Memory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pmem20

Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory

Andrew R.A. Conway

Version of record first published: 15 Oct 2010.

MEMORY, 1996, *4* (6), 577–590

Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory

Andrew R.A. Conway
University of South Carolina, USA

Randall W. Engle
Georgia Institute of Technology, USA

The causes of the positive relationship between comprehension and measures of working memory capacity remain unclear. This study tests three hypotheses for the relationship by equating the difficulty, for 48 individual subjects, of processing demands in complex working memory tasks. Even with difficulty of processing equated, the relationship between number of words recalled in the working memory measure and comprehension remained high and significant. The results favour a general capacity view. We suggest that high working memory span subjects have more limited-capacity attentional resources available to them than low span subjects and that individual differences in working memory capacity will have implications for any task that requires controlled effortful processing.

INTRODUCTION

In the two decades that have followed the seminal work of Baddeley and Hitch (1974), evidence supporting the relationship between working memory capacity and cognitive performance has steadily accumulated (for a review, see Engle, 1995). However, it remains unclear exactly why this relationship occurs. The purpose of the current study is to test three competing hypotheses that have been proposed to account for the relationship between working memory capacity and reading comprehension. As such, this introduction will proceed with a brief review of the three competing hypotheses.

Pascual-Leone (1970) argued that keeping schemes active requires attentional control or mental energy and that the amount of mental power or M-space increases developmentally as a result of biological or epigenetic factors. Case

Requests for reprints should be sent to Andrew R.A. Conway, Department of Psychology, University of South Carolina, Columbia, SC 29208, USA. Email: conwaya@black.cla.sc.edu

This work was supported by Grant RO1-HD27490-01A1 from the National Institute of Child Health and Human Development and Grant F49620-93-1-0336 from the Air Force Office of Scientific Research. We would like to thank Stephen Tuholski for assistance with designing the materials for the study.

(1974) extended the ideas of Pascual-Leone to suggest that differences in M-space are responsible for individual as well as developmental differences in cognition. However, he argued that increases in measured M-space do not result from an increase in attentional resources but as a result of a speed-up in mental operations as they become more automatic. The attentional resources freed by the automatisation of mental operations can be used to keep other schemes in the active portion of memory. Although the Neo-Piagetian approach has been primarily used to understand the development of cognition, the ideas may also be helpful in efforts to explain individual differences at a given stage of development. We have referred to this approach to the relationship between working memory capacity and higher-level cognition as the *general processing hypothesis* because Case (1985) viewed the operations that become automatised as general to a wide variety of tasks (Engle, Cantor, & Carullo, 1992).

Baddeley and Hitch (1974) argued that working memory is a complex system used both for the storage of information and for the computational processing of that information. They proposed the central executive as a flexible but limited-capacity work space. The central executive is used for both storage and processing—consequently, when greater effort is required to process information, less capacity remains for the storage of that information. They also proposed a variety of data representation systems including one for speech information called the articulatory loop and one for visual and spatial information called the visuo-spatial sketchpad. Both Case's theory and Baddeley and Hitch's theory propose a moment-to-moment trade-off between resources allocated for storage and resources allocated for processing.

Following the logic of Baddeley and Hitch (1974) and Case (1974), Daneman and Carpenter (1980, 1983) hypothesised that the correlation between working memory capacity and higher-level tasks like reading comprehension will only occur if the processing component of the working memory task is of the same type as is required by the higher-level task. This would lead to the same type of trade-off in the higher-level task as would occur in the working memory task. They used a measure of working memory that required both processing and storage of information. Subjects read aloud sets of sentences and, at the end of a 3–7-sentence set, they were required to recall the last word of each sentence. Daneman and Carpenter (1980) hypothesised that the processing or mental operations required to read the sentences would vary in efficiency across individuals and that a reader with more efficient processes would have more working memory capacity available for storage than would a reader with less efficient processes. Thus, good readers should recall more of the last words than poor readers because they have more automatised reading operations. We therefore call this idea the *task specific hypothesis.* Daneman and Carpenter (1980) found that the number of words recalled in the reading span measure of working memory correlated quite well with global measures of reading such as the Verbal Scholastic Aptitude Test (VSAT) as well as with more molecular measures such as the ability to correctly attribute a delayed pronominal reference.

