Journal of Educational Psychology
1990, Vol. 82, No. 4, 799-804

0022-0663/90/$00.75

Is "Working Memory Capacity" Just Another Name for Word Knowledge?

Randall W. Engle, John K. Nations, and Judy Cantor
University of South Carolina

This study is concerned with whether the correlation between complex working memory spans and reading comprehension occurs because the complex spans reflect the capacity of a structural working memory that plays a causal role in comprehension or because a third factor, word knowledge, plays a causal role in both the span tasks and comprehension. If the latter hypothesis is correct, the correlation between word span and reading comprehension should be large when span is tested with low-frequency words but should not occur when span is tested with very familiar words. Ninety college students were tested on a simple and a complex version of the word span task with high- and low-frequency words. The Verbal Scholastic Aptitude Test (VSAT) was used as a measure of reading comprehension. The correlation between span and VSAT was somewhat higher when span was tested with low-frequency words, but was significant with both low- and high-frequency words. This suggests that both word knowledge and a content-free working memory play a causal role in the relationship between word span and higher level cognitive tasks.

The limited-capacity working memory is a critical element in the explanation of such common and important cognitive tasks as learning, problem solving, and reading and listening comprehension. In the case of comprehension, a reader would be more likely to recall a previously read proposition the longer it was represented in working memory. Furthermore, a reader with a large working memory capacity would be better able to integrate related facts that occur at different places in a text. For example, when the reader encounters a pronoun, the noun referent is more likely to still be represented in the working memory if the reader has a large working memory capacity (Daneman & Carpenter, 1980).

Daneman and Carpenter (1980) argued that the overall executive capacity of working memory does not differ across individuals. However, individuals would differ in the storage component of working memory as a consequence of how much attention their reading processes require. Readers whose reading-specific skills are well learned and automatized would need little of the working memory capacity for processing and would have more residual capacity for storage. The capacity of the short-term storage would be task-specific. A good and a poor reader with corresponding large and small storage capacity while reading might have the same, or even reversed, level of storage capacity in a task unrelated to reading.

Daneman and Carpenter reasoned that for a span measure of short-term storage to correlate with a measure of higher level cognition, the span task must require the same strategies and processes as those used during the higher level tasks. To this end, they used a span task in which subjects read a series of sentences aloud and, at the end of the series, tried to recall the last word from each sentence. Thus, the task was in many ways identical to a simple word span task except that sentences were read aloud between to-be-recalled words. They found that this reading span measure of working memory correlated with several reading comprehension measures including the Verbal Scholastic Aptitude Test (VSAT). They assumed that this correlation occurred because the reading span task required the subject to use the same reading-specific processes used during the reading comprehension test. Moreover, traditional memory span and comprehension are uncorrelated because they do not require common complex cognitive processes.

Recent work, however, suggests that individual differences in working memory capacity are more general (Turner & Engle, 1989). The reading span task is really a dual task with reading the sentences aloud being a background task to the primary task, remembering a list of words. If Daneman and Carpenter's theory is correct, any background task that forces the subject to use skills unlike those used during reading should not predict performance on a comprehension test. However, another possibility is that total executive processing space varies across individuals independently of reading skill. Older tasks of short-term memory, such as the digit span, may not correlate with reading comprehension (Guyer & Friedman, 1975) because these tasks place a premium on superficial coding and rehearsal strategies that probably play a minimal role in comprehending text. Rehearsal increases the number of digits recalled. But whether an individual knows when and how to rehearse would tell us nothing about storage space available to that individual. The reading part of the reading span task may simply serve to prevent grouping, coding, and rehearsal that allows the span score to more accurately reflect the true storage space available to the subject.

This research was supported by AFOSR 870069 to Randall W. Engle. We would like to thank Russel Onley for help in running subjects, Linda LaPointe for help with stimuli and analysis, and Lee Kirkpatrick for advice about the analysis.

Correspondence concerning this article should be addressed to Randall W. Engle, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

Turner and Engle (1989) compared the correlation of different span measures with reading comprehension. In this study, we asked whether the background element of the complex span task must invoke reading-specific skills for the span score to predict reading comprehension.

In addition to simple word and digit tasks, we used complex span tasks that required subjects to either read sentences aloud or solve a string of arithmetic operations as a background to the span task itself. The reading span was similar to the procedure used by Daneman and Carpenter. Subjects saw one sentence at a time and read it aloud. In one condition, the subjects recalled the last words from each sentence in the set. In another condition, however, the sentence was followed by a digit and the subject tried to recall the digits that followed each sentence in the set.

