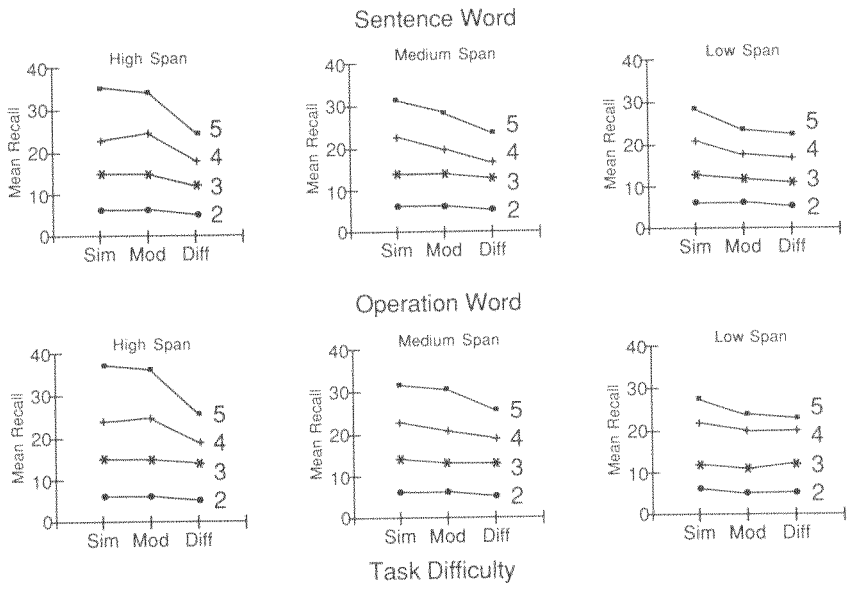

FIG. 1. Mean recall as a function of span task, set size, difficulty of the background component, and whether the subject was high, medium, or low span.

but not low span groups as a function of difficulty suggests that subjects in the high span groups are allocating more processing time than low span groups as difficulty level increases. If so, then the response time for the verification tasks should reflect this difference. However, this was not the case. In high, medium, and low span groups the response time to verify sentences and operations increased in a similar fashion as the level of difficulty increased. The high span subjects in the Sentence Word task showed an increase in response time from 2.8 to 5.5 s (from easy to difficult conditions) and the low span subjects showed an increase from 2.9 to 5.8 s. The corresponding increase in the Operations Word task was from 3.1 to 8.5 s for high span subjects and from 3.5 to 9.8 s for the low span subjects.

For both Sentence Word and Operation Word tasks, span declined with increasing difficulty. This suggests a possible trade-off in processing between the background task and the span task. Are good readers more or less likely to trade off performance in the span for the increasingly difficult background task? This was tested by calculating a difference score between the Simple and Difficult conditions for each subject and testing for the correlation between this score and the reading comprehension measures. The correlations were all nonsignificant and near zero. For the Sentence Word task, the correlations were .02 with the Nelson–Denny, .03 with the VSAT, and .01 with the QSAT. For the Operation Word task, the correlations were −.01 with the Nelson–Denny, .17 with the VSAT, and .06 with the QSAT. This suggests that the good and poor readers did not differentially trade off between background task and span task.

Span intercorrelations. If the Sentence Word and Operations Word complex spans are reflecting the same underlying WM capacity, then they ought to intercorrelate. Table 9 shows that the Sentence Word and Operations Word spans, at least across the simple and moderate levels of difficulty, correlated significantly. Only 3 of the 15 possible correlations were not significant and all involved a difficult task as a member of the pair.

GENERAL DISCUSSION

The primary question motivating this

TABLE 9
CORRELATIONS AMONG COMPLEX MEMORY SPANS

	Sentence			Operation		
	Simple	Moderate	Difficult	Simple	Moderate	Difficult
Sentence Word						
Simple	—					
Moderate	.52***	—				
Difficult	.37**	.27	—			
Operations Word						
Simple	.61***	.48**	.42**	—		
Moderate	.50***	.46**	.15	.68***	—	
Difficult	.55***	.45**	.22	.65***	.71***	—

* $p < .05$.
** $p < .005$.
*** $p < .0001$.

study was whether the correlation observed between the complex memory span and reading comprehension is a result of a rather specific interaction between particular skills of the individual and the task being performed at the moment or the result of a relatively immutable capacity which transcends the specific task. If memory span predicts reading comprehension when measured against a background of both reading- and arithmetic-related tasks, then the measure can be considered to be independent of specific task skills. Experiments 1 and 2 clearly demonstrated that good readers remembered more words and digits than poor readers, regardless of whether the background task required reading or arithmetic skills. The Sentence Word, Sentence Digit, Operations Word, and Operations Digit spans in Experiment 1, and the moderate Sentence Word and Operations Word spans in Experiment 2 all predicted reading comprehension.