Another possible explanation for the relationship between working memory capacity and comprehension is that high span subjects simply have more attentional resources to draw on than low span subjects, independent of the task involved. According to this view, which we call the *general capacity hypothesis*, high working memory capacity individuals will have more attentional resources to perform a task regardless of the specific nature of the task. Of course, individuals will also vary in efficiency of their mental operations in a specific task, but, other things being equal, high working memory capacity individuals will still have more attentional resources available to them than low working memory capacity individuals. Thus, there should be a relationship between working memory capacity and reading comprehension regardless of the specific processing component of the span task. All that is necessary is that the processing component place some demand on attentional resources. Turner and Engle (1989) tested this hypothesis by varying the processing component of the reading span task. Instead of having subjects read sentences, they had subjects perform mathematical operations. In this ‘‘operation span task’’, the subject performs simple mathematical operations while maintaining words for later recall. Each operation is presented with a word and after each set of operation–word strings, the subject recalls the words. This task bears much surface similarity to the reading span task except that, instead of reading, the subject performs mathematical operations. Working memory capacity or operation span is defined as the number of words the subject can recall while successfully performing the mathematical problems. Turner and Engle (1989) found that operation span correlated with VSAT as well as reading span. Furthermore, operation span and reading span accounted for about the same variance in comprehension. Engle, Cantor, and Carullo (1992) provided further support for the general capacity hypothesis in a study in which they examined performance on a moving window version of the operation and reading span tasks.

The task specific hypothesis, the general processing hypothesis, and the general capacity hypothesis all predict a correlation between reading span and VSAT. However, the hypotheses differ on two other predictions. First, the general capacity and the general processing hypotheses predict that operation span will also correlate with VSAT (Turner & Engle, 1989). The task specific hypothesis would not predict this correlation. Second, when viewing time on the processing component of the span tasks is partialled out of the correlation between span and VSAT, the general capacity view predicts that the correlation will remain significant. The task specific and general processing hypotheses both predict that partialling out viewing time would eliminate or diminish the correlation between span and VSAT.

The results of Engle, Cantor, and Carullo (1992) clearly supported the general capacity hypothesis. Significant correlations were found between reading span, operation span, and VSAT. Furthermore, when viewing time was partialled out of the correlation between span and VSAT, the correlation remained significant. Therefore, while statistically controlling for the time spent on the processing component of the span tasks, the storage component of the span tasks still predicted comprehension ability. This clearly does not support either the task specific hypothesis or the general processing hypothesis.

Our approach in the current study is similar to that of Engle, Cantor, and Carullo (1992). However, instead of statistically controlling for processing efficiency, we hoped to equate, across subjects, the processing demands of an operation span task. The logic for the experiment is simple. If the relationship between working memory span and comprehension is driven by the trade-off between processing and storage, then equating the difficulty of the span task should eliminate the relationship. In contrast, if the relationship between working memory span and comprehension is driven by attentional resources above and beyond the trade-off between processing and storage, then equating the difficulty of the span task should not affect the relationship.

In order to equate processing across subjects, we first determined each subject's capability on operations exactly like those used in the operation span task. Therefore, we had subjects perform mathematical operations of varying difficulty. From their performance on these operations, we designed three operation span tasks in which the mathematical operations were ''tailored'' to the mathematical ability of the subject.

