Individual Differences in Working Memory and Comprehension: A Test of Four Hypotheses

Randall W. Engle, Judy Cantor, and Julie J. Carullo
University of South Carolina

A relationship has consistently been found between measures of working memory and reading comprehension. Four hypotheses for this relationship were tested in 3 experiments. In the first 2 experiments, a moving window procedure was used to present the operation–word and reading span tasks. High- and low-span subjects did not differentially trade off time on the elements of the tasks and the to-be-remembered word. Furthermore, the correlation between span and comprehension was undiminished when the viewing times were partialed out. Experiment 3 compared a traditional experimenter-paced simple word-span and a subject-paced span in their relationship with comprehension. The experimenter-paced word-span correlated with comprehension but the subject-paced span did not. The results of all 3 experiments support a general capacity explanation for the relationship between working memory and comprehension.

The idea that there are limits on attentional resources or immediate memory capacity is an old one. Nearly a century ago, James Mark Baldwin (1894) proposed *attention span* as the maximum number of mental elements to which a child can attend at any one time. He assumed that attention span is limited by neurological factors and changes, with age, as a result of physical development in the brain. The rate of this physical change sets a limit on the rate of the child's intellectual development. Piaget's (1926) *field of centration* and Pascual-Leone's (1970) *mental power* (or *M-Space*) appear to be different names for the same construct. Pascual-Leone, like Baldwin, assumed that M-Space increased with age as a result of changes in the brain that are determined by biological or epigenetic factors. He argued that keeping schemes active requires mental energy, and the amount of energy (i.e., mental power) increases developmentally. Thus, the number of schemes that a person can keep active increases developmentally.

More recently, Case (1974, 1985) proposed a theory of the development of working memory that built on these concepts from Baldwin (1894), Piaget (1926), and Pascual-Leone (1970). He suggested that much of the intellectual change observed as a child matures is a direct consequence of the child's ability to process more information in a given period of time. Case argued that

> maximum mental effort or M-power is also assumed to vary both within and across age groups. Within age groups, differences are assumed to result largely from biological factors, and to be

at least partially responsible for producing differences in what Spearman (1927) labeled "content-free intelligence (*g*)." (1974, p. 548)

According to Case (1974, 1985), mental operations, such as representing ongoing motor movement or representing the current sensory-perceptual state and comparing it with a goal state, all require energy from the fixed pool of total resources, or mental space. Any space not allocated to the operations can then be used for maintaining the products of this mental processing and for representing schemes or knowledge structures retrieved from long-term memory. In this framework, short-term storage space is then defined as the amount of residual space available after the person has allocated that necessary to perform the mental operations required in the task.

Case (1974, 1985) argued that mental operations become faster as a consequence of biological or epigenetic factors in development, occur with less effort, and occur with less drain on the total processing space. As the operations become faster and more efficient, there is more and more storage space available. He argued that the reason digit span nearly doubles in children from 5 to 12 years of age (Engle & Marshall, 1983) is not that the total processing space increases, but that the operations required to encode and combine the digits become faster and less of a drain on the fixed mental resources. This leaves more residual resources for storage. Case seemed to believe that operational efficiency is fairly general, having stated in his 1985 book that "the factor which ties together tests of OS (operation space) and STSS (short term storage space) is not a highly specific one" (p. 371). We call this view the *general processing hypothesis*. Accordingly, both intra-individual developmental changes and within-age-group individual differences in intellectual functioning are the result of a gradual decrease in the proportion of total processing space required for mental operations.

Case's (1974, 1985) view of M-space is also reflected in the adult literature on working memory. The ideas of Baddeley and Hitch (1974), for example, parallel those of Case in

suggesting that working memory is important to both storage and processing (see also Baddeley, 1986). They argued that the central executive of working memory is a flexible work space with limited capacity that is used for processing incoming information and storing the products of that processing. As a result, when greater effort is required to process information, less capacity remains to store the products.

Daneman and Carpenter (1980, 1983) also closely followed Case's (1974, 1985) ideas when describing their own view of adult individual differences in working memory. They hypothesized that working memory is used to represent the strategies and skills used in a complex mental task such as reading, with the remaining capacity used to store the resulting products of the reading-specific operations. They suggested that while information was being processed and the products of this processing stored, the two functions, processing and storage, must compete for the limited capacity in working memory. Furthermore, it was argued that individual differences in reading comprehension are due to readers' differing in the efficiency of their processing skills. Consequently, good readers have more functional working memory capacity for products simply because they have more efficient reading skills. With this framework, it is predicted that individual differences in working memory capacity are peculiar to the particular task being performed at the moment. We therefore call the approach taken by Daneman and Carpenter the *task-specific hypothesis.*

To test the task-specific view, Daneman and Carpenter (1980) used a measure of working memory capacity that required both processing and storage in working memory. This "reading-span" task was really two tasks performed concurrently. One required subjects to read series of unrelated sentences. It was assumed that processing the sentences would limit the storage capacity of working memory. The second task, performed concurrently with the first, was to maintain the last word of each sentence in the series. Subsequent recall of these words would then reflect the amount of residual capacity that was available after meeting the demands for processing. The number of last words recalled against the background of reading was defined as the subjects' span. According to Daneman and Carpenter, this measure reflects individual working memory storage capacity, specific to reading situations. Therefore, the performance on this task should strongly predict performance on other reading-related tasks. They found that the reading-span measure correlated highly with three different reading comprehension measures: answering fact questions ($r = .72$), pronoun reference questions ($r = .90$), and the Verbal Scholastic Aptitude Test (VSAT; Experiment 1, $r = .59$; Experiment 2, $r = .49$). However, a simple word-span test given to the same subjects did not significantly correlate with any of the reading comprehension measures. In two later studies, Daneman and Carpenter (1983) found significant correlations of .46 and .58 between reading span and VSAT.

Another attempt to demonstrate processing efficiency differences in working memory was reported by Daneman and Tardiff (1987). They argued that task-specific processing skill, not storage, is the real cause of individual differences in working memory capacity and the concomitant correlation

with comprehension. Therefore, individual differences in working memory can be tested by measuring processing efficiency directly, even when simultaneous storage is not required. They studied the relation between verbal abilities and three different span tasks: verbal span, math span, and spatial span. Each task had both a processing and a storage component. The processing component of the verbal-span task required subjects to view sets of 4 one-syllable words (e.g., *gang sine wed lock*) and to combine the 2 that would make the single word that did not have a syllable boundary between the 2 smaller words. In this case, the word would be *sinewed.* The storage component required the subjects to recall the resulting larger words after two to four sets of words had been presented. The processing of the math span required the subject to view sets of three numbers (e.g., 26 9 72) and to combine the two that would result in a number evenly divisible by three. In this case, that would be 972. The storage component required the subjects to recall the resulting numbers. The processing component of the spatial span required subjects to view sets of cards. Each card contained three 3 × 3 tic-tac-toe grids, with red and blue tokens occupying certain cells. The subjects were to imagine the three grids as representing the top, middle, and bottom layers of a three-dimensional tic-tac-toe game. They were to mentally combine the three layers and locate the winning line by touching the three tokens with their finger. After a set of two to four cards, the storage component of the spatial-span task required the subjects to recall the winning lines for all the cards in the set. Corresponding to the three span tasks were three additional "storage-free" tests that measured performance on the processing component when recall was not required.

Daneman and Tardiff (1987) showed that none of the spatial measures correlated above zero with verbal ability. For both the verbal span and the math span, when storage was required, the storage component correlated with verbal ability (.61 and .51, respectively), as did the processing component (.62 and .33, respectively). The most striking finding, according to Daneman and Tardiff, was that processing performed in the verbal and math storage-free measures correlated with verbal ability as strongly as processing performed when storage was required. As a result, they concluded that individual differences in processing accounts for all the interesting individual differences in comprehension and verbal ability. Furthermore, they argued that the concept of temporary storage was unnecessary in accounting for these differences and, in fact, might "stunt theoretical advancement" (p. 504). Finally, because only the math and verbal tasks predicted verbal ability, Daneman and Tardiff concluded that there are separate working-memory resources for language- and nonlanguage-based information processing.

There are several points about the Daneman and Tardiff (1987) methods and conclusions that concern us. First, the descriptive data indicated that spatial-processing performance was at ceiling with virtually no variability between subjects. This reduces the ability to find any individual differences in their task. They do mention that performance on the spatial task predicted other spatial-criterion measures, but none of these data were presented. We admit the possibility for separate verbal and nonverbal resource pools (cf. Wickens, 1984),

but this view cannot be supported using the data from Daneman and Tardiff.

Second, it is probably not prudent to directly compare the three tasks they used. Turner and Engle (1989) showed that the difficulty of the processing portion of the span task is an important variable driving whether a task predicts comprehension. They found that only moderately difficult tasks, in both the reading- and operations-span tasks, were successful in predicting comprehension (see also Just & Carpenter, 1992). Daneman and Tardiff's (1987) tasks were not equated for difficulty. The accuracy data showed that the spatial processing was easiest. Math-processing accuracy was nearly one standard deviation below spatial processing, and the verbal component was nearly another standard deviation below that. Therefore, one of the reasons that their tasks may have differed in predictive power is because they also differed in task difficulty.

Third, Daneman and Tardiff (1987) made the argument that performance in the processing component of the verbal and math spans accounts for all the variance in verbal ability. This is because the correlations associated with the storage-free tasks were as high for the tasks requiring storage. However, they did not statistically test whether storage-free processing accounted for the same variance in verbal ability as the measures obtained when storage was required. If their hypothesis is correct, then the correlations between verbal ability and the storage and processing components for the task requiring recall should disappear when the effect of the task requiring only processing is removed. Unfortunately, in this case, even if these partial correlations were computed, the results would be suspect because their storage-free verbal and math tasks relied heavily on temporary storage, and for the tasks requiring storage, the span score is totally dependent on the processing task. Subjects could not recall an item correctly in the span task if they had not performed the processing component correctly.

Finally, Daneman and Tardiff's (1987) verbal-processing task placed a very heavy emphasis on word knowledge. Many of the single-syllable words and most of the combined to-be-remembered words were low to very low in frequency of usage. Individuals with poor word knowledge would be unlikely to ever arrive at *sinewed* given *sine* and *wed*, thus, their memory score would necessarily be lower than an individual with more extensive word knowledge. Dixon, LeFevre, and Twilley (1988) and Engle, Nations, and Cantor (1990) showed that word knowledge contributes significantly to the relationship between working memory and comprehension, above that occurring when the effects of word knowledge are minimized. It would not be surprising, therefore, if the verbal-processing component accounted for variance above and beyond that contributed by the math span.

An alternative explanation for the relationship between the working-memory measure and comprehension is that people are good readers because they have a larger working-memory capacity, independent of the specific task being performed. If that is true, the working-memory span should predict higher level tasks requiring working memory for their successful completion, regardless of the background task used with the span test. As such, the memory-span test could be embedded in any secondary processing task and still predict success in higher level tasks. This view sees individual differences in working memory as reflecting a relatively stable characteristic of the subjects. We call this the *general capacity hypothesis* (Cantor, Engle, & Hamilton, 1991; Engle, Nations, & Cantor, 1990; La Pointe & Engle, 1990; Turner & Engle, 1986, 1989).

Turner and Engle (1986, 1989) provided the first evidence for the general capacity hypothesis of working memory. They demonstrated that the relationship between working-memory capacity and comprehension is not specific to reading processes per se. They measured working-memory capacity by using a span task in which the background or processing component was an arithmetic operation, such as "$(4 / 2) - 1 = 1$? SNOW." Subjects read and verified an operation and then read a word after the operation (in this case *SNOW*). After a series of operation–word strings had been presented— for example, $(4 / 2) - 1 = 1$? SNOW, $(3 \times 1) + 4 = 7$? TABLE—the subjects recalled the words that had followed each operation, in this case, *SNOW* and *TABLE*. The number of operation–word strings in a sequence increased, and memory span was the number of words recalled.