In the operation span task subjects saw a string such as "(4 × 2) − 3 = 6 snow" presented on the monitor of a computer. After the string was read and responded to, subjects were presented another string, such as "(9/3) + 1 = 4 tiger." In one condition the operation was followed by a to-be-recalled word, as in the examples, and in another condition the to-be-recalled item was the digit that followed the equal sign in the operation. The subject's task was to respond "yes" or "no" depending on the correctness of the arithmetic string and to recall the words or digits at the end of each progressively longer set of operations. In the example, the subject would be required to read the operations, respond "no" following the first string, "yes" following the second, and then to recall "snow" and "tiger."

This task would seemingly require different skills than reading a passage for comprehension, and should not predict a criterion test of reading comprehension. The subjects also completed the Nelson-Denny Reading Comprehension test. We obtained subjects' VSAT scores from university records.

If Daneman and Carpenter's theory is correct, the correlation between span score and the reading comprehension measures should only be significant for the reading span because the operation span task does not force the subject to use reading-specific skills. Turner and Engle (1989) found, however, that scores from the operation-word task correlated with both of the reading comprehension measures as highly as did the reading span scores. This finding suggests that working memory capacity is not dependent on the particular strategy used to accomplish the task at hand. Turner and Engle's regression analysis suggested that operation span and reading span tasks tap the same underlying process.

The explanation we have offered for that correlation based on Turner and Engle (1989), is that capacity differences in working memory are a characteristic of the individual and serve a primitive causal role in the performance of higher level cognitive tasks such as reading comprehension. Like all correlations, however, this one is open to many alternative explanations. One simple and attractive hypothesis is that individual differences on both the complex word span task and reading comprehension tests are simply a reflection of individual differences in some aspect of word knowledge. Baddeley and Scott (1971) showed that performance in a short-term memory task is a function of the frequency of the word stimuli. Thus, the same set of words might functionally be low-frequency for subjects with low word knowledge and high-frequency for subjects high in word knowledge. Because individual differences in word knowledge should obviously affect higher level verbal tasks such as reading comprehension, word span would accidentally correlate with reading comprehension and other tasks of a verbal nature.

Several lines of evidence make this argument difficult to discount. One is that nearly all the published reports of significant correlations between measures of working memory and higher level cognitive tasks have used verbally oriented tasks as the criterion variable. Furthermore, nearly all the predictor tasks have been word span tasks. The one exception was the reading digit span and operation digit span tasks used in Turner and Engle (1989), but even with those tasks, the correlations with reading comprehension, although significant, were lower than for the complex word span tasks.

There are several different ways to test the hypothesis that the memory span–comprehension connection is driven by the common variable of word knowledge. One approach would be to test for word knowledge and to partial the variance due to word knowledge out of the correlation in question. This approach was used recently by Dixon, LeFevre, and Twilley (1989). They obtained three measures each of working memory capacity, word knowledge, and reading skill. Their regression analyses and causal modeling led them to conclude that word knowledge and working memory capacity both contributed independently to reading comprehension.

We report here a different approach to this question. If word knowledge is responsible for the high correlation between reading comprehension and word span, then the correlation should be very high when low-frequency words are used, because these words would be familiar to subjects high in word knowledge and unfamiliar to subjects low in word knowledge. However, the correlation should disappear when the words used in the span task are high-frequency words, because there should be little subject variability in word knowledge.

If, however, individuals differ in total executive processing and storage space available to them, and this is the primary causal factor in the relationship between word span performance and comprehension, then a different pattern of correlations should occur. The correlation between reading comprehension and span should be significant with both high- and low-frequency words. If, as Dixon, LeFevre, and Twilley (1989) argued, word familiarity plays an independent role in both span and reading, then the correlation might be somewhat higher with low-frequency words than with high-frequency words, but both should be significant.

We used the operation span and the simple word span as our measures of working memory. In every comparison in our lab, the operation span has predicted higher level cognition at least as well as the reading span and it allows greater control over the difficulty of the task than does the reading span task. As our criterion reading comprehension task, we used the Verbal Scholastic Aptitude Test. The VSAT has several advantages for our purposes. It is highly reliable and the scores were already available. Furthermore, Turner and Engle's (1989) research has shown the VSAT to highly correlate with less global tests of reading such as the Nelson-Denny.

Even those elements of the test that are not specific to reading, such as vocabulary, appear to depend on working memory capacity (Daneman & Green, 1986).