The three hypotheses outlined earlier make different predictions regarding the correlations between our new operation span tasks and VSAT. The task specific hypothesis would not predict a correlation between VSAT and our operation span tasks with processing demand equated. This is because the view argues that individuals differ in span because of their differing ability to perform the processing component of the task. Therefore, if each subject is at the same point on the performance function for the processing component, the individual differences in the span score should disappear and the relationship between the span score and reading comprehension should disappear. Similarly, the general processing hypothesis would predict the absence of significant correlations between VSAT and our new operation span tasks with processing demand equated. This is because individual differences in span are argued to result from individual differences in the amount of operation space required by the processing portion of the task. Therefore, if each subject uses the same amount of operation space, they will each have the same amount of residual storage space for remembering words. Unlike the other two views, the general capacity hypothesis would still predict significant correlations between VSAT and our new operation span tasks with processing demand equated. This is because the view argues that individuals differ in the total amount of attentional resources available to them. Therefore, regardless of the demand of the processing component of the task, individual differences in span will remain.

METHOD

Subjects

Forty-eight undergraduates from the University of South Carolina participated in the study. All were tested individually in each of the three sessions, received course credit for participation, and signed permission for access to their Scholastic Aptitude Test (SAT) scores from university files. To ensure a wide range of comprehension skill, we chose subjects based on their Verbal SAT score. We specified five VSAT intervals; 200–340, 350–440, 450–540, 550–640, and 650–800; and chose 6, 12, 12, 12, and 6 subjects from each interval, respectively.

Materials

All the tasks reported here were conducted using an IBM PS/2 computer and a VGA monitor. The original operation span task was programmed using Turbo Pascal software. The mathematical operations and the new operation span tasks were programmed using Micro Experimental Laboratory (MEL) software (Schneider, 1988).

Procedure

Each subject participated in three experimental sessions. In the first session the subject performed the original operation span task and a backward letter task, both of which are normally administered to hundreds of subjects each semester in our lab. The backward letter task is not totally germane to the current problem but the results are presented for completeness. In the second session the subject performed a series of mathematical operations to determine the points at which they would achieve approximately 75%, 85%, and 95% accuracy. The series of operations was designed as a hierarchy in terms of difficulty. In the third session the subject performed three new operation span tasks in which the difficulty of the mathematical operations was manipulated to conform to the levels of difficulty ascertained in the second session.

Original Operation Span Task. This task was the same operation span task previously used in our lab (Conway & Engle, 1994). For each subject, a pool of 66 mathematical operations was randomly paired with a pool of 66 to-be-remembered words (taken from LaPointe & Engle, 1990). During the task, subjects were presented with operation–word strings, e.g. (8/4) + 2 = 4 ? BIRD. Each operation required the subject to multiply or divide two integers and then add or subtract a third integer, i.e. (8/4) + 2 = 4. The integers ranged from 1 to 10.

The subject was to read the operation aloud, say ‘‘yes’’ or ‘‘no’’, to indicate if the number to the right of the equal sign was the correct answer, and then say the word aloud. After the subject said the word, the experimenter immediately pressed a key, and another operation–word string was presented. This process continued until a question mark cued the subject to write the to-be-remembered words, in order, on a response sheet. The number of operation–word strings per series varied from two to six. Three series of each length were performed, and the order of series length was randomised. The first three series, each of length 2, served as practice. A subject's span score was the sum of the correctly recalled words for trials that were perfectly recalled in correct order. For example, if a subject recalled all the series of length 2 in correct order and one of the series of length 3 in correct order, their span score would be 9 (2 + 2 + 2 + 3). This score was originally reported by Turner and Engle (1989), and consistently correlates with VSAT (Cantor & Engle, 1993; Cantor, Engle, & Hamilton, 1991; Engle, Cantor, & Carullo, 1992; Engle, Nations, & Cantor, 1990; LaPointe & Engle, 1990). Each subject's accuracy on the operations was also recorded. If accuracy was below 85%, the subject was not used in the experiment.