The most important finding of this study was that the number of words recalled in the operation–word-span task correlated with reading comprehension as well as did Daneman and Carpenter's (1980, 1983) reading span. Furthermore, the operation–word-span and reading–word-span tasks accounted for the same variance in comprehension. These results clearly failed to support task-specific theory. As predicted by the general capacity theory, the two complex span measures of working-memory capacity effectively predicted individual differences in comprehension independent of the type of processing task in which they were embedded. Similar results have been demonstrated numerous times in subsequent studies (Daneman & Tardiff, 1987; Engle, Nations, & Cantor, 1990; La Pointe & Engle, 1990; Salthouse, Mitchell, Skovronek, & Babcock, 1989).

Turner and Engle (1989), and the replications that followed, clearly demonstrated that the relationship between span and comprehension is not task-specific. However, these studies do not completely eliminate the possibility that processing efficiency at a more general level drives this relationship. For example, Daneman and Tardiff (1987) argued that there are global information-processing parameters that are important to all language-based tasks, including reading and math. Accordingly, it is the efficiency with which people process language or symbolic information that determines working-memory capacity. The reading and operation spans would both reflect this efficiency and therefore account for the same variance in comprehension. With this view, it is still possible that processing efficiency, not storage, mediates the relationship between span and comprehension.

One way of addressing the relationship between processing, storage, and comprehension is to study high- and low-span individuals. Using this approach, Carpenter and Just (1989) studied subjects of high and low reading span to determine whether they allocate their resources differently in the reading-span task. They identified 6 high- and 6 low-span subjects and recorded their eye movements as they performed the span task for many trials. Gaze time in the reading-span task was

assumed to reflect processing time (Just & Carpenter, 1987). They found that high-span subjects spent less time reading the sentences than low-span subjects, suggesting more efficient processing by the high-span subjects. High-span subjects also gazed longer at the end word of each sentence (i.e., the to-be-remembered word) than did low-span subjects. These results supported their contention that processing efficiency is directly related to recall of stored items. High-span subjects were better processors and could therefore store more subsequent information.

In an attempt to identify the specific processing differences between high- and low-span subjects, Carpenter and Just (1989) obtained two measures of processing efficiency for all the words in the sentences, except the span items. The word-encoding parameter was the slope of gaze time as a function of word length, and the lexical-access parameter was the slope of gaze time as a function of word frequency. They found no differences between high- and low-span subjects on the word-encoding parameter. However, high-span subjects spent less time than low-span subjects on lexical access, that is, the time assumed to reflect access to the meaning of a word. From these findings, Carpenter and Just argued that high-span subjects read faster because they accessed the meaning of words faster than low-span subjects and, as a result, had more residual resources to memorize the sentence-final words in a reading-span task than low-span subjects. In other words, the high-span subjects recalled more words because of more skilled processing, not because of greater storage capacity.

The slope of the frequency function, which Carpenter and Just (1989) attributed to lexical access time, also differed depending on whether subjects were reading under memory load. In the reading-span task, no memory load exists when the first sentence of a group of sentences is read. Memory load increases as each subsequent sentence in a group is read, because there are more sentence-final words to be remembered. Carpenter and Just found that low-span subjects showed the same lexical-access parameter under load and no load. They spent more time on low-frequency words than high-frequency words under both load and no-load conditions. Under no load, high-span subjects also gazed longer at low-frequency words than at high-frequency words. However, high-span subjects behaved differently under load. They spent the same amount of time processing low-frequency words and high-frequency words. High-span subjects apparently chose not to allocate the necessary additional resources to the low-frequency words in the sentences and instead chose to allocate those resources to remembering the span words. Carpenter and Just concluded that high-span subjects processed sentences more superficially as memory load increased, whereas low-span subjects continued to process sentences for meaning as memory load increased. They argued that only the high-span subjects altered their processing of words to accommodate the strain that remembering words placed on memory limitations.

One could interpret the Carpenter and Just (1989) data to justify yet a fourth explanation of the relationship between working-memory scores and higher level cognitive tasks, such as reading comprehension. The relationship could occur because subjects vary in how strategic they are in all cognitive tasks. High-span subjects might score high in the working-memory tasks because they more intelligently allocate their resources across the two components of the task to maximize their span scores. These same subjects might also be more likely to use intelligent strategies in reading so their comprehension scores would be higher. In other words, the high-span subjects may just be more planful, and not have access to more working-memory resources per se. We call this fourth view the *strategic allocation hypothesis.*

A caveat about the Carpenter and Just (1989) procedure is necessary. They showed that high- and low-span subjects behave differently while performing the reading span. However, the procedure required for eye-movement analysis dictated that a few subjects were tested for many, many trials on the reading-span task. It is likely that well-practiced subjects adopted processing strategies for performing this task that were different from those used initially. More important, we do not know whether the processing differences observed between high- and low-span subjects mediated the relationship between span and comprehension. To answer this question, they would have needed to statistically remove the individual differences associated with the processing task from the relationship between span and comprehension.

One other feature about this methodology that concerns us is that Just and Carpenter (1987) previously showed that in a normal reading task, subjects tend to spend more time on the last word in the sentence. They interpreted this increase in gaze time as time used to wrap up sentence processing. Thus, because the last word in the sentence was also the span word, it is impossible to say whether the extra time that high-span subjects spent on the to-be-remembered word was due to the word's being rehearsed or extra time spent in comprehending the sentence.

What Is the Relationship Between Processing, Storage, and Verbal Ability?

Several questions can be posed at this point. Do high- and low-working memory span individuals, as measured by the complex span tasks, also differ in their performance on the processing components of those span tasks? Do the individual differences on the processing components of the span tasks account for the relationship between span and comprehension? That is, would the relationship between span and comprehension disappear if the performance on the processing component was statistically removed? Is there a strategic trade-off between the two components of the span task, such that increases in span score are associated with decreased performance on the processing component? If so, would the span–comprehension relationship disappear when the variance due to this trade-off is partialed out of the relationship?

The goal of the present studies was to investigate the triadic relationship between storage, processing, and comprehension in a task in which the speed and accuracy of processing could be measured independently of storage. We specifically wished to test four different hypotheses for the working-memory-span–comprehension relationship: general processing, task-specific processing, strategic allocation, and general capacity. We assumed that both general processing and task-specific

processing hypotheses would be supported if the following events occurred: (a) viewing times (VTs) for processing, even without storage required, correlates with both span and VSAT; (b) VTs for processing with storage also required correlates with both span and VSAT; and (c) the span–VSAT correlation is eliminated when the effects due to VTs are partialed out. Both general processing and task-specific hypotheses would fail to the extent that these findings do not occur. The strategic allocation hypothesis would be supported if (a) a negative correlation occurs between accuracy on the processing component and span; (b) a negative correlation occurs between VTs for the processing component and span; (c) a positive correlation occurs between the VT for the to-be-remembered span word and the span score; (d) low-span subjects would not change the VT on the processing component as memory load increases, but high-span subjects decrease the processing time as load increases; and (e) the span–VSAT correlation is eliminated when the VTs for the processing and the to-be-remembered word are partialed out. The general capacity hypothesis would be supported if (a) the span–VSAT correlation occurs, regardless of the nature of the processing component (either reading or arithmetic), and (b) the span–VSAT correlation remains significant when the VTs for both the processing component and the to-be-remembered word are partialed out.

The first 2 studies presented here consisted of subject-paced, complex, memory-span tasks in which the time to perform the processing portion of the task was measured both with and without the requirement to recall to-be-remembered words. The first study used an operation–word-span task similar to that used by Turner and Engle (1989), and the second study used a reading-span task modeled after that of Daneman and Carpenter (1980). In both studies, the processing component and the to-be-remembered words were presented one element at a time using a moving-window technique. Thus, the time to process each element of the complex tasks could be measured in the normal performance of the span task, with and without recall required.

Results from studies using the moving-window technique to explore reading behavior have yielded results generally consistent with eye-movement data. Just, Carpenter, and Woolley (1982), for example, reported that reading-time data with moving-window presentation share many of the same characteristics as reading-time data when eye movements are recorded. They found that mean time to view a word in a moving-window presentation of text was highly correlated with mean gaze duration. From such findings, it is reasonable to assume that if the time someone spends gazing at words in a text indicates the time spent processing those words, then the time individuals spend viewing each word in a moving-window presentation of text also represents the time subjects spend processing these components. Consequently, the moving-window technique was considered an appropriate procedure for examining processing as subjects performed in complex span tasks.

The operation–word-span task was used in Experiment 1 for several reasons. We have repeatedly shown that the operation–word task correlates with verbal abilities in general and, reading comprehension specifically, as does the reading-span task. In addition, the difficulty of the task is much easier to monitor, and accuracy of processing on the operation–word task is much easier to measure than processing in the reading-span task. Finally, the operation component can be constructed so that the same number of elements is used in each operation, making it easier to compare processing time on the elements of the processing component. This was particularly important because we wanted to compare the processing time on the elements of the operation as memory load increased. We also argue that if the goal is to measure working-memory capacity independent of the criterion task, then the operation span is preferable to the reading-span task. Successful performance of the reading span is much more likely to depend on specific processes common to those used in general reading-comprehension tasks.

Experiment 1

Method

Subjects

Seventy students from introductory psychology classes at the University of South Carolina participated for class credit. Subjects were selected from those students giving permission for the release of their SAT scores. To ensure a wide range of comprehension skills, 10 students were chosen at each of seven intervals from the 200 to the 800 score range for the VSAT. The intervals were defined as 200–350, 351–400, 401–450, 451–500, 501–550, 551–600, and 601–800. All subjects had normal to corrected normal vision and were native English speakers. Subjects were included in the study if their accuracy in the processing portion of the span task remained above 85% correct. Three of the original subjects failed to keep their operation accuracy above 85% and were therefore replaced.

Stimuli and Procedures

All stimuli were presented centered on a VGA monitor of an IBM PS/2 Model 50 computer in 14-point type, printed in white and shown on a black background. All tasks were programmed with Micro Experimental Laboratory, Prerelease System Version 120 (Psychology Software Tools Inc., 1988).

Two lists of common, one-syllable words were generated, one with 20 words and one with 66 words. Examples of words used are *taste, knife,* and *clown.* A pool of 132 operations were constructed to fulfill the following constraints: (a) The first component of the operation included two numbers that were either multiplied or divided. The numbers for the first component ranged from 1 to 20, and the resulting answer ranged from 1 to 10. (b) The second component involved either addition or subtraction of a single-digit number, ranging from 1 to 9. (c) The final answer was also a single-digit number from 1 to 9. An example of the type of operation used is "(6 / 2) − 1 = _____." The operations were similar to those in the moderately difficult condition of the Turner and Engle (1989) study.

Operation–word span task. The span task (i.e., with recall) required subjects to pace themselves through an operation string, solve the operation, and then read an unrelated word. Each time subjects pressed the plus key on the numeric keypad, the next part of an operation was displayed. These operations were shown in five parts: (a) the left parenthesis and a number, (b) a multiplication or division sign, (c) a number and the right parenthesis, (d) an addition or

subtraction sign, and (e) a number from one to nine. The last part of the operation string was followed by "= _____." When this answer blank appeared, subjects typed in their answer using the keys zero through nine on the numeric keypad. After typing in an answer, a word was displayed until the subject pressed the plus key to end the sequence.