Method

Subjects

Subjects were 90 undergraduates from the University of South Carolina who volunteered to participate and received course credit. Prior to participating in this study, the subjects had agreed to allow our lab to obtain their SAT scores from university records. The procedures met all APA ethical principles for use of human subjects.

Subjects were chosen on the basis of seven intervals of the VSAT to give a rectilinear distribution of scores. These intervals and the number of subjects obtained are presented in Table 1.

Materials

Four different word pools, matched for number of syllables, were constructed from a book of word frequencies (Carroll, Davies, & Richman, 1971). Two pools were composed only of high-frequency words appearing over 500 times per million in the English language. Examples of high-frequency words we used are *under*, *cat*, *difficult*, *after*, *interesting*, *green*, *show*, and *easy*. The remaining two pools were made up of low-frequency words appearing less than 30 times per million. Examples of low frequency words we used are *suite*, *wry*, *sextant*, *putrid*, *flora*, *implement*, *plausible*, and *ballast*.

We also used two sets of mathematical operations identical to those constructed by La Pointe and Engle (1990) to generate stimuli for their operation span task. Each operation required the multiplication or division of two integers, the result of which was to be added to or subtracted from a third integer (e.g., "Is [10/2] − 1 = 6?"). Integers were randomly generated with the constraints that (1) those to the left of the equal sign be digits between 1 and 10, and (2) the correct final product be a whole number. An answer was always provided, and trials were approximately equally divided as to whether or not this answer was correct. The to-be-recalled word appeared to the right of the operation.

Tasks

All tasks were presented with an IBM XT computer on a monochrome monitor. The tasks are described in the following paragraphs.

Simple word spans. There were two word span tasks, one using high-frequency words and the other using low-frequency words. Words were presented one at a time, centered on the screen, at a rate of one per second. Each task consisted of 18 trials that were three to eight words in length. The order of these trials was randomly selected for each subject, so that trial length was unpredictable. The termination of a trial was indicated by the presentation of a question mark. At this signal, the subject was to begin written serial-recall of the words. There was no time limit for recall. The subject indicated to an experimenter when to begin the next trial, which was then initiated by a key press.

Operation spans. As with the word span tasks, there were two operation span tasks, corresponding to high- and low-frequency words. Simultaneously presented on the screen were an operation followed by a to-be-remembered word. Subjects read the operation aloud, provided an oral *yes* or *no* verification as to whether the answer provided was correct, and then read aloud the word. As soon as the to-be-remembered word was read, an experimenter pressed a key that initiated either another operation–word pair or a question mark. The question mark signalled that written serial recall of the words for that set of operations was to begin. Recall was untimed, and a new trial was begun when the subjects indicated that they were finished writing. For each task there were 18 trials that ranged in length from two to seven operation–word combinations. Again, the order of trials was randomly selected for each subject so that list length was unpredictable.

Table 1
Mean and Standard Deviation of Verbal Scholastic Aptitude Test (VSAT) for Subjects in Different VSAT Intervals

VSAT interval	*M*	*SD*	Number of subjects
250–350	320.00	33.99	10
351–400	376.86	13.80	16
401–450	430.00	17.72	12
451–500	481.54	14.05	13
501–550	527.86	14.24	14
551–600	572.31	15.36	13
601–800	650.83	35.79	12

Procedure

Each subject participated individually in a single session for approximately an hour and a half, and received all four span tasks. Each of these tasks sampled exclusively from one of the four word pools. Across subjects, each word pool was approximately equally represented in each task, and task order was also approximately counterbalanced.

Scoring

Three different scoring procedures were used, but all led to the same conclusions. We will present the results of the one we call the Absolute Span score. This score was the sum of all words recalled on perfectly recalled trials. Thus, if a subject recalled perfectly all three trials of length 2, all three trials of length 3, but only recalled two trials perfectly at length 4 and recalled none of the trials at longer list lengths perfectly, the span score would be 17.

Results

Because of a problem in data collection, verification errors for the operation span tasks could only be calculated for 70 subjects but the subjects were evenly distributed over the intervals, with data from 10(±1) subjects in each VSAT interval. Overall errors were very low, averaging around one percent for operations followed by high-frequency words (.79%) and about the same for operations followed by low-frequency words (1.12%). Even though error data for 21 subjects were not available, there is no reason to expect that their data were any different from those of the sample examined.