Backward Letter Task. The backward letter task consisted of auditory presentation of strings of random letters, chosen from the pool of all consonants except w. The letters were recorded in a female voice at a rate of one letter per second and the word "recall" was spoken in the same voice after the last letter. The lists of letters varied in length from two to eight, with three trials at each length. The subject was required to write the list in the reverse order on an answer sheet. If a subject could not recall a letter, they were to leave a blank space for that letter. The same scoring procedure was used as with the operation span tasks.

Mathematical Operations. The purpose of this session was to determine each subject's performance on mathematical operations of varying difficulty (see Table 1). The subject's performance during this session determined the operations to be used in the subsequent operation span tasks. Before performing the mathematical operations, the subject was given "number recognition" trials to familiarise them with the keyboard. A number was presented in the centre of the computer screen and the subject pressed the corresponding key on the numeric keypad on the right-hand side of the keyboard. Each subject performed 20 of these trials.

Each subject then performed 375 operations in 25 blocks of 15 trials. Each block contained one operation from each of the 15 types of mathematical operations selected in a pilot study[1]. The order of presentation of the 15 types within a block was random.

[1] A pilot study with approximately 100 subjects was conducted to select the mathematical operations used in sessions 2 and 3. A series of 20 types of operations that we intuited to range in difficulty from very easy to very difficult were used. Each subject received 15 operations of each of the 20 types at a rate of three seconds per operation. The subject was to type the correct digit solution to the operation within the three-second period or the item was counted as an error. The pilot study verified the intuitive order of difficulty of the operations but found that five of the types of operations were either too difficult for our subjects to solve in three seconds or were indiscriminable from other types of operations. This left a series of 15 types of mathematical operations that ranged in difficulty from "2 + 5 = ?" to "(22 + 34)/7 = ?". These types of operations were used in the study reported here and are shown in Table 1.

TABLE 1
Types of Mathematical Operations Used

Form	*a*	*b*	*c*
(a + b)	R (1, 9)	R (1, 9)	—
(a – b)	R (1, 9)	R (1, 9)	—
(a + b + c)	R (1, 9)	R (1, 9)	R (1, 9)
(a – b – c)	R (1, 20)	R (1, 20)	R (1, 20)
(a – b – c)	R (20, 50)	R (1, 50)	R (1, 50)
(a – b)	R (51, 99)	R (a / 10 * 10, a / 10 * 10 + 8)	
(a – b)	R (50, 99)	R ((a / 10) * 10 + 1, (a / 10 – 1) * 10 + 9)	
(a / b)	b * R (2, 9)	R (2, 9)	—
(a + b – c)	R (1, 9)	R (1, 9)	R (1, 17)
(a / b)	b * R (2, 9)	R (11, 19)	—
(a + b – c)	R (11, 19)	R (11, 19)	R (13, 37)
(a* b) – c	R (2, 6)	R (2, 6)	R (3, 35)
(a* b) – c	R (7, 11)	R (7, 11)	R (48, 120)
(a / b) – c	b * R (2, 9)	R (2, 9)	R (1, 8)
(a + b) / c	temp-b where temp = c * R (2, 9)	R (2, a)	R (2, 9)

The form of the operation is followed by the range of possible integer values for a, b, and c. The values for a, b, and c were chosen such that the answer of the operation would be an integer between 1 and 9. Formation of the last operation type listed in the Table, (a + b) / c, required an algorithm that first assigned a value to c [R (2, 9)], then a temporary value to a [c * R (2, 9)], then a value to b [R (2, a)], and then a final value for a (a – b), based on the value of b.

An operation appeared on the computer screen (e.g. 2 + 3 = ?) and the subject's task was to enter the answer using the numeric keypad on the right-hand side of the keyboard within three seconds of the onset of the operation. If the subject did not respond in three seconds, the trial was scored as an error and the next trial began.