One possible explanation of these findings is that good readers also have good quantitative skills leading to a spurious correlation between the operations tasks and reading comprehension. However, the relationship between Operations Word and reading comprehension was still present when quantitative skills measured by the Quantitative SAT were factored out of the association. Thus, a complex span reflecting WM capacity does not need to be "reading" related to generate a significant correlation with reading comprehension. In fact, the Operations Word task contributed significantly to the comprehension variation even when entered into the linear regression after the Sentence Word task.

Experiment 1 demonstrated the absence of a significant correlation between reading comprehension and the simple digit and word spans. This led to the question of why the complex spans, but not the simple spans, predicted reading comprehension. Experiment 2 considered the possibility that the complex span is a more accurate measure of WM capacity because it prevents the use of memory strategies, such as grouping and rote rehearsal. If so, then varying the difficulty level of the background task, against which the complex span is measured should affect the complex span/comprehension correlations. That is exactly what was found in Experiment 2. The Sentence Word/comprehension and Operations Word/comprehension correlations were a function of the different levels of difficulty. Further, it is important to note that these two sets of correlations exhibited similar inverted U-shaped functions. That is, when the difficulty level of the reading-related (Sentence Word) or arithmetic-related (Operations Word) background tasks was moderate, the span/comprehension correlations were higher in magnitude than when the background tasks were

very simple, or, were very difficult. The similarity of the Operations Word and Sentence Word correlational patterns further supports the argument that the tie between the complex span and comprehension is at least somewhat independent of the type of background.

What is not clear from our results is just how far we can push this generalization about the independence of the nature of the background task and the criterion task. Working memory may be a unitary individual characteristic, independent of the nature of the task in which the individual makes use of it.

On the other hand, it remains possible that the complex span task must make use of a verbal, if not "reading," background task in order for the span to predict comprehension. The operations task, while not "reading" as we normally think of it, certainly makes use of a verbal code for the numbers and operations. This would fit with research and theorizing on attentional resources by Wickens (1984) who proposes that attentional resources for verbal and spatial tasks are independent. It may be the case that working memory capacity for an individual is not peculiar to a particular task but may be peculiar for either tasks that are generally either verbal or spatial. Another possibility is that only verbal tasks will prevent verbal rehearsal and only spatial tasks will prevent spatial rehearsal. Both of these ideas would seem compatible with Wickens views on resource allocation.

Results that may support this general view of working memory were recently reported by Daneman and Tardiff (1987). They used three different tasks for their 36 subjects: a verbal parsing task with the last word of each parsing trial remembered as the span item, a number division task with the resultant multidigit answer for each trial as the span item, and a 3-dimensional tic tac toe game requiring the use of visual imagery with the resultant winning path as the span item. Daneman and Tardiff found that both the *word span* and the *digit span* correlated with reading comprehension but the *imagery span* did not.

This result may reflect separate verbal and spatial working memory processes but there are several reasons to be cautious about the findings. For one, the three tasks seem to vary greatly in their general difficulty, both in the processing component and in the span component and, as we have seen, the difficulty of the background task will affect the relationship between span and criterion performance. Second, the nature of the background task and the nature of the to-be-remembered span were confounded. Using the multiple task procedure to answer theoretical questions requires that one of the components of the compound task be held constant while the nature of the other is varied. Further, if the findings of Experiment 1 on the differences between words and digits as span items are correct, then the failure to find significant correlations for the imagery task could be simply because of the nature of the span items.

Why then do complex spans, but not simple spans, predict reading comprehension? One possibility is that performing the background task against which the complex span is measured, limits the subjects ability to use memory strategies such as rehearsal or grouping of the to-be-remembered information. This is not far from the view of working memory and short-term memory taken by Anderson (1983). He views working memory as "the temporary knowledge structures currently being attended and the active parts of long-term memory," page 118. Word and digit span, as traditionally measured, reflect a combination of those few elements kept active above threshold without the benefit of rehearsal *plus* those elements maintained through rehearsal. From this analysis, working memory consists of the subset of all knowledge structures that are activated above some ambient baseline and short-term memory is the subset of those elements that are activated above some threshold and can thus be

maintained without rehearsal. The threshold corresponds to what we typically think of as conscious awareness. It may be that the background task in the complex span prevents the use of those helpful memory strategies commonly used in the simple span task. The relatively slow rate of presentation typically used in memory span studies certainly allows rehearsal and grouping strategies to be used by nearly all adult subjects to circumvent the capacity limitations of short-term memory. The background task in the complex span tasks makes it more difficult to use these strategies and thus gives a clearer picture of the "true" capacity of short-term memory.