Each part of the operation–word sequence was shown where it would have appeared if the entire operation–word sequence had been presented all at once centered on the screen. Viewing times between successive keypresses were collected for each of the seven parts of an operation–word sequence. Therefore, the times spent viewing each of the five parts of an operation, solving it, and reading the word were recorded. Subjects were encouraged to respond as quickly and as accurately to the operations as they could. They spoke each part of an operation aloud as it appeared on the screen, spoke and typed their answer to the operation, and then read aloud the unrelated word that ended the operation–word sequence. Subjects pressed a key as soon as they had read the word, which initiated the first part of a new operation or displayed the prompt "Recall Words." If the recall prompt was presented, subjects attempted written serial recall of the words from the immediately preceding group. They were given as much time as they desired for recall and were encouraged to guess if unsure.

The number of operation–word sequences presented before recall varied from two to six with 3 trials given at each level. Consequently, the span task consisted of 15 trials. The order of presentation of these trials was randomized to prevent subjects from being able to anticipate the number of words they would have to remember on any given trial. Furthermore, each subject received a different random order of operation–word pairs.

Two practice trials of three operation–word sequences were given before the span task began. The experimenter modeled the task on the first practice trial, and the subject performed the task on the second practice trial.

Operation task without recall. For this task, subjects solved 20 numerical operations each of which was followed by an unrelated word. They were instructed not to retain the words. Subjects paced through the operations in the same manner as the span task, but there was, of course, no recall cue. The times spent viewing each part of the operation, answering the problem, and reading the unrelated word were recorded.

Each subject was tested individually on all tasks in a single 1-hr session, first on the operation task without recall then on the operation-span task. An experimenter remained present during the entire session to ensure that task instructions were being followed.

Results

Scoring

Scores on the VSAT and quantitative SAT (QSAT) were used as measures of verbal ability and math ability, respectively. Possible scores on each of these tests range from 200 to 800.

Span measure. A word was considered correct if it were recalled in the appropriate trial, regardless of its position within that trial. Overall span performance was then determined by totaling the number of words subjects recalled from those trials on which all words from the trial were correctly recalled. The maximum span score using this method is 60.

Viewing time. (a) *Operation task without recall:* Mean VTs were computed for each subject for each of the five parts of an operation string, the answer, and also the word. Al-

though subjects received 20 operation–word sequences in the operation task without recall, the initial 4 sequences were treated as practice and those data were omitted from the analyses. (b) *Operation–word span:* Overall mean times for the operation–word-span task with recall were computed for each of the six parts of the operation and for the to-be-remembered word. For each subject, 60 observations contributed to these mean VT scores. (c) *VTs as a function of memory load:* When subjects encountered the first word in a group of operation–word sequences, they were not trying to remember other words, and consequently, memory load is considered to be zero. However, as they encountered each subsequent operation–word string, subjects were trying to remember previously presented words. Thus, memory load increased by a factor of one as each operation–word string was presented.

To determine the effect of memory load on processing strategy, VTs on the span task were analyzed as a function of the number of words the subjects were trying to remember at the moment. Mean time to respond was calculated for each part of an operation–word sequence when it was presented in the first, second, third, fourth, fifth, and sixth position within a trial, regardless of the group size of the trial. Collapsing over group size of the trials led to the mean for each item for each subject, including 15 observations for the first and second operation–word sequences within a trial, 12 for the third position, 9 for the fourth position, 6 for the fifth position, and 3 for the sixth position.

Task accuracy. Accuracy for the storage-free operation task was the percentage of correct answers out of the last 16 problems given. Accuracy for the operation task with recall was the percentage of correct answers to the operations out of the 60 that subjects solved in the span task.

Analyses

It is often the case that Cronbach's (1957) distinction between experimental and correlational methods and analyses extends to the study of individual differences in working memory. One approach is more analysis of variance (ANOVA) oriented. This approach is to identify subjects with high and low scores on complex span tasks, treat them as two independent and homogeneous groups, and see whether the two groups behave differently to one or more task manipulations (Carpenter & Just, 1989; Just & Carpenter, 1992; Whitney, Richie, & Clark, in press). This approach lends itself to studying the effects of different independent variables, but has several deficits. Grouping subjects throws away useful information about the variability of subjects within each group and all those subjects in the middle whose data were discarded. It also is more difficult with this approach to study the contribution of the variables to common and unique variance.

The other approach, more correlation-oriented, is to study subjects over the entire range of working memory capacity and to study relationships among other variables of interest and the extent to which they lead to unique and common variance (Cantor, Engle, & Hamilton, 1991; Daneman & Carpenter, 1980, 1983; Turner & Engle, 1989). We feel these approaches are strongest when used together and, conse-

quently, use them both here to answer different aspects of the central question.

Processing Differences Between High- and Low-Span Subjects

If processing efficiency is an important variable in distinguishing high- from low-span subjects, then these subjects should differ in the speed with which they solve the operations in our tasks. Moreover, if Daneman and Tardiff (1987) were correct in assuming that storage is an unnecessary component in uncovering these differences, then the time to perform an operation when no recall is required should distinguish high- and low-span subjects with the same power as the processing measure when storage is required.

To test this issue, 20 high- and 20 low-span subjects were identified, based on their recall scores from the operation span with recall. Mean VTs (in milliseconds) for these subjects were submitted to a three-way ANOVA, with the variables being task type (recall required or not required), task component (the six parts of the operation plus the to-be-remembered word), and span level (high and low). Any post hoc

analyses were computed with the Tukey Honestly Significant Difference and, where reported, reflect significance at least at the .05 level.

The data are shown in Figure 1, reflecting the three-way interaction of Task Type × Task Component × Span Level, $F(6, 228) = 2.65$, $p < .02$, $MS_e = 35,293.76$. Several points are worth noting. First, in the storage-free task—that is, the operation task without recall—high and low subjects show the same pattern of VT across all components of the task. Even though the high-span subjects appeared to be slightly faster overall, there were no statistically significant differences between high- and low-span subjects at any of these task components. On the operation-span task requiring recall, the high- and low-span subjects were even more similar on the components of the operation task. However, VT for the span word was quite different for the high- and low-span subjects when recall was required. Low-span subjects only increased their VT for the span word by a nonsignificant 112 ms. High-span subjects, however, increased time on the span word by 538 ms when recall was required. This suggests that high-span subjects differentially allocated their processing time across the two tasks, but that low-span subjects did not.
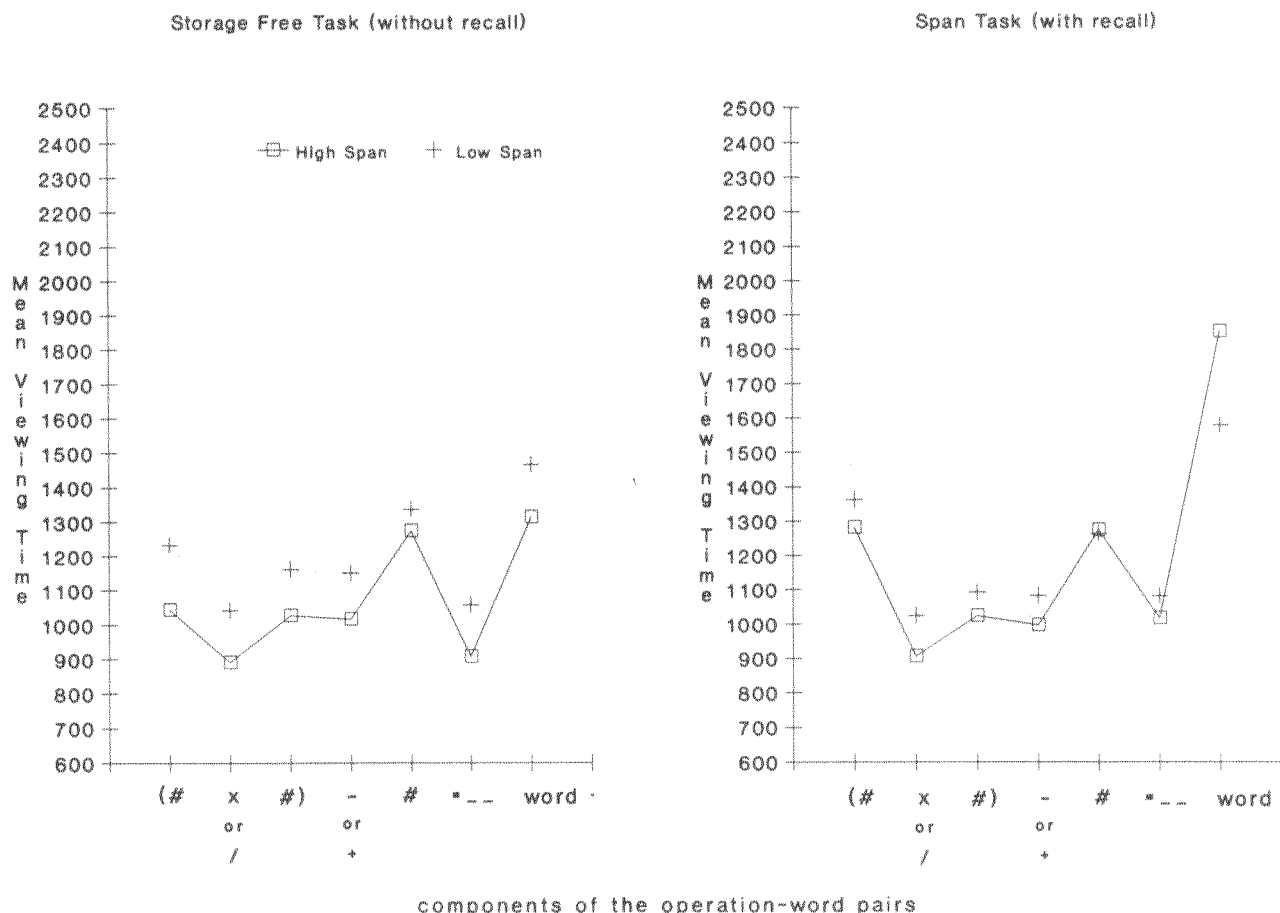

Figure 1. Mean viewing times for low- and high-span subjects in the operation task with and without recall required and as a function of the elements of the operation and the to-be-recalled word in Experiment 1.
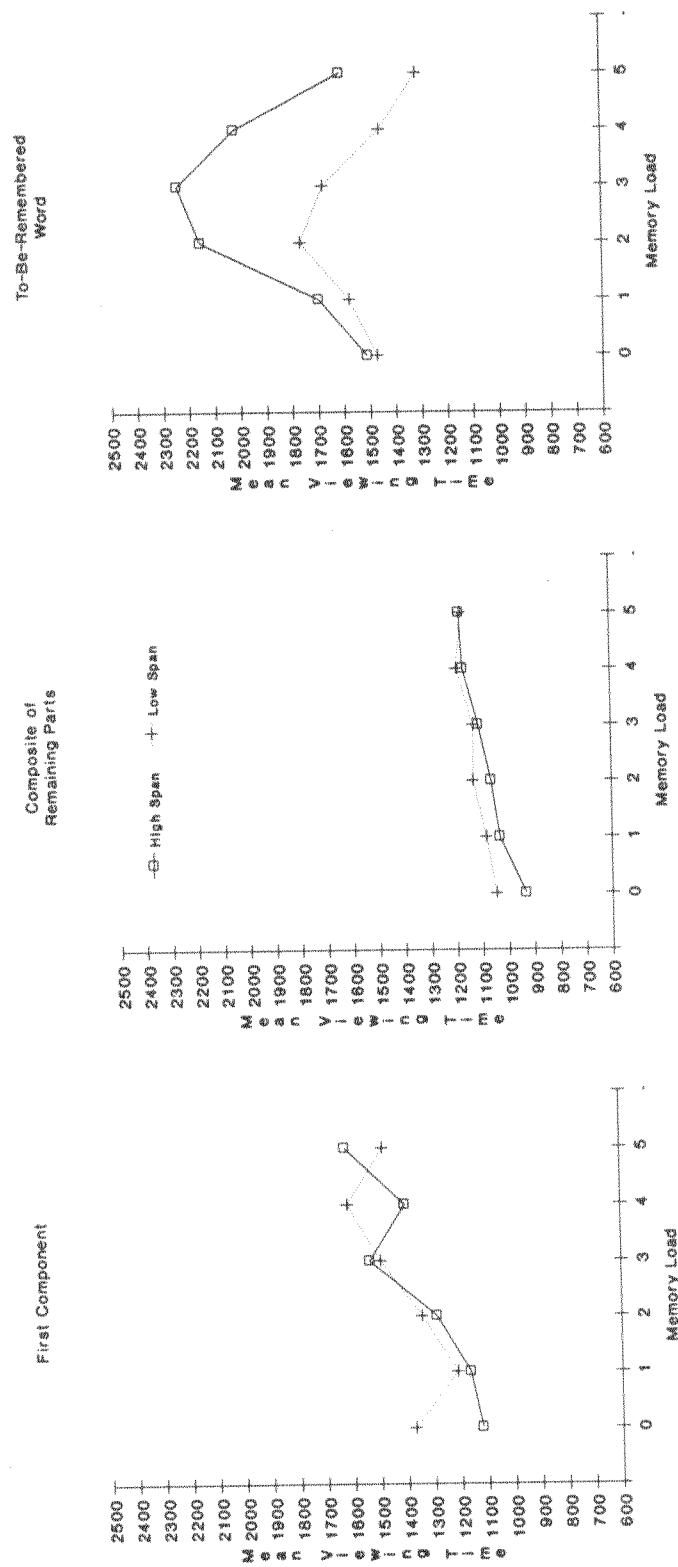
Figure 2. Mean viewing times for the time on the first element of the operation, a composite of Parts 2–6, and the *word* for low- and high-span subjects as a function of memory load in Experiment 1.