Response accuracy was recorded by the computer. If the subject made fewer than three errors (92% accuracy or better) on an operation type, that operation type was designated as the operation type to be used in the "easy" span task for that subject. If the subject made three, four, or five errors (between 80% and 88% accuracy) on an operation type, that operation type was designated as the operation type to be used in the "moderate" span task for that subject. If the subject made six, seven, or eight errors (between 68% and 76% accuracy) on an operation type, that operation type was designated as the operation type to be used in the "difficult" span task for that subject. If more than one operation type qualified for use in the span tasks (i.e. the subject responded at 100% accuracy on more than one operation type) then the operation type defined as more difficult by the pilot study was chosen as the operation type to be used in the span task. If no operation type qualified for either the easy, moderate, or difficult span task then the subject did not participate in the study.

Operation Span Tasks with Maths Difficulty Manipulated and Controlled for Each Subject. Each subject performed three operation span tasks; easy, moderate, and difficult. The procedure for each span task was exactly the same as the procedure for the original operation span task (described earlier). The only difference between the tasks was the type of mathematical operations used. For the ''easy'' span task, the subject received the operation type on which he or she made fewer than three errors in the previous session. For the ''moderate'' span task, the subject received the operation type on which he or she made three, four, or five errors in the previous session. For the ''difficult'' span task, the subject received the operation type on which he or she made six, seven, or eight errors in the previous session. The order of the three tasks was counterbalanced across subjects within each VSAT range.

Three pools of 66 high-frequency concrete nouns (taken from Carrol, Davies, & Richman, 1971) were randomised for the easy, moderate, and difficult span tasks. Therefore, an individual subject received different words for the easy, moderate, and difficult span tasks, but the same words and the same order of words were used for each subject.

In addition to obtaining each subject's span score, we recorded the time the subject spent reading the operation and word. This ''viewing time'' began when the experimenter pressed a key to present the operation–word pair and ended when the experimenter pressed a key indicating the subject had finished reading the operation–word pair, which led to the presentation of the next operation–word pair. During this time, the subject was to read the mathematical operation aloud, say ''yes'' or ''no'' to indicate whether the given answer was correct or incorrect, and say the word.

RESULTS

Descriptive statistics for the dependent measures of greatest interest are reported in Table 2. As can be seen, error rates were relatively low and varied only slightly as a function of difficulty. This was supported by a one-way repeated measures ANOVA on error rate. The main effect for difficulty was marginally significant, $F(2,90) = 2.87$, $P = 0.06$, $MS_e = 4.31$. Simple comparisons showed that significantly fewer errors were made in the easy span task (M = 1.06) than in the difficult span task, (M = 2.02) $F(1,45) = 8.02$, $P < 0.01$. No other simple comparisons were significant.

Our manipulation of difficulty was successful because subjects were slower in the difficult span task than in the moderate span task, and faster in the easy span task than in the moderate span task. This was supported by a one way repeated measures ANOVA on viewing time[2]. The main effect for difficulty was significant, $F(2,90) = 19.36$, $P < 0.01$, $MS_e = 507{,}472$ and pair-wise comparisons

[2] The viewing time data for two subjects were not recorded because of a computer error. One subject was from the 650–800 VSAT range and the other was from the 550–640 VSAT range.

TABLE 2
Descriptive Statistics

	Span		Viewing Time		Error Rate	
Easy	19.65	(10.25)	5077	(966)	1.06	(2.21)
Moderate	18.58	(11.34)	5546	(1121)	1.31	(2.34)
Difficult	18.10	(12.49)	6002	(948)	2.02	(2.25)
Original	13.21	(7.02)				
Backward Letter	35.02	(14.02)				
VSAT	507.90	(116.40)				

Mean and (standard deviation). The span and backward letter measures are the sum of the correctly recalled items for trials that were perfectly recalled in correct order. The viewing time data are in milliseconds and the error rate data are proportions.

showed that all levels of difficulty significantly differed from one another (for all, $P < 0.01$).