A preliminary analysis indicated no effects associated with task order, so this factor was not considered further. An analysis of variance (ANOVA) was computed, with the factors being word frequency (high or low), task type (simple span or operation span), and VSAT interval (1–7), with interval contrasted for orthogonal polynomial trends.

Overall, we expected that high-frequency words would be better remembered than low-frequency words, and that retention would be better in the simple span tasks than in the operation spans. Table 2 confirms the expected pattern of results. The analysis showed main effects for frequency, $F(1, 83) = 292.7$, $p < .001$, $MS_e = 43.9$, and for task type, $F(1, 83) = 95.3$, $p < .001$, $MS_e = 64.0$. Interestingly, the difference between high- and low-frequency words was somewhat less pronounced in the operations tasks, resulting in a frequency by task type interaction, $F(1, 83) = 126.2$, $p < .001$, $MS_e = 28.0$. There was a main effect for VSAT interval, $F(6, 83) = 6.6$, $p < .001$, $MS_e = 173.3$, which also showed a significant linear trend, $F(1, 83) = 39.4$, $p < .001$, $MS_e = 173.3$. There were no significant global interactions associated with VSAT interval, but there was a linear interaction between task type and interval, $F(1, 83) = 5.7$, $p < .02$, $MS_e = 64.0$. As can be seen from Table 2, the slope of the simple span across interval is steeper than that of the operation span, but this is probably an artifact of the greater range in the simple span scores.

Correlations

Individuals who score high on tests of vocabulary skills tend to score high on working memory span tasks (Daneman and Green, 1986) and also perform better on standardized tests of reading comprehension. Clearly, there is a relationship between word span, word knowledge, and verbal comprehension. On the basis of the theories we described previously, we expected that the data would result in one of two interpretable patterns of correlations.

One possibility is that the relationship between comprehension and span is mediated by a word knowledge factor. In general, people who know more words may be more familiar with the words in span tasks and in text, thereby producing better memory scores and greater text comprehension than people with less word knowledge. This would produce an indirect relationship between span and comprehension. When the words in the span task are highly familiar to all subjects, however, the span scores should not correlate with comprehension. Only those tasks with low-frequency words should predict verbal comprehension.

An alternative position is that individuals differ in the executive processing and storage space available to them and that this difference has an influence on comprehension independent of word knowledge. According to this view, individual differences in working memory capacity will become manifest in any task that requires immediate memory, including all memory span tasks administered in this study and tests of reading comprehension, such as the VSAT. It should also be noted that this explanation does not ignore the role that word knowledge would play in comprehension. Clearly, reading will be disrupted by unfamiliar words, and subjects are likely to bring to the tasks well established individual differences in word knowledge. Comprehension is probably influenced by many factors, with working memory capacity and word knowledge being among them. We would therefore expect that tasks reflecting both word knowledge and working memory capacity will be the best predictors of comprehension. For this study, such tasks involve low-frequency words. Nonetheless, if working memory capacity has an independent relationship with comprehension, we would also expect that the span tasks involving high-frequency words would still predict VSAT.

Table 2
Mean and Standard Deviation As a Function of Word Frequency, Span Task, and Verbal Scholastic Aptitude Test (VSAT) Interval

VSAT interval	Simple span				Operation span			
	High		Low		High		Low	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
250–350	25.0	11.2	8.5	3.6	11.3	7.1	6.3	4.0
351–400	23.7	11.1	8.8	5.7	14.8	10.8	9.9	5.4
401–450	28.5	9.26	11.1	5.7	14.8	8.2	9.8	5.6
451–500	32.2	9.6	12.9	4.5	19.4	8.9	11.9	5.8
501–550	36.9	13.9	13.9	6.8	20.6	8.8	12.5	4.2
551–600	40.7	13.6	18.7	5.8	18.2	9.9	15.2	9.2
601–800	38.8	14.6	23.3	8.2	24.3	7.2	17.7	11.3

In general, Table 2 shows that for both the simple word and operation tasks the number of items recalled increased as VSAT increased, and that this trend is found with both high- and low-frequency words. The results of the corresponding correlational analyses are presented in Table 3.

The analyses show that both the simple span and operation span tasks with high-frequency words predict VSAT. These results do not support the notion that individual differences in word knowledge alone are responsible for the relationship between working memory span and comprehension. This finding suggests that working memory has an influence on verbal comprehension independent of word knowledge. As predicted, tasks involving low-frequency words tended to be stronger predictors of VSAT than those with high-frequency words. It appears that a combination of working memory capacity and word knowledge best predicts verbal comprehension. In support of this notion, the correlation between simple word span and VSAT was significantly greater with low-frequency words than with high-frequency words, $t(88) = 2.97$, $p < .01$ (one-tailed). For the operations span tasks, there was relatively little difference between the correlations involving high- and low-frequency words.