In this same light, high-span subjects significantly increased their VTs on the first component of the operation when recall was required, but low-span subjects did not. We suspect that subjects were rehearsing the to-be-remembered word both when the word was present on the screen and when the first component of the next operation was being displayed. This would suggest that one difference between high- and low-span subjects is the amount of rehearsal they use. We recently argued that rehearsal does inflate the span score, but does not drive the relationship between span and abilities (Cantor, Engle, & Hamilton, 1991). We return to this issue in the correlation analysis, which is more suited to answering this question. For now, the punchline of these analyses is that the storage-free processing task does not distinguish between high- and low-span subjects. Additionally, although the processing times on the background operation task show differences between high- and low-span subjects, these differences seem to reflect processing of the to-be-remembered word, not the operation itself.

It should be noted that the analysis produced some other statistically significant effects, all of which are best explained in the context of the previously mentioned three-way interaction. Overall, both high- and low-span subjects spent longer on the first and preterminal components of the operation than on other components. This was reflected in an effect of task component, $F(6, 228) = 40.42, p < .01, MS_e = 82,377.24$. In addition, subjects took longer on the span task requiring recall than on the storage-free task, $F(1, 38) = 771, p < .01, MS_e = 77,153.83$.

Viewing Times as a Function Of Memory Load

When subjects were presented with the first operation–word string of a trial, memory load was zero because there were no previous words to maintain in memory. As trial length increased, so did memory load. Subjects were presented with up to six operation–word strings before recall so memory load varied from 0 to 5.

Our previous analysis indicated that high- and low-span subjects differed in their time on the to-be-remembered word when recall was required. High-span subjects also spent more time on the first component of the operation when recall was required. However, the remaining components of the background task did not differ between high- and low-span subjects, and a preliminary analysis based on load also showed no effect for the other components of the operation. As a result, a composite reflecting all of these remaining parts of the operation was computed. This composite was the average time spent viewing the second through sixth components.

A three-way ANOVA was performed on mean VTs as a function of span (high and low), operation part (first operation component, composite of remaining components, and to-be-remembered word), and memory load (0–5). The results are shown in Figure 2, reflecting the significant three-way interaction, $F(10, 380) = 2.09, p < .05, MS_e = 95,900.62$.

There is one finding that clearly stands out in this analysis. The major difference between high- and low-span subjects across memory load is captured by the time spent on the to-be-remembered word. Low-span subjects showed peak per-

formance on the to-be-remembered word when load was at 2 and then the VTs declined with increased load. High-span subjects spent significantly longer on the word than low-span subjects at loads 2 through 5. Unlike what would be predicted by the task-specific hypothesis, processing times on the operation did not distinguish high- from low-span subjects. All other effects in this analysis were significant except the overall main effect of span ($F < 1$). Because these effects are not meaningful beyond the context of the three-way interaction, they are listed but not discussed further. The significant effects were operation part, $F(2, 76) = 34.7, p < .001, MS_e = 631,474$; the Part × Span interaction, $F(2, 76) = 4.52, p < .05, MS_e = 631,474$; the load main effect, $F(5, 190) = 16.54, p < .001, MS_e = 90,739$; the Span × Load interaction, $F(5, 190) = 3.96, p < .01, MS_e = 90,739$; and the Part × Load interaction, $F(10, 380) = 8.64, p < .001, MS_e = 95,900$.

If one thing is clear from the two analyses presented to this point, it is that high- and low-span subjects do not differ on the processing component of the operation–word task, regardless of whether storage is required. How they do differ is in the amount of time spent on the to-be-remembered words. High-span subjects simply spend more time on the words when recall is required, and they increase the time on the words, as memory load increases, more than do the low-span subjects.

But these analyses do not tell us whether the typically observed correlation between span and verbal ability is a result of some factor reflected in this differential time on the to-be-remembered words. The strategic allocation hypothesis argues that high-span subjects are simply more strategic in all tasks, including reading and listening comprehension tasks. If this is true, then there is nothing specifically important about individual differences in working memory capacity and comprehension. Furthermore, if the time spent on the words is factored out of the relationship between span and comprehension, that relationship should disappear. This question is best answered using correlational procedures.

Correlational Analysis

Descriptive statistics. The descriptive statistics for VSAT, QSAT, the absolute span score, and task accuracy are reported in Table 1.

Span reliability. A reliability score based on three composite scores was computed for the operation-span task. In

Table 1

Descriptive Statistics for VSAT, QSAT, Absolute Span Score, and Task Accuracy

Measure	M	SD	Minimum score	Maximum score
VSAT	476.3	111.9	210	680
QSAT	521.0	98.3	260	710
Absolute span	25.2	9.2	9	49
Task accuracy				
Operation without recall	98%	3.6	88%	100%
Operation with recall	96%	2.8	87%	99%

Note. VSAT = verbal SAT (Scholastic Aptitude Test); QSAT = quantitative SAT. $N = 70$.

the span task, subjects received three trials each of two, three, four, five, and six operation–word sequences. Three separate scores each reflecting the total number of words recalled from one trial at each group level were calculated. Thus, each of the three composite scores was made up of performance in five trials (i.e., one trial at each group level) of the span task. The reliability score for the span task, as determined using Cronbach's alpha, was .74.

Accuracy data. Did accuracy in performing the operation component bear any relationship to span score, comprehension, or both? Correlations among task accuracy (for the operation task without recall and the operation task with recall), VSAT, QSAT, and the span score were calculated. These correlations are reported in Table 2.

Accuracy in solving operations in the operation task without recall did correlate with VSAT and QSAT despite the ceiling on operation accuracy. But QSAT and VSAT were highly correlated in the present study (.69), and when QSAT was controlled, the correlation between VSAT and operation accuracy was nonsignificant ($r = .17$, $p > .05$). Performance in the operation component of the task with recall did not correlate with VSAT or QSAT. If accuracy reflects processing efficiency, the Daneman and Tardiff (1987) hypothesis would predict that performance on the processing component should predict the span score. However, the span score was not related to operation accuracy in either task.

Correlations between working-memory span and comprehension. The main goal of this study was to examine the relationship found among complex memory span measures and comprehension as indexed by VSAT. The questions addressed with the correlation analysis, therefore, were whether a relationship existed between the operation–word span and VSAT and, if so, whether processing time in the span task drives that relationship.

A significant correlation of .34 ($p < .05$) was found between VSAT and the number of words recalled in the operation-span task. This correlation is slightly smaller than found in previous studies, but is consistent in indicating that working memory, as measured by complex memory-span tasks, significantly relates to comprehension as indexed by VSAT. Although span was correlated with VSAT, it was not correlated with QSAT in the present study ($r = .21$).

Correlations among viewing times and comprehension. Carpenter and Just (1989) showed that high-span individuals spend less time on the processing portion of the tasks and more time on the words than low-span individuals. This suggests that the VT on the operations should correlate with span and VSAT performance. Moreover, according to the task-specific hypothesis, there should be a negative correlation among operation VTs and span, and a negative correlation among operation VTs and VSAT.

Storage-free operation task. Correlations among VTs for the operation task without recall, VSAT, and QSAT are reported in Table 3. The VTs for the six elements of the operation without recall were all significantly correlated with VSAT and QSAT. However, these times were not correlated with span. Furthermore, the time subjects spent viewing words was not correlated with VSAT, QSAT, or span. Remember that the words were not to be recalled in this task. Overall,

Table 2
Correlations Among Task Accuracy, VSAT, QSAT, and Span Score

	Task accuracy	
Measure	Operation without recall	Operation with recall
VSAT	.36*	.15
QSAT	.35*	.20
Absolute span	.21	.20

Note. VSAT = verbal SAT (Scholastic Aptitude Test); QSAT = quantitative SAT. $N = 70$.
* $p < .05$.

the important correlations to note for the operation task without recall are those among the VTs, VSAT, and span. These correlations ranged from −.31 to −.51, indicating that subjects with higher VSAT scores responded faster to operations. However, the time spent viewing the operations components was not correlated with span, which ranged from −.01 to −.20.

Operation task with recall required. The correlations among VTs for the operation-span task, VSAT, and QSAT are reported in Table 4. With few exceptions, the results were similar to those with the operation task without recall. Viewing times for the elements of the operation were significantly related to VSAT and QSAT at almost the same level as for the storage-free operations. A partial correlation controlling for QSAT was .21 and not significant ($p > .05$), which suggests that math ability probably underlies the relationship that was found between VSAT and time spent processing operations in the span task.

Again, contrary to the Daneman and Tardiff (1987) hypothesis, the VTs were not correlated with span. The correlations among VTs and memory span were all nonsignificant and ranged from −.07 to .01, indicating that VTs on the operations were not related to the number of words recalled.

What appears to be important to performance in the span task was the amount of time subjects spent on the to-be-remembered word. Time spent on the word was correlated .29 ($p < .05$) with memory span. This correlation indicates that as one might expect, the more time subjects spent on the word, the more words they recalled. If this relationship was bought at the expense of time spent on the operations, then the correlations between time spent on the parts of the operations and span should have been significantly negative. None of these correlations were above .1 and none were significant.

The real punchline of the present study is revealed in the correlations among the VTs, span scores, and VSAT. Significant negative correlations, which ranged from −.29 to −.53, were found among time spent on each part of an operation and VSAT. However, the time spent looking at a word was not significantly correlated with VSAT. Taken together, these correlations indicate that the higher an individual's VSAT score, the faster that individual was in proceeding through the operations. However, high-VSAT subjects did not spend any more time processing the to-be-remembered words than low-VSAT subjects.

To summarize, subjects who adopted a strategy of looking at the words longer recalled more words in the operation–

Table 3
Correlations Among Viewing Times for the Operation Task Without Recall, VSAT, QSAT, and Span Score

Measure	Operation task without recall						
	(#	×/	#)	+--	#	=__	word
VSAT	-.36*	-.51*	-.47*	-.50*	-.31*	-.47*	-.12
QSAT	-.31*	-.55*	-.48*	-.59*	-.37*	-.48*	-.23
Span	-.17	-.20	-.16	-.15	-.01	-.13	-.06

Note. VSAT = verbal SAT (Scholastic Aptitude Test); QSAT = quantitative SAT. $N = 70$.
* $p < .05$.

word-span task, regardless of how much time they spent viewing and solving the operations. However, subjects with higher comprehension skills were no more likely to adopt a strategy of looking at the words longer than were subjects with low comprehension skills.