Cohen and Sandberg's (1977) findings support the notion that any measure reflecting the WM limitation important in higher level cognition must dissociate the influence of memory strategies from the span measure. They demonstrated a relationship between IQ and the recency portion of a digit-recall function. In one experiment they gave their subjects a probed serial recall task in which subjects were presented with auditory or visual lists of nine digits. They also varied whether the digits were presented at 1 digit/s or 4 digits/s. After presentation, the subjects received a cue to serially recall the first, middle, or last three digits in the list. And, as expected, they found that digits were recalled better at the beginning and the end of the lists, regardless of the rate of presentation. They then tested the relationship between IQ and recall from each of the three portions of the list. The recency portion of the recall function was the only one to show reliable correlations with IQ, with low IQ subjects showing a greater deficit than high IQ subjects on items at the end of the lists of digits. Further, the most reliable correlations occurred between IQ and the recency portion of the recall function when the presentation rate was 4 digits/s, too fast to allow much rehearsal. However, when they tested the correlation between IQ and the primacy recall, they found NO systematic relationship. They suggested that items at the beginning of the list are represented by an empty STM and that capacity was then available for the use of rehearsal strategies. On the other hand, when the items at the end of the list occur, the subject has already allocated existing resources and this prevents the use of grouping, rehearsal, and other memory strategies.

In a later paper, Cohen and Sandberg (1980) included the above data in a factor analysis, from which they argued that the most critical factors determining a relationship between their subjects' STM capacity measures and IQ are whether items are encoded in order and whether STM is empty or partially filled at the time of encoding the to-be-remembered items. The important point is that the primacy portion of Cohen and Sandberg's probed serial recall tasks, remembering items from the beginning of a list, can be compared to the simple digit and word span tasks used in Experiment 1, wherein capacity was also available to rehearse the items. And, the recency portion of Cohen and Sandberg's tasks can be compared to the complex spans used in Experiments 1 and 2, in which the background task prevents subjects from using memory strategies to increase the number of items recalled. The implication is that the span measured while the subject is performing any rehearsal preventative task would more accurately reflect the capacity of the buffer.

Why, then, does the size of the correlations between the complex spans and comprehension appear to depend on whether the items to-be-remembered are digits or words? Experiment 1 demonstrated the magnitude of the correlations between comprehension and the complex word spans (Sentence Word and Operations Word) to be nearly twice the magnitude of those between comprehension and the complex digit spans (Sentence Digit and Operations Digit). Since the same background tasks were used (i.e., verification of sentences in Sentence Word and Sentence Digit and verification of operation strings in

Operations Word and Operations Digit), differences in the to-be-remembered items appear to be crucial. This finding was unexpected and any account is decidedly post hoc in nature.

One possibility may be that memory strategies are more available or more automated for digits than for words. Intuitively, it would seem that opportunities to apply grouping strategies to a list of unrelated digits occur more often in the real world. And, to the extent that strategies are repeatedly used for processing specific types of information, the strategies become automatic processes. If so, strategies used for grouping digits may be simply more automatic than those used for grouping words, and consequently, would be more difficult to inhibit by the background task.

How do the data fit with existing theories of WM? Baddeley and Hitch (1974) initially formalized the notion of a working memory system, identifying processing and structural components of a limited working memory capacity important in performing higher level cognitive tasks. They proposed that part of the limited capacity in the system is used for processing incoming information (i.e., the central executive), with the remaining resources (i.e., the articulatory loop and visuo-spatial scratch pad) used for storage of the products of that processing. While processing information, the central executive is considered a controller, selecting the most beneficial strategies for integrating information from several different sources. As such, the central executive is used for decision making and controlling the amount of resources to be allocated among the requirements of the ongoing information processing (Baddeley, 1986).

If the complex span is reflecting a WM system, such as proposed by Baddeley and Hitch (1974), then which part of WM is primarily responsible for the background task? Further, which component does the complex span measure? To restate this question, *are the complex spans reflecting individual differences in the central executive, or are they reflecting differences in the articulatory loop, or some interaction of the two?* Baddeley and his colleagues (Baddeley, 1986) have argued that the articulatory loop is a time based system, which plays an important role in the performance of the simple span task. Recent unpublished work from our lab (LaPointe & Engle, 1988) shows that word length, a variable thought to reflect the articulatory loop, leads to the same effect in complex span tasks as in the simple word span. Span scores are higher with short words than with long words in both simple and complex word span tasks suggesting that time-based articulatory coding is important to both. It is possible that the background task prevents grouping and other rehearsal strategies that are irrelevant to reading while leaving the articulatory coding intact. Rehearsal and grouping, as used in memory tasks, are not typical reading skills but the use of articulatory coding probably is important, even if not essential for decoding the individual word or phrase. Whether articulatory coding must occur in the complex span tasks for the span scores to predict reading comprehension is not yet clear but we hope to be able to answer that question soon.