We recognize that verbal comprehension reflects many factors, two of which were demonstrated in this study. Represented in Figure 1 is a model of how these two hypothetical constructs, working memory and word knowledge, might influence VSAT. This model was tested with the EQS package for structural equation modelling (Bentler, 1985) and sufficiently captured our data: $\chi^2(3) = 5.89$, $p = .11$, with Bentler-Bonet normed and non-normed fit indexes of .97 and .95 respectively. Also in Figure 1 are the standardized weights associated with the EQS Maximum Likelihood solution.

For the tasks administered, individual differences in working memory capacity are best captured by the span tasks using high-frequency words because these tasks are not confounded by individual differences in word knowledge. That is not to suggest that these tasks are the best predictors of verbal comprehension, but simply that they are purer measures of working memory capacity limitations. Alternatively, performance on tasks that use very low-frequency words is viewed as most strongly indexing a word knowledge factor. Notice that

Table 3
Correlations Between Verbal Scholastic Aptitude Test and Different Span Tasks as a Function of Word Frequency

Word frequency	*r*
Simple–high	.45*
Simple–low	.63*
Operation–high	.38*
Operation–low	.44*

Note. *n* for all correlations = 90.
*$p < .001$.

there are significant, albeit small, paths from working memory to the low-frequency word tasks. These paths are meant to reflect the assumption that all our tasks require immediate memory.

The notion that capacity and knowledge both influence comprehension is represented by the paths drawn from working memory and word knowledge to verbal SAT. These account for the assertion that tasks reflecting both these factors will be better predictors of verbal comprehension.

There is one more path in the model that has not been directly tested in this study. We assume that individual differences in working memory capacity directly influence the development of word knowledge. Working memory has been associated with the ability to disambiguate text, such as recovering from garden path passages (Daneman & Carpenter, 1983). Daneman and Green (1986) showed that working memory capacity was important in vocabulary learning, particularly in drawing inferences about the meaning of novel words from the context in which words were used.

Discussion

This study was an attempt to test one explanation for the frequently observed correlation between word span and such educationally relevant tasks as reading comprehension. We have entertained two simple explanations of this observed relationship. One is that individuals differ in a working memory capacity independent of the contents of that working memory. This basic capacity difference plays a causal role in differences in the comprehension of discourse in the ways described by Just and Carpenter (1980) and Kintsch and van Dijk (1978). An alternative explanation for the correlation between word span and comprehension, and the one tested here, is that a third factor, word knowledge, plays a causal role in both word span tasks and in the comprehension of text. This is admittedly a broad category, including many different skills and types of knowledge, but we operationalized it here in terms of basic word familiarity. Our idea was that if two subjects differed in their knowledge about words, they would perform differently if they received the same words in a word span task and in a reading comprehension task.

Our logic was that we should observe a very strong correlation between word span and comprehension if we used low-frequency words, but that the correlation should be near zero when high-frequency words were used. If the correlation in