Partial Correlations

Another way to determine the extent to which processing is important to the relationship between working memory and VSAT is to control for the time that subjects spent processing information in the operation–word-span task. The question was whether the relationship found between span and VSAT occurs because of the speed with which individuals process information. If the speed or efficiency with which individuals process information is important to the relationship between span and VSAT, the relationship between span and VSAT should disappear when the time spent processing operation–word strings is partialed out.

Controlling for the time spent on the components of the operation in the span task and looking at the relation between span and VSAT still resulted in a significant partial correlation ($r = .36$), $F(1, 67) = 10.07$, $p < .05$. If rehearsal were a common factor in the relation found between span and VSAT, then the correlation between span and VSAT should decrease significantly when the time subjects spend rehearsing words is held constant. In the present study, one measure of rehearsal time for subjects is the time subjects spent viewing the word in an operation–word sequence. A partial correlation was computed between VSAT and span, holding constant the time individuals spent on the words. The resulting correlation was .38, $F(1, 67) = 11.35$, $p < .05$, and not significantly higher than the zero-order correlation of .34. That this partial correlation between span and VSAT was significant suggests that rehearsal is not important to the relationship found between

span and VSAT. These partial correlations imply that the relationship between span and comprehension exists over and above the relationship that span and VSAT share with time spent processing the operations and reading the words.

As suggested earlier, subjects may have also rehearsed the to-be-remembered words as they looked at the first component of an operation as well as when they were viewing the word itself. If time spent on both the word and also the first part of the operation indicate rehearsal, and rehearsal is a common factor in the relationship that occurs between span and comprehension, then the correlation between span and VSAT should be weakened when time on the word and on the first component of the next operation is controlled. A composite score for the time spent on "#" and the "word" was calculated and partialed out of the relationship of VSAT and span. A significant partial correlation of .41 resulted, $F(1, 67) = 13.46$, $p < .05$. Again, it was in the wrong direction for the Daneman and Tardiff (1987) hypothesis, but was not significantly larger than the zero-order correlation. This partial correlation casts doubt on any suggestion that high-span subjects are simply more strategic than low-span subjects and that this is the factor responsible for the span–comprehension correlation. It also supports the idea that rehearsal of words in a span task may weaken the relationship between span and comprehension.

Discussion

High- and low-span subjects did not differ in the way they performed the storage-free operation task. Even with the requirement to recall, high- and low-span subjects did not differ on the time spent on the operation-processing component. They did, however, spend different amounts of time on the to-be-remembered word. Although high- and low-span

Table 4
Correlations Among Viewing Times for the Operation Span Task, VSAT, QSAT, and Span Score

Measure	Operation span task						
	(#	×/	#)	+--	#	=__	word
VSAT	-.45*	-.53*	-.46*	-.48*	-.29*	-.33*	-.09
QSAT	-.44*	-.57*	-.50*	-.53*	-.30*	-.35*	-.06
Span	-.04	-.09	-.03	-.03	.01	-.07	.29*

Note. VSAT = verbal SAT (Scholastic Aptitude Test); QSAT = quantitative SAT. $N = 70$.
* $p < .05$.

subjects spent about the same amount of time processing the to-be-remembered words when there was no memory load, as load increased, the high-span subjects increased time on the words more than the low-span subjects.

The correlational analysis showed that VT on the operation component, both in the storage-free version and when recall was not required, correlated with VSAT. Even when the VT on the operation was partialed out, the correlation between span and VSAT was still significant. There was also a correlation between the amount of time that subjects spent on the to-be-remembered word and the span score. When this was partialed out, the correlation between span and VSAT was still significant. So, the correlation between span and VSAT went from .34 to .36, to .38, to .41, as more and more of the processing times were partialed out. These results are quite counter to the view that processing efficiency differences mediate the relationship between span and comprehension and therefore fail to support either the general processing or task-specific processing hypotheses. We also show no evidence for the view that strategic allocation of resources accounts for this relationship. On the other hand, finding that the reading-comprehension measure correlated with the operation–word span and that the correlation remained significant, even with processing time for the operation and for time spent rehearsing the span word partialed out, clearly supports the general capacity hypothesis.

Experiment 2

We just argued that the operation-span task is actually better than the reading-span task for testing general working-memory resources independent of specific processes. The results of Experiment 1 support that contention. It could be argued, however, that the results of this experiment are peculiar to the operation–word-span task and would not generalize to other forms of complex working-memory tasks, such as the reading span. To head off this concern, we conducted Experiment 2 in which subjects performed a version of the reading-span task using a moving-window presentation format. We were interested in whether VTs were different for high- and low-span subjects on the processing component and storage component. Unfortunately, in the version of the reading-span task developed by Daneman and Carpenter (1980), the to-be-remembered word was part of the processing component. Therefore, we used a version of the reading span in which the to-be-remembered word was separate and independent of the sentence. La Pointe and Engle (1990) showed that this version correlated with VSAT .34, if the to-be-remembered words were three to four syllables in length, and .54, if the words were a single syllable in length.

Method

Subjects

Seventy introductory psychology students participated for class credit. They were not the same subjects used in the previous experiment. Subjects were selected on the basis of the same VSAT intervals used in Experiment 1. Because of the small number of students found in the lowest VSAT interval (i.e., 200–350), 7 students were tested in that interval and 13 students were tested in the next to lowest interval (350–400). Ten subjects were tested in each of the remaining intervals.

Stimuli and Procedures

Subjects completed one sentence–word task without recall and one sentence–word-span task requiring recall, in that order. Both tasks were subject-paced, using the moving-window procedure. Subjects were tested individually in one 30–45-min session. The stimuli were presented on the color monitor of an IBM PS/2, Model 50 computer in 14-point type. Stimuli were shown in white on a black background. Both tasks were programmed with Micro Experimental Laboratory, Prerelease system 120 (Psychology Software Tools Inc., 1988).

One list of 20 and one list of 80 common one-syllable words were constructed. Words were printed in all capital letters so that they were easily distinguishable from the sentences they followed. Examples of words used include *STREET, MIND,* and *LUCK.* One set of 20 and one set of 80 sentences were used in the storage-free task and the task that required recall of the word. Sentences were chosen from La Pointe and Engle (1990, Experiment 2) and were 11–16 words in length. Content questions about the sentences were developed for both tasks. Six questions were used for the task without recall, and 20 questions were used in the span task.

Sentence–word task without recall. Subjects read 20 sentences with normal punctuation, each of which was followed by a word unrelated to the sentence and typed in all capitals. Subjects paced through the sentence–word sequence by pressing the plus sign key on the numeric keypad. Each time they pressed the key, a word was displayed on the screen where it would have been displayed if the entire sequence had been presented at one time. One example sequence is *The old lady was too weak to open the heavy church door. FOOD.* Subjects were instructed to read the sentence and following word aloud. They were encouraged to read the sentences for meaning and were told that they would be asked questions about sentences they had read at varying points during the task. Questions were asked after Sentences 2, 6, 9, 14, 17, and 20.

All subjects received the same order of sentence–word sequences and question presentation. The time spent viewing each word was recorded. However, subjects were not informed that their keypresses were being timed.

Sentence–word-span task. Subjects paced through sentence–word sequences and were asked questions as in the other task. In addition, subjects recalled the words following the sentences after reading a group of sentence–word sequences.

Trials were groups of two to six sentence–word sequences, and there were four trials at each group size. Trials were randomly ordered so that subjects would not be able to anticipate the number of words that they would receive before they had to recall words. To ensure that subjects were paying attention to the sentences they were reading, a true or false question was given after each trial. The question asked about one of the sentences in the trial. One random order of presentation of sentence–word sequences and group sizes was chosen.

Subjects paced through the sentence–word task in the same manner as in the task without recall. The end of a trial was marked by the presentation of a display of "Recall Words," at which time subjects were to write down the words that had followed the sentences. They were given as much time as they needed and were told to try to write the words in the correct serial order and to guess if unsure.

After subjects finished recalling the words, they pressed a key that initiated the display of a true or false question about one of the sentences from the previous trial. Subjects responded to the question by pressing either a key on the numeric keypad labeled *T* for true or one labeled *F* for false. After their response, they pressed a key to

begin the next trial of sentence–word sequences. One practice trial of three sentence–word sequences and a question were given before the span task was begun.

Results

Scoring

Span was calculated in the same manner as in Experiment 1. Viewing time was recorded for each word of the sentences and for the to-be-remembered word. Because the sentences ranged in length from 11 to 16 words, it was not possible to obtain the mean time on each component of a sentence. As a result, we derived several measures of VT that collapsed across sentence length.

Experiment 1 showed that high- and low-span subjects differed in their time to view both the first component of an operation and the to-be-remembered word. Consequently, for the current experiment, we calculated the mean VT for the first word of the sentence, the mean for the last word of the sentence, a composite mean for all the words in between the first and last word, and the mean for the to-be-remembered word. Thus, there were four VTs for each task.

For each task, the proportion of correct true or false answers to questions about sentences was calculated. Accuracy was above 85% for both tasks. An ANOVA with span and task as variables indicated no differences in accuracy between high-

and low-span subjects or across tasks and no interaction between these variables. All Fs were less than 1.0.

Processing differences between high- and low-span subjects. As with Experiment 1, 20 high- and 20 low-span subjects were identified, representing the highest and lowest portions of our distribution. The mean VTs were submitted to a three-way ANOVA reflecting the variables of span (high and low), task (with or without recall), and sentence component (first word, middle words, last word, and to-be-remembered word).

The most important outcome of this analysis is depicted in Figure 3, reflecting the significant interaction of Span × Task Type × Task Component, $F(3, 114) = 3.05$, $p < .04$, $MS_e = 796,028$. Unlike Experiment 1, there was a time measure in the task without recall that distinguished high- from low-span subjects. However, this measure was not part of the sentence. High-span subjects spent significantly longer viewing the unrelated words following the sentences. This was surprising because the words did not have to be recalled and this task was always presented before the task that required recall. It is quite possible that the high-span subjects used the time when the span word was present as additional sentence wrap-up time because they were to answer questions about the sentences later.