The number of words recalled in the operation span task did not vary as a function of difficulty. This was supported by a one-way repeated measures ANOVA on operation span. The main effect for difficulty was not significant, $F(2,90) < 1$, $MS_e = 27.85$.

We calculated reliability measures for our operation span tasks with mathematical difficulty manipulated. In each of our operation span tasks, the subject was presented with 15 series of operation–word pairs. These series varied in length from two to six operation–word pairs per series and there were three series of each length, 2, 3, 4, 5, and 6. Therefore, for each operation span task, we calculated three submeasures, each derived from five operation–word pair series of length 2, 3, 4, 5, and 6. We calculated Cronbach's alpha for the easy, moderate, and difficult span tasks based on these submeasures. Cronbach's alpha for the easy, moderate, and difficult tasks is 0.80, 0.84, and 0.84 respectively.

Intercorrelations among the span measures and VSAT are reported in Table 3. All of the correlations in the Table are significant (for all, $P < 0.01$). Performance on the original operation span task correlates highly with performance on the span tasks in which we manipulated mathematical difficulty. Also, we found the intercorrelations between the new span tasks to be highly significant. Most importantly, *all* of the span tasks; original, easy, moderate, and difficult; significantly correlate with VSAT. This suggests that individual differences in span are not accounted for by differing ability on the processing component of complex span tasks, such as operation span. These results support the general capacity hypothesis and fail to support both the task specific hypothesis and the general processing hypothesis.

Regression analyses

Although the intercorrelations among the various span measures are all considerable and significant, we can ask whether the measures account for

TABLE 3
Intercorrelations Between Span Tasks and VSAT

	VSAT	*Original*	*Easy*	*Moderate*	*Difficult*
Original	0.59				
Easy	0.62	0.54			
Moderate	0.49	0.68	0.69		
Difficult	0.54	0.68	0.72	0.82	
Backward Letter	0.44	0.41	0.33	0.43	0.37

common variance in VSAT. There are several ways we can converge on an answer to this question. One way is to use a forward selection procedure to determine the amount of new variance the measures account for in VSAT. Table 4 shows the results of the forward selection procedure. The easy operation span task accounted for 33% of the variance in VSAT, the original operation span accounted for an additional 10%, backward letter accounted for an additional 3%, and the moderate and difficult accounted for 1% additional each but the latter two were not significant. All of the measures combined accounted for 48% of the variance in VSAT but the bulk of that was contributed by a single measure, the easy operation span.

Viewing Time

To determine whether the efficiency of processing for an individual in mathematical operations played any role in the relationship between the number of words recalled and the Verbal SAT, we calculated partial correlations between VSAT and our span measures while statistically controlling for viewing time. If the general processing hypothesis is correct, these correlations should become non-significant. The partial correlations are reported in Table 5. The correlations between the span tasks and VSAT remain virtually unchanged and, obviously, significant when viewing time is partialled out (for all, $P < 0.01$). Therefore, the significant correlations between the span tasks and VSAT are not due to the amount of time required to process the operation–word pair.

TABLE 4
Results of Regression Analyses of Variance in VSAT

Variable	*Partial R^2*	*Model R^2*	*F*	*P*
Forward Selection Results				
Easy	0.33	0.33	21.24	0.0001
Original	0.10	0.43	7.60	0.009
Backward letter	0.03	0.46	2.30	0.14
Moderate	0.01	0.47	1.10	0.30
Difficult	0.01	0.48	0.59	0.45

TABLE 5
Correlations Between VSAT and Span Tasks Before and After Partialling Out Viewing Time

	Original	*Easy*	*Moderate*	*Difficult*
VSAT (before)	0.59	0.62	0.49	0.54
VSAT (after)	—	0.60	0.48	0.52

DISCUSSION

The purpose of this study was to investigate the relationship between working memory capacity and reading comprehension, and to provide a test of three competing hypotheses proposed to account for this relationship. We used the operation span task because it was possible to systematically vary the difficulty of the processing component of the task. We equated the processing demand of the operation span task across subjects and systematically manipulated the level of difficulty across three conditions. The correlations between these three conditions and reading comprehension, as operationalised by VSAT, ranged from 0.49 to 0.62 and did not differ statistically from the original version of the operation span which correlated 0.59 with VSAT—for all pair-wise comparisons, $t(45) < 1.43$, $P > 0.10$. Further, these correlations were undiminished when we partialled out the time that subjects spent viewing the operation–word string. The general capacity hypothesis can explain these results but the task specific and general processing hypotheses cannot.

The general capacity model of working memory was first proposed by Engle, Cantor, and Carullo (1992). The model assumed that working memory consists of knowledge units in long-term declarative memory which are currently active beyond some critical threshold. The model also assumed that knowledge units vary in their level of activation and that the total amount of activation available to the system is limited. The total amount of activation available to each individual varies, and it is this variance that causes individual differences in working memory capacity. Cantor and Engle (1993) provided support for the general capacity model by reporting that the amount of activation available to long-term memory, as measured by the fact retrieval task (Anderson, 1974), statistically accounted for the correlation between operation span and VSAT.

A recent study conducted in our lab (Conway & Engle, 1994), however, has convinced us that it is not sufficient to simply say that high- and low-span subjects differ in the total amount of activation available to them. A further qualification for the general capacity model is that individual differences will only reveal themselves in tasks that force the subject to engage in controlled effortful processing. If the task allows for automatic processing, then the limited-capacity resource we call working memory will not be taxed. Indeed, Conway and Engle (1994) found that individual differences in working memory capacity were important in a memory search task that required controlled processing, but were not important in a memory search task that allowed for automatic processing. Thus, we now believe that individual differences on the complex WM measures correspond to differences in general, controlled, effortful, attentional resources.

The question remains, why do we find that operation span predicts VSAT, even when the processing demand of the task is equated for each subject? The operation span task, regardless of the demand of the processing component, requires the subject to switch attention constantly from one aspect of the task to another. Subjects must perform a mathematical operation and then encode a word, perform a mathematical operation, and encode a word, and so on, until they are asked to recall the words. This type of attention switching requires the subject to engage in controlled effortful processing. We agree with Baddeley and Hitch (1974) and Daneman and Carpenter (1980), that tapping both processing and storage is necessary for a span task to be a good measure of a central executive or working memory capacity. However, we argue that it is not the demand of the processing component that is critical. We argue that the simple existence of a processing component beyond the storage component is what is required for a span task to be a good measure of working memory and a good predictor of more complex cognitive behaviour, such as reading comprehension. Of course, the processing component has to be demanding enough that it forces the subject to shift attention away from the storage component of the task and to engage in controlled effortful processing. Support for this argument comes from our finding that viewing time was a function of level of difficulty but the number of words recalled was not. Subjects spent nearly one second longer processing the operation–word pair in the difficult span task than in the easy span task, yet the number of words recalled in each task was not statistically different. If the demand of the processing component of the task was the critical determinant of span, then we should have found the number of words recalled to be a function of the level of difficulty of the operations. However, if attention switching is the critical determinant of span, as we argue, then level of difficulty will not have an effect on the number of words recalled, as we found.

Towse and Hitch (1995) recently reported evidence in support of an attention-switching interpretation of developmental differences in performance on the counting span task. They independently manipulated counting difficulty and counting time in the counting span task and found that difficulty did not have an effect on span when time of counting was controlled. They argue, as we have, that performance on span tasks such as reading span, operation–word span, and counting span is not driven by a trade-off between resources allocated to processing and storage. Their view differs from ours however, in that they argue that the timing of the processing component of the task is critical to span performance. Thus, according to their view, span performance is driven by the time spent away from the storage component. The attention-switch itself is not critical; the time between successive switches is.

Our data do not support their view. We found that