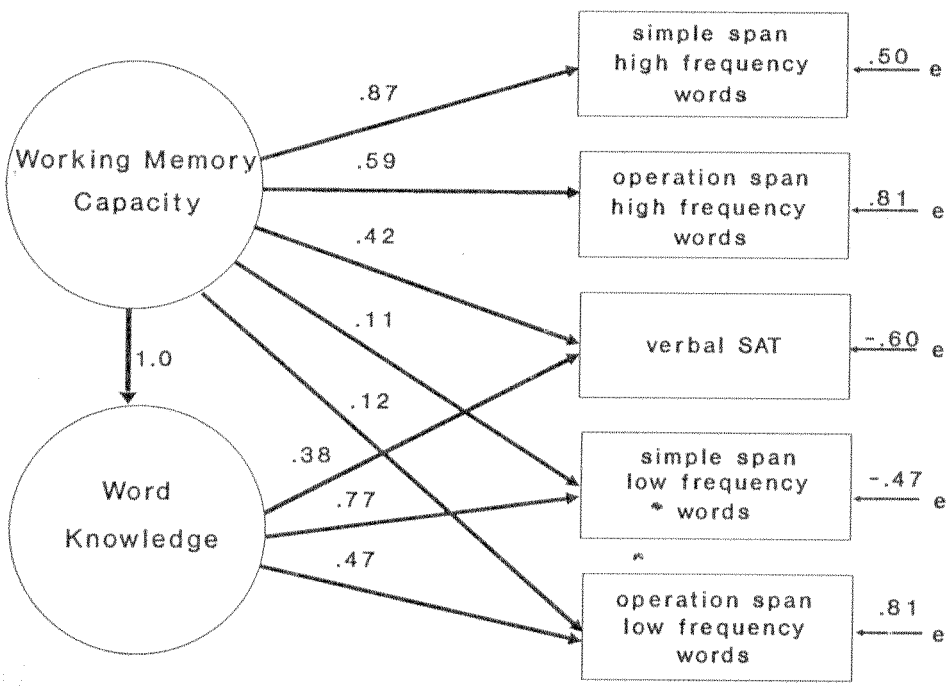

Figure 1. Structural model indicating the relationship of the span tasks and the Verbal Scholastic Aptitude Test (VSAT) to the underlying hypothetical constructs of Working Memory Capacity and Word Knowledge. (The direction of each arrow [path] represents the assumed causal direction between the construct and the measured variables. The standard coefficients above each path indicate the magnitude of the relationships. These coefficients are usually interpreted as similar to standardized regression weights.)

question is entirely a result of differential capacity in working memory, then we should see no difference between the correlations observed with low- and high-frequency words. A third possibility was that both word knowledge and working memory capacity play a causal role in both span and comprehension tasks. In that case, the correlation between span and comprehension would be high with low-frequency words and somewhat lower but still significant with high-frequency words.

This third scenario is exactly what we found. Although the correlation between the simple word span with high-frequency words and the VSAT was significant and moderately high, the correlation between the word span task with low-frequency words and VSAT was significantly higher. For the operation span tasks, argued to be a better measure of working memory than simple word or digit span, there was little difference between the correlations involving low- and high-frequency words.

It is important to point out here that the central focus of this research was on the nature of working memory and the relationship between working memory and reading comprehension. The research was not directed at estimating the relative importance of working memory capacity and word knowledge in the complex task we call reading. We argue that the results presented here lend strong support to the idea that word knowledge per se does not explain the relationship between word span and higher level cognitive tasks.

But if this relationship between working memory tasks and comprehension is caused by individual differences in working memory capacity, what is the nature of that difference and how is it manifested in the higher level tasks like reading comprehension? Although we have argued here that the difference is not a result of word knowledge, it is still possible that the relationship between working memory and higher level cognition is peculiar to verbal tasks. Baddeley (1986) reviews extensive evidence that articulatory coding plays an important role in simple word span tasks and also in some primitive higher level tasks. La Pointe and Engle (1990) presented evidence that the complex working memory tasks are also sensitive to variables that have been presumed to drive articulatory coding.

It is possible that the complex word spans measure the amount of verbal material that can be retained at a superficial level, a level that would probably be coded in some articulatory form. This possibility fits with the conclusions of Dixon, LeFevre, and Twilley (1989), who found that the reading span did not correlate with comprehension in a task that required that inferences be drawn based on a mental model of real-world knowledge. They speculated that the reading span measures the retention of surface level text and will only predict comprehension in tasks that also require a superficial understanding of the material. However, recent work from Engle, Carullo, and Collins (1989) has shown that, for college student subjects, the word span correlated most highly with inference and cause–effect questions, both of which require deeper level knowledge than fact or pronoun reference questions and both of which correlate most highly with the Verbal SAT. Thus, whether simple and complex word spans best reflect the capability to retain superficial text information or the capacity to build a deeper mental model (Johnson-Laird, 1983) awaits a resolution. It is likely that if individuals differ in their capacity to retain surface information, this would affect their ability to construct a more global mental model as well.

References

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

Baddeley, A. D., & Scott, D. (1971). Word frequency and the unit-sequence interference hypothesis in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 35–40.

Bentler, P. M. (1985). *Theory and implementation of EQS.* Los Angeles: BMDP Statistical Software.

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book.* New York: Houghton Mifflin.

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

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

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

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

Engle, R. W., Carullo, J. J., & Collins, K. W. (1989). *Individual differences in the role of working memory in comprehension and following directions in children and adults.* Manuscript submitted for publication.

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

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

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

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

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.

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

Received August 7, 1989
Revision received March 8, 1990
Accepted May 29, 1990 ■