In this same task, the processing-efficiency parameters associated with reading sentences did not show a difference
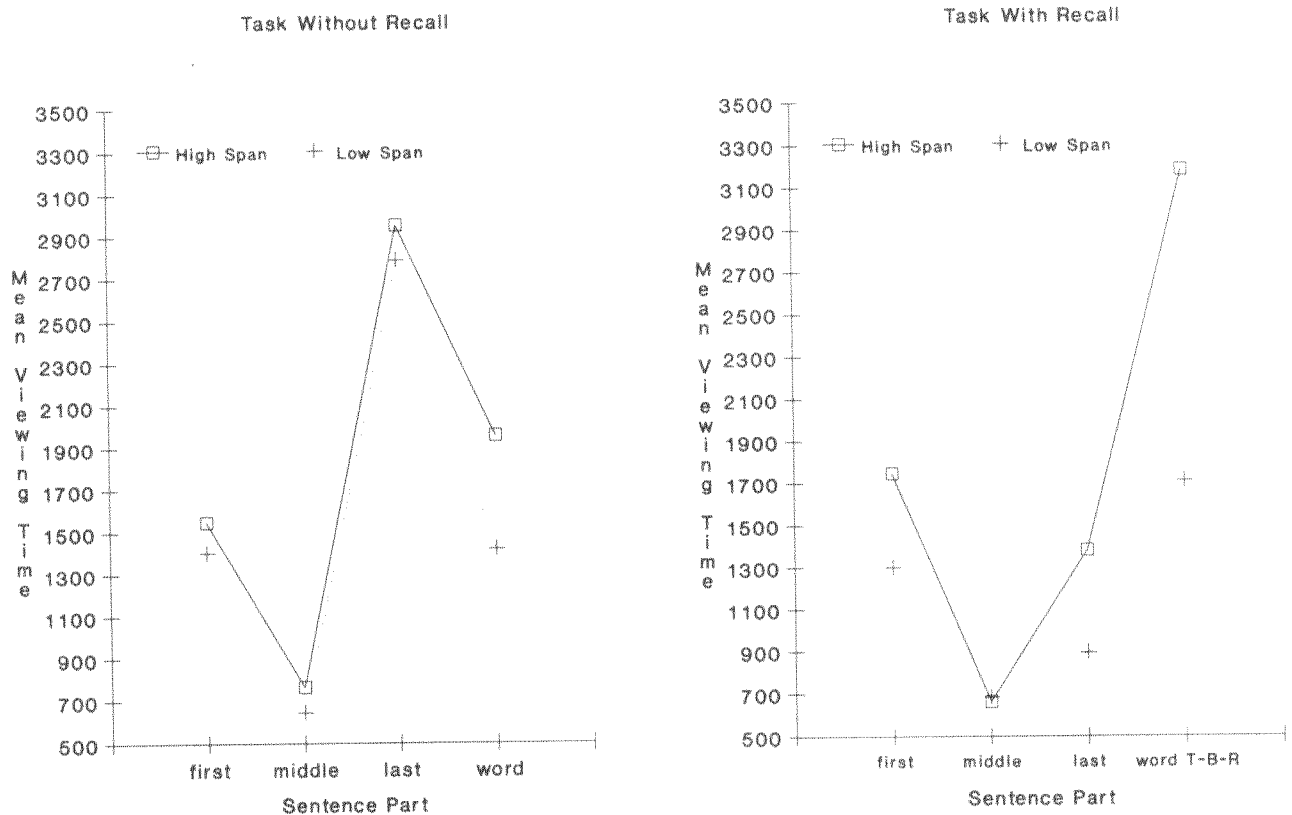

Figure 3. Mean viewing times for low- and high-span subjects in the reading task with and without recall required and as a function of the elements of the sentence and the to-be-remembered word in Experiment 2.

between high- and low-span subjects. This finding parallels the results of Experiment 1 and provides additional evidence against the notion that processing efficiency is the important factor in individual differences in working memory, even when storage is not required.

When recall was required, high-span subjects spent significantly longer on the first word of the sentence than did low-span subjects. This is also consistent with the results of Experiment 1 and probably reflects different rehearsal strategies by the high- and low-span subjects. When recall was required, high-span subjects also spent more time on the final words of sentences than did low-span subjects. Similar findings were reported by Carpenter and Just (1989), but in their task, the final words were also the to-be-remembered words. We have controlled for this confound and still find that high- and low-span subjects differ in the time spent on the final word of the sentence.

Several other important effects can be seen in Figure 3. Time spent on the final words of the sentences decreased for both high- and low-span subjects when recall was necessary, relative to when recall was not required. This produced a significant Task Type × Task Component interaction, $F(3, 114) = 27.62$, $p < .01$, $MS_e = 796,028$, and may indicate that when subjects are working under an extraneous memory load they spend less time in the wrap-up phase of sentence processing. There was also a main effect of span, with high-span subjects viewing longer than low-span subjects, $F(1, 38) = 14.00$, $p < .01$, $MS_e = 1,194,983$. There was also a main effect of task component, $F(3, 114) = 32.58$, $p < .03$, $MS_e = 873,520$, and an interaction of Span × Task Component, $F(3, 114) = 3.16$, $p < .03$, $MS_e = 873,520$. These are best described by the three-way interaction and are not discussed further.

Thus far, we have found no evidence for the notion that processing efficiency produces individual differences in working memory, with or without storage required. There do appear to be differences when recall is required, and these differences are found at the first and last words of the sentences and for the to-be-remembered words. We next look at how these parameters are affected by memory load.

Viewing times as a function of memory load. Mean VTs for the sentence–word-span task were submitted to a three-way ANOVA with the variables of span (high and low), load, (0–5), and task component (first, middle, and last words, and the to-be-remembered words). Figure 4 shows the significant three-way interaction, $F(15, 570) = 3.19$, $p < .01$, $MS_e = 356,782$. The first panel shows that high- and low-span subjects treated the first word of the sentence differently as load increased. Low-span subjects showed only a slight increase in VT for the first word as load increased, whereas high-span subjects showed a more precipitous increase with load. The second panel shows that high- and low-span subjects spent the same amount of time on the words in the middle of the sentence, and neither showed any change as a function of load.

The time spent on the last word of the sentence–word-span task was quite revealing. Just and Carpenter (1980) and Rayner, Sereno, Morris, Schmauder, and Clifton (1989) showed that when reading normal sentences, gaze time increases for the last word. This has been interpreted as wrap-up time used to finish processing of the sentence. When there was no memory load in the present study, high- and low-span subjects spent exactly the same amount of time on the final word in the sentence, which corresponds to the findings with the storage-free sentence–word task. As the memory load from the to-be-remembered words increased, however, VT for the high- and low-span subjects diverged. Low-span subjects did not spend much more time on the final word under load than under no load. High-span subjects, on the other hand, spent much more time on the final word under load than under no load.

Finally, as with the previous experiment using the operation span, high-span subjects spent significantly longer viewing the to-be-remembered words than did low-span subjects. Again, we suspect that this difference, along with differences associated with the first words of the sentence, reflects rehearsal strategies. Particularly interesting is the time spent on the to-be-remembered word. The pattern of VTs for high- and low-span subjects, across memory load, looks remarkably similar to the results of Experiment 1, and we reexamine this finding in Experiment 3.

It should be noted that the memory-load analysis for this experiment showed all effects to be statistically significant. Because these effects are qualified by the three-way interaction, they are simply listed here without discussion. Overall, high-span subjects showed longer VTs than did low-span subjects, $F(1, 38) = 18.63$, $p < .01$, $MS_e = 4,866,491$. There were also significant main effects associated with load, $F(5, 190) = 11.66$, $p < .01$, $MS_e = 296,283$, and with task component, $F(3, 114) = 48.45$, $p < .01$, $MS_e = 2,979,186$. Span interacted with load, $F(5, 190) = 6.62$, $p < .01$, $MS_e = 296,283$, and with task component, $F(3, 114) = 9.62$, $p < .01$, $MS_e = 2,979,186$, and the Span × Task Component × Load interaction was also significant, $F(15, 570) = 6.30$, $p < .01$, $MS_e = 356,782$.

Correlational Analysis

As with the previous experiment, the analyses of time spent on elements of the task showed differences between high- and low-span subjects. Again, however, these analyses do not tell whether any of these differences drive the relationship between span and verbal ability. To examine this issue, the data from all 70 subjects were submitted to correlational analyses.

Descriptive statistics. The descriptive statistics are shown in Table 5 for VSAT, sentence–word span, and the accuracy for the questions in the storage-free sentence–word task and the task that required recall.

Span reliability. The reliability score for the sentence–word-span task, calculated as in Experiment 1 and using Cronbach's alpha, was .89.

Accuracy data. Did accuracy in answering questions about the sentences bear any relationship to span score, to comprehension, or to both? Correlations among task accuracy (for the sentence–word task without recall and with recall) and VSAT and the span score were calculated. Although the four correlations among task accuracy in both tasks and VSAT and span were significant, they were all low (.21–.23). The important finding was that even when accuracy was

First Word

Middle Words
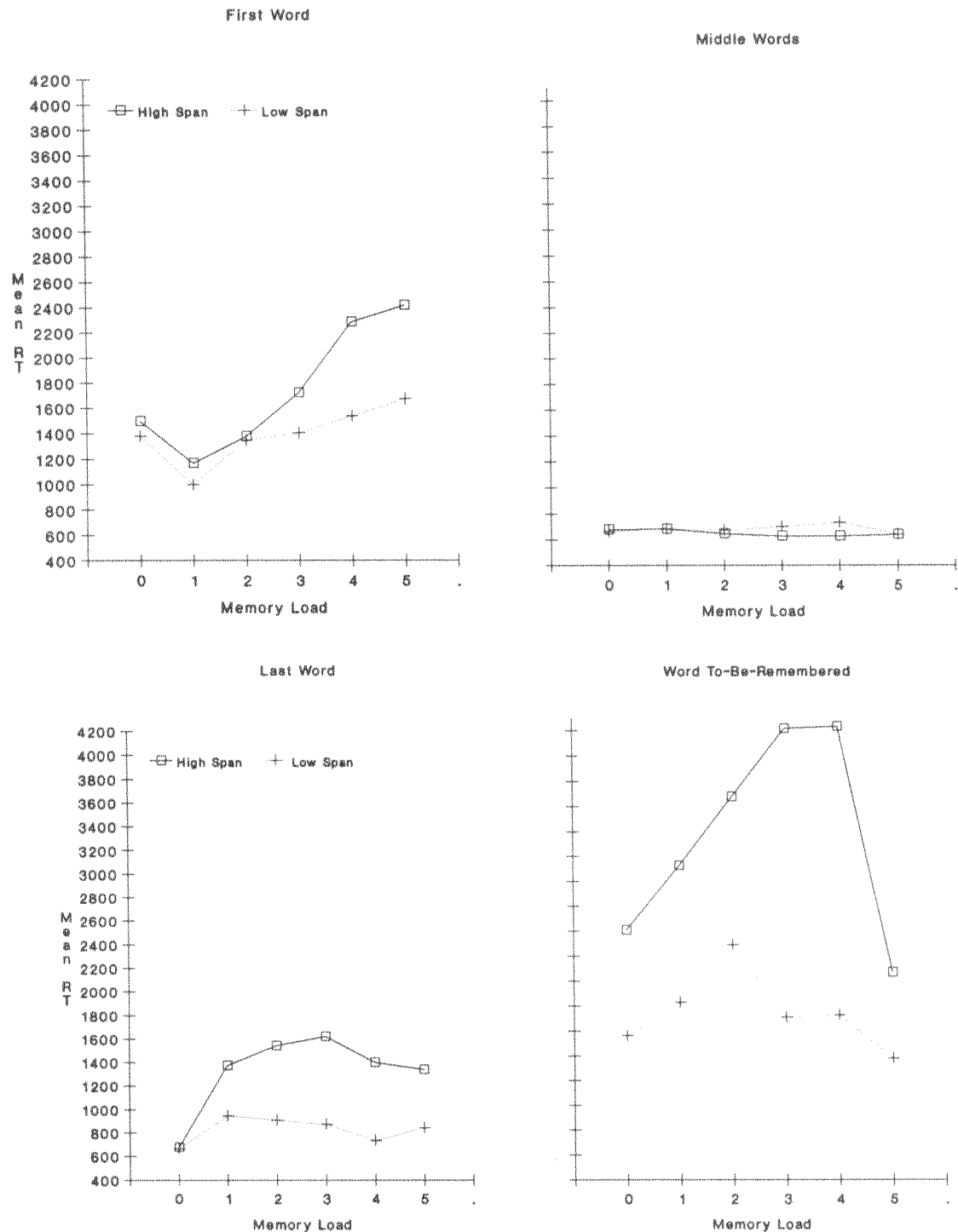
Last Word

Word To-Be-Remembered

Figure 4. Mean viewing times for the elements of the first word of the sentence, a composite of the middle words, the last word of the sentence, and the to-be-remembered word as a function of memory load in Experiment 2.

partialed out of the relationship between span and VSAT, the correlation was still significant ($r = .37$, $p < .001$), for both the storage-free task accuracy and span accuracy.

Correlations between sentence–word span and comprehension. The span score in the sentence–word-span task requir-

ing recall significantly predicted VSAT ($r = .40$), a finding that is now well established. It should be pointed out that this is not statistically higher than the correlation between the operation–word span and VSAT in Experiment 1 ($Z = .23$, $p > .05$).

Table 5

Descriptive Statistics for VSAT and Span and Performance on the Questions After the Sentence–Word Task

Measure	M	SD	Minimum	Maximum
VSAT	480.00	102.90	280	680
Span	27.20	15.07	0	65
Content question accuracy				
Without recall	85.55	15.18	60	100
With recall	87.75	8.25	70	100

Note. VSAT = verbal SAT (Scholastic Aptitude Test).

Table 6 presents the intercorrelations among span, VSAT, and VTs across the different tasks. In general, the number of words recalled in the span task correlated positively with processing time on both the storage-free sentence–word task and the task requiring recall. Higher span individuals tended to spend more time viewing the words in the sentence and the to-be-remembered word, regardless of whether recall was required. The time spent processing sentences did not, with one minor exception, correlate with VSAT.

Partial correlations. Although the results show that time spent on parts of the sentences and the to-be-remembered word predicted span, the real question is whether those differences accounted for the relationship between span and VSAT. In fact, the correlation between span and VSAT remained significant and essentially unchanged when the VTs for all the components were partialed out of the correlation. Removing the covariation attributable to time on the first word, middle words, last word, or to-be-remembered word—either singly or in combinations—from the correlation between span and VSAT did not change the magnitude of the correlation. Remember that the zero-order correlation between span and VSAT was .40. When the VTs for all four variables were removed, the partial correlation was still .40 and significant ($p < .001$).

Discussion

High- and low-span subjects did not differ in the time to read the sentence portion of the storage-free task. When recall was required, high-span subjects spent more time on the first word of the sentence, the last word of the sentence, and the to-be-remembered word. When memory load was zero in the sentence-span task, there was no difference between high- and low-span subjects in the time spent on the to-be-remembered word. As memory load increased, high-span subjects increased

the time on the to-be-remembered word, but low-span subjects showed only a slight increase. These results were nearly identical to those found in Experiment 1.

The correlational analysis showed that VT on the sentence component, both in the storage-free version and when recall was required, correlated with span, but not with VSAT. When the VTs were partialed out of the relationship between span and VSAT, the correlation was virtually unchanged.

Experiment 3

We argued that some short-term memory (STM) tasks encourage low-level rehearsal by using digits, letters, or the same small pool of words over and over. These tasks obscure the relationships among STM capacity and measures of higher level cognition. Essentially, rehearsal inflates span, and the span score then would not reflect the number of schemes or structures the subjects could keep active in complex tasks that do not afford rehearsal. Recent studies from our laboratory supported this notion by demonstrating that performance on a traditional word-span task reliably predicts verbal abilities when an unlimited pool of words is sampled without replacement, such that no word is repeated in the span test (Cantor et al., 1991; Engle, Nations, & Cantor, 1990; La Pointe & Engle, 1990). Additionally, we showed that when the recall scores on an STM task are broken down by serial position, scores on the first few items (i.e., those that are most likely to be rehearsed) do not correlate with ability measures. However, scores on end-of-the-list items, those least likely to be rehearsed, do consistently predict measures of verbal abilities (Cantor et al., 1991; Cohen & Sandberg, 1977).

For this experiment, we adapted the STM word span for use with the moving-window technique so that we could measure the amount of time subjects spent rehearsing words in a simple word-span task. With this technique, subjects control the rate of word presentation and can, if they choose, delay presentation to rehearse previously presented stimuli. If our assumptions about the effects of rehearsal on the span-comprehension relationship are correct, then three predictions should be supported. First, recall performance on the subject-paced task should be higher than on the rapidly paced version typical of word-span tasks, because the former would encourage rehearsal. Second, because of this rehearsal, there should be no correlation between word span in the self-paced task and VSAT. Third, if VT is also an indication of rehearsal, then the time spent viewing words should show no correlation with VSAT.

Table 6

Intercorrelations Among Span, VSAT, and Viewing Times Across the Different Tasks

Measure	Task without recall				Task with recall			
	First word	Middle word	Last word	TBR word	First word	Middle word	Last word	TBR word
Span	.07	.32*	.21*	.44**	.28*	.10	.37*	.44**
VSAT	.01	.20*	−.06	.16	−.01	−.10	.16	.15

Note. VSAT = verbal SAT (Scholastic Aptitude Test); TBR = to-be-remembered. $N = 70$.
* $p < .05$. ** $p < .01$.

Method

Subjects

Subjects were 40 University of South Carolina students who received class credit for participation and agreed to have their SAT scores verified from university records.

Design and Materials

Subjects completed two, simple word-span tasks: one that was self-paced and one in which the words were presented at one word per second. Presentation of the two span tasks was counterbalanced.

The to-be-remembered words were presented in 14-point type in the center of a color monitor driven by an IBM PS/2 Model 50 computer. They were presented as white letters against a black background and were read aloud by the subject. Both tasks were programmed with MEL (Psychology Software Tools Inc., 1988). Two lists of 165 common two-syllable words were used in the span tasks. The two lists were counterbalanced over the two tasks.

Experimenter-paced word-span task. Words were displayed at a one word per second rate, with the order of words randomized for each subject. Trials ranged from three to eight words in length, with five trials at each of the lengths. The span task began with three words per trial and progressed to eight words per trial. Five trials were completed at a given length before the number of words in a trial was increased. After a group of words, a question mark was displayed in the center of the screen, and subjects were to write the words in the correct serial order. They were given as much time as they needed for recall.

Subject-paced word-span task. This task was identical to the standard word-span task, except that subjects pressed a key to view each word of a single trial and the question mark that cued recall. The time spent viewing the individual words was recorded by the computer, but subjects were not informed of this until debriefing.

Results

Scoring

The span score and VT was measured and scored in the same manner as in the previous two experiments.

Reliability of the Span Score

Reliability was measured as in the previous experiments, except the composite score included one from each length (3–8), and there were five such composites. The Cronbach's alpha was .92 for the experimenter-paced task and .91 for the subject-paced version.

Descriptive Statistics

The descriptive statistics for both tasks are presented in Table 7. Mean VT refers to the average time spent on the words, ignoring serial position and trial length. As predicted, subjects recalled more words in the subject-paced task than in the one that was experimenter paced, $t(39) = 2.34$, $p < .05$. We suggest that this difference reflects the increased opportunity for rehearsal when the task is subject paced.

Table 7

Descriptive Statistics for VSAT, Standard Word Span, Subject-Paced Word Span, and Viewing Times

Measure	M	SD	Minimum	Maximum
VSAT	502.70	118.70	290.00	800.00
Standard word span	49.60	20.60	23.00	128.00
Subject-paced span	57.60	24.70	19.00	104.00
Viewing times	2.48	1.36	1.04	7.28

Note. VSAT = verbal SAT (Scholastic Aptitude Test). Viewing times are in seconds.

The 12 highest and 13 lowest span subjects were selected from the self-paced task. The experimenter-paced and subject-paced span scores for these subjects are shown in Figure 5. The high-span and low-span subjects performed very differently under the two conditions. High-span subjects increased their recall when given the opportunity to set their own pace. However, the low-span subjects showed a slight decrement in recall under self-paced conditions. These effects were confirmed by an ANOVA with main effects of task, $F(1, 23) = 14.0$, $p < .01$, $MS_e = 201.0$; span, $F(1, 23) = 100.3$, $p < .001$, $MS_e = 223.5$; and the interaction of task and span, $F(1, 23) = 14.0$, $p < .01$, $MS_e = 201.0$.

Viewing Times

Assuming that the time subjects spend viewing words in the span task reflects rehearsal, we next examined whether
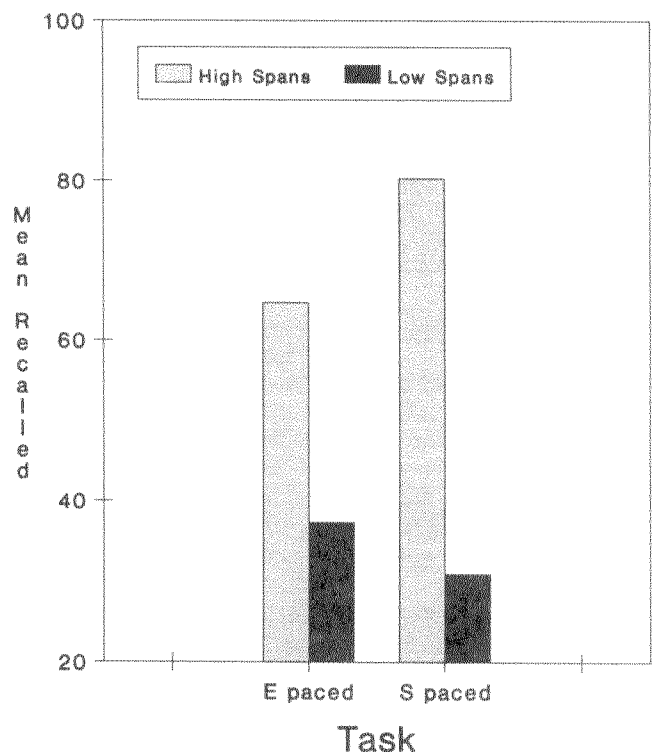

Figure 5. Mean words recalled as a function of task for low- and high-span subjects in Experiment 3. (E = experimenter; S = subject.)

subjects differ in rehearsal depending on span size and memory load. A two-way ANOVA was computed on mean VTs as a function of Span (high or low, based on the subject-paced task) and Memory Load (0–7). The results, shown in Figure 6, reflect the significant Span × Load interaction, $F(7, 161) = 2.19$, $p < .05$, $MS_e = 1,406,592.9$. As load increased, high-span subjects increased time spent on words at a faster rate than low-span subjects. These functions look strikingly similar to the results in Experiments 1 and 2. We argued that those differences indicated differential rehearsal strategies, and we believe that the viewing times on the STM span reflect the same strategies.

It should be noted that this analysis produced a significant main effect for span size, $F(1, 23) = 4.58$, $p < .05$, $MS_e = 2,447,089.7$, and for memory load, $F(7, 161) = 3.90$, $p < .01$, $MS_e = 1,406,592.9$.

Correlations

The zero-order correlations among span in the standard word-span task, span in the subject-paced task, average time on words, and VSAT are presented in Table 8. The traditional word span significantly correlated with VSAT ($r = .49$).

We predicted that the subject-paced span task would encourage rehearsal, and the corresponding span score would not correlate with VSAT. We also argued that VT was an indication of rehearsal and would also show no relationship to VSAT. Both of these hypotheses were supported ($rs = .21$ and .03, respectively).

The only score that predicted VSAT was the number of items recalled in the standard word-span task. In addition,

Table 8

Correlations Among the Standard Word Span, Subject-Paced Word Span, Mean Viewing Time in the Subject-Paced Task, and the VSAT

Measure	Standard span	Subject-paced span	Viewing time	VSAT
Standard span	-----			
Subject-paced span	.56*	-----		
Viewing time	−.02	.49*	-----	
VSAT	.49*	.21	.03	-----

Note. VSAT = verbal SAT (Scholastic Aptitude Test).
* $p < .05$.

partial correlations showed that this relationship did not change when controlling for the time spent on the subject-paced task ($r = .49$) or the number of items recalled on the subject-paced task ($r = .46$). Even when both of these parameters are controlled, the correlation between span on the experimenter-controlled task and VSAT did not significantly change ($r = .46$).

One point should be noted when comparing the results of this experiment with those from the first 2 studies. Allowing self-paced presentation in the first 2 studies did not eliminate the correlation between the number of words recalled and VSAT. However, in Experiment 3, self-paced presentation eliminated the correlation.

It is our contention that even though rehearsal was possible in Experiments 1 and 2, as was evidenced by the increased VTs for the to-be-recalled words, this rehearsal was interrupted by regular, periodic shifts of attention to the background task. We later argue that the measurement of how many items can be retained when those items occur in the context of frequent shifts of attention to other areas of memory are an important aspect of what the complex working-memory tasks are measuring.

General Discussion

The primary goal of these studies was to test four hypotheses about the relationship between the number of items recalled in complex working-memory tasks and reading comprehension as measured by the VSAT. The general processing hypothesis assumes that individual differences on tests of working memory occur because of increases in the efficiency of general cognitive operations. The task-specific hypothesis assumes that differences in storage occur entirely as a result of differences in the skill of operations specific to the task being performed at the moment. The strategic allocation hypothesis assumes that the span–VSAT correlation occurs because of individual differences in the intelligent use of strategies in all cognitive tasks. The general capacity hypothesis, proposed by Turner and Engle (1986, 1989), assumes that the span–VSAT correlation occurs because the complex working-memory span provides a measure of storage capacity general to a wide variety of tasks, at least within the verbal domain.

Both the general processing and task-specific processing hypotheses needed to show (a) a relationship between the measures of proficiency in the processing component of the
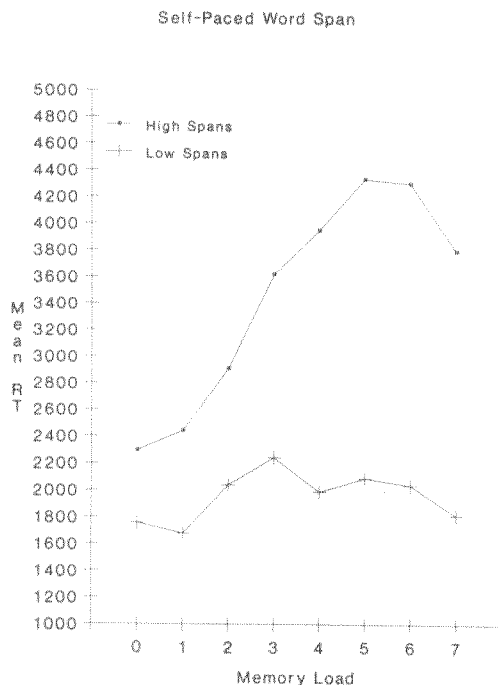
Self-Paced Word Span

Figure 6. Mean viewing times for low- and high-span subjects as a function of memory load in Experiment 3.

span task and VSAT, (b) a relationship between the processing component and span, and (c) that the relationship between span and VSAT is eliminated when the measure of processing skill was partialed out. In Experiments 1 and 2, performance on the storage-free processing component did not distinguish between high- and low-span subjects. When the VTs on the processing components were partialed out of the span–VSAT relationship, that correlation was still significant and undiminished. This strongly suggests that processing skill, at least for those processes and at that level of difficulty measured here, contributes in no way to the relationship between the working-memory span and comprehension.

To support the strategic allocation hypothesis, it would be necessary to show (a) a negative correlation between performance on the processing component and the span score, (b) a positive correlation between time on the to-be-remembered word and the span score, and (c) that high-span subjects traded off processing time for time on the to-be-remembered word more as memory load increased than did low-span subjects. The only one of these findings observed was the positive correlation between the time on the to-be-remembered word and recall of that word. However, there was actually an increase for the high-span subjects in time on the processing components compared with the low-span subjects, at least for Experiment 2. More important, when the VT on the processing components and the time spent rehearsing the to-be-remembered word was partialed out, the relationship between span and VSAT was still significant and undiminished. Counter to the predictions of the strategic allocation hypothesis, high-span subjects seemed to be able to make task-sensitive adjustments in the allocation of their resources, whereas the low-span subjects did not. As memory load increased in all three experiments, the high-span subjects continuously increased the time on the to-be-remembered word and, in the second experiment, on the elements of the processing task as well.

Overall, the data appear to rule out the hypothesis that the relationship between working memory and comprehension is mediated by individual differences in the efficiency with which individuals perform procedures required in the task or in how they allocate their resources. Differences in the time spent on the elements in a complex span task—due either to the efficiency with which individuals process information or to more conscious use of resource allocation strategies while performing cognitive tasks—do not account for the relationship between span and comprehension. The correlation between span and VSAT occurred both with a complex task requiring reading and one requiring the solution of arithmetic strings. More important, this relationship was not reduced when variance attributable to virtually every other element in the complex task was partialed out. This supports the general capacity hypothesis.

A General Capacity Architecture of Working Memory

Before we present our ideas about how individuals vary in working memory, we must talk briefly about our vision of the structure or architecture of working memory. The ideas presented here parallel theories proposed by Anderson (1983),

Schneider and Detweiler (1987), Cowan (1988), and Just and Carpenter (1992). We assume that working memory is much more extensive than assumed by traditional buffer theories (e.g., Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). It consists of those temporary or permanent knowledge units in long-term declarative memory that are currently active. Short-term memory is that information that is maintained at a surface level of coding and is within the grasp of immediate consciousness. Thus, STM is a subset of working memory, which is a subset of long-term declarative memory.

We further assume that knowledge units vary in their ambient level of activation and the total amount of activation in the system is limited. Activation is the fuel or resource that drives the system. Working memory consists of those knowledge units that have recently been activated either from objects in the environment or as a result of productions and are in various states of loss of activation through either decay or inhibition.

A further characteristic of working memory that appears to be essential in everyday life is the ability to shift attention away from the current task, say, as a result of an interruption or distraction, then back to the task to recover the relevant task information and status of task variables at the time of interruption. There are various ways this could be implemented under either symbolic architecture, such as ACT*, or connectionist architecture. Schneider and Detweiler (1987) proposed, as part of their connectionist model of working memory, a context-storage mechanism that codes a set of information by the context, broadly defined, in which it occurred. This mechanism would allow for a much larger working memory and for the recovery of information that has been flushed from consciousness by new interrupting information. Another notion that appears applicable to this function is the concept of retrieval structure proposed by Chase and Ericsson (1982). Retrieval structures are meaning-based links that allow rapid access of information in long-term memory, thus extending the contents of working memory, and have been most clearly demonstrated in the context of individuals with skilled memory. Regardless of the instantiation of this feature, we assume that there is a limitation in either the number of contexts or the number of links resulting from the limitation in total activation.

What is the role of processing under this view? We assume that procedures only make demands on working memory to the extent that (a) they are so poorly learned that a search of declarative memory is required for a production to match the contents of working memory or (b) the procedure produces intermediate and final products that must be stored as a temporary declarative memory unit in working memory.

How does this theory allow for individual differences of the type demonstrated in this article? We assume that individuals differ in the total level of activation available to their system. The amount of total activation available is an abiding characteristic of the system and would change relatively little with changes in the knowledge structure. Thus, an individual might develop a knowledge base that would allow the use of retrieval structures to drastically increase the amount of information stored and recalled. However, an analysis of retrieval would show a relatively small number of links at each level of the

hierarchy (Ericsson & Staszewski, 1989). Broadbent (1975) and Schneider and Detweiler (1987) provided further support for the view that working memory limitations are still observed even when those limitations are extended or apparently circumvented.

We would argue that complex working-memory tasks, such as those used here, measure the amount of activation available to the individual on a moment-to-moment basis. The tasks all demand that the individual attend to the processing component, then to the span item, then to the processing component of the next sentence or operation and then back to the span item, and so on until recall is required. This models the very feature we just described, and we would argue that the number of span items recalled in this context reflects the amount of activation available for maintaining the temporary structures for retention of the span items while concomitantly switching attention back and forth to other interpolated activity.

The Role of STM

Does the simple span task tell us anything about individual differences that is important to higher level cognition? We would argue that some forms of the simple span task reflect individual differences in the use of articulatory coding. To the extent that articulatory coding is important to the higher level task, the simple spans would predict the ability of the subject to keep information active at this superficial level of coding. The digit span is not a good measure because it is sensitive to rehearsal, grouping, and recognition of patterns that are idiosyncratic to digits, and these elaborative strategies are probably not generalizable to cognitive tasks, such as reading. The word span, running digit and word span, and probed recall of digits and words (Cohen & Sandberg, 1977) probably all reflect the amount of information that can be kept active using articulatory coding.

The evidence for that assertion comes from two of our previous articles. Cantor, Engle, and Hamilton (1991) assessed the relationships among complex span measures of working memory, measures of STM, and comprehension as measured by the VSAT. Subjects performed six tasks: a digit span; a word span; a reading span with words; a reading span with digits; and two probed-recall tasks, one with words and one with digits. The probed-recall task presented subjects with a 9-item list followed by a cue to recall the first 3, the second 3, or the final 3 items in the list. We assumed, following Cohen and Sandberg (1977), that the initial three positions would receive the most rehearsal, but the final positions would most accurately reflect the information maintained in an active state without rehearsal.

All of the tasks, except the digit span, correlated with VSAT. This included both reading spans (digits and words), the simple word span, and the probed recall of the last positions for both words and digits. However, a factor analysis showed that the simple span tasks and probed-recall tasks loaded on one factor, whereas the complex span tasks loaded on a second factor. Multiple regression showed that composite scores formed for the STM variable, and the working-memory variable independently contributed variance in VSAT. Together,

the two variables accounted for 42% of the variance in VSAT. This suggests quite strongly that STM and working memory are separate structures and that individual differences in both are important to comprehension. Another finding that is relevant to this issue comes from a series of experiments reported by La Pointe and Engle (1990). They showed that articulatory suppression, which presumably eliminates the articulatory code, caused the correlation between simple word span and VSAT to become nonsignificant, but had no effect on the correlation between the complex span and VSAT. This suggests that individual differences in the use of articulatory coding account for the relationship between the simple word span and comprehension, but not for the relationship between the complex span and comprehension.

We would argue that the STM variable is important in comprehension because of the retention of surface-level codes, such as the exact words of a recent clause or phrase (Jarvella, 1971; Sachs, 1967). The working-memory variable is important in building the mental model or gist of the developing story. Greater working-memory capacity would mean that an individual could comprehend a more complex story; retain subtle subplots, even if irrelevant to the main plot; and connect related propositions, even those separated by time and context.

References

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). San Diego, CA: Academic Press.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). San Diego, CA: Academic Press.

Baldwin, J. M. (1894). *Mental development in the child and the race.* New York: Macmillan.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory* (pp. 3–18). New York: Wiley.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15,* 229–246.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Erlbaum.

Case, R. (1974). Structures and strictures, some functional limitations on the course of cognitive growth. *Cognitive Psychology, 6,* 544–573.

Case, R. (1985). *Intellectual development: Birth to adulthood.* San Diego, CA: Academic Press.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 16, pp. 1–58). San Diego, CA: Academic Press.

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

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human infor-

mation-processing system. *Psychological Bulletin, 104,* 163–191.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

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., & Tardiff, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 491–508). Hillsdale, NJ: Erlbaum.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80,* 465–472.

Engle, R. W., & Marshall, K. (1983). Do developmental changes in digit span result from acquisition strategies. *Journal of Experimental Child Psychology, 36,* 429–436.

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is working memory capacity just another name for word knowledge? *Journal of Educational Psychology, 82,* 799–804.

Ericsson, K. A., & Staszewski, J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–268). Hillsdale, NJ: Erlbaum.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10,* 409–416.

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

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Boston: Allyn & Bacon.

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

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228–238.

La Pointe L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 63,* 301–345.

Piaget, J. (1926). *The language and thought of the child.* London: Routledge & Kegan Paul.

Psychology Software Tools Inc. (1988). *Micro Experimental Laboratory: Prerelease system version 120.* Pittsburgh, PA: Author.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4,* 21–49.

Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437–442.

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 507–516.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 53–119). San Diego, CA: Academic Press.

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

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

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

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). San Diego, CA: Academic Press.

Whitney, P., Ritchie, B. R., & Clark, M. B. (in press). Working memory and the use of inferences in text comprehension. *Discourse Processes.*