Another question that must be left hanging is why the Operations Word task added to the variance accounted for in the linear regression above that accounted for by the Sentence Word task. One theoretically uninteresting possibility is that the Sentence Word task in Experiment 1 was simply more difficult than was the Operations Word task and we see from Experiment 2 that difficulty level is important. The Sentence Word task required the subject to verify the correctness of sentences. This could be done easily but the reading aloud of the rearranged sentences was difficult and confusing to our subjects. This possibly made the overall task more difficult. This is supported by Table 1 which shows that the mean span for the Sentence Word task was lower and the standard deviation larger

than for the Operations Word task. But for now, we just do not have a solid explanation of this finding.

REFERENCES

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

BADDELEY, A. D. (1986). *Working memory*. Oxford Psychology Series Vol. 11. Oxford: Oxford Univ. Press (Clarendon).

BADDELEY, A. D., & HITCH, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

BADDELEY, A. D., & LIEBERMAN, K. (1980). Spatial working memory. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp 521–530). Hillsdale, NJ: Erlbaum.

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.

BRAINERD, C. J., & KINGMA, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. *Developmental Review*, **4**, 311–377.

BRAINERD, C. J., & KINGMA, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, **17**, 210–247.

BROADBENT, D. E. (1958). *Perception and communication*. Oxford: Oxford University Press.

CARPENTER, P. A., & JUST, M. A. (1988). The role of working memory in language comprehension, In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Erlbaum.

CASE, R. (1974). Mental strategies, mental capacity, and instruction: A neo-Piageitan investigation. *Journal of Experimental Child Psychology*, **133**, 386–404.

CHASE, W. G., & ERICSSON, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16). New York: Academic Press.

COHEN, R. L., & SANDBERG, T. (1977). Relation between intelligence and short term memory. *Cognitive Psychology*, **9**, 534–554.

COHEN, R. L., & SANDBERG, T. (1980). Intelligence and short-term memory: A clandestine relationship. *Intelligence*, **4**, 319–331.

COLTHEART, V. (1972). The effects of acoustic and semantic similarity on concept identification. *Quarterly Journal of Experimental Psychology*, **24**, 55–65.

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

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

DANEMAN, M., & GREEN, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory & Language*, **25**, 1–18.

DANEMAN, M., & TARDIF, T. (1987). Working memory and reading skill re-examined. In M. Coltheart (Ed.), *Attention and Performance XII*, London: Erlbaum.

FRANCIS, W. N., & KUCERA, H. (1982). *Frequency Analysis of English Language*. Boston, MA: Houghton Mifflin.

HITCH, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, **10**, 302–323.

HOWARD, D. (1983). *Cognitive Psychology: Memory, Language, and Thought*. New York: Macmillan.

JUST, M. A., & CARPENTER, P. A. (1987). *The psychology of reading and language comprehension*. Newton, MA: Allyn & Bacon.

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

KLAPP, S. T., MARSHBURN, E. A., & LESTER, P. T. (1983). Short term memory does not involve the "working memory" of intellectual processing: The demise of a common assumption. *Journal of Experimental Psychology: General*, **112**, 240–264.

LAPOINTE, L., & ENGLE, R. W. (1988). [Word length effects in complex and simple memory spans.] Unpublished manuscript.

MARASCUILO, L. A., & LEVIN, J. R. (1983). *Multivariate Statistics in the Social Sciences: A Researcher's Guide*. Monterey, CA: Brooks/Cole.

MARTIN, M. (1978). Memory span as a measure of individual differences in memory capacity. *Memory & Cognition*, **6**, 194–198.

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.

MILLER, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81–97.

MILLER, G. A., & MCKEAN, K. O. (1964). A chronometric study of some relations between sentences.

Quarterly Journal of Experimental Psychology, **16**, 297–308.

NELSON, M. S., & DENNY, E. D. (1973). *The Nelson–Denny Reading Test*. Boston: Houghton Mifflin.

NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice–Hall.

PERFETTI, C. A., & LESGOLD, A. M. (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.

SALAME, J., & 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.

SCHNEIDER, W., & DETWELLER, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21). New York: Academic Press.

SHANNON, C. E., & WEAVER, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

TURNER, M. L., & ENGLE, R. W. (1986). Working memory. *Proceedings of the Human Factors Society*, **30**, 1273–1277.

VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

WAUGH, N. C., & NORMAN, D. A. (1965). Primary memory. *Psychological Review*, **72**, 89–104.

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

WICKENS, C. D. (1984). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, NJ: Erlbaum.

(Received February 17, 1988)

(Revision received September 28, 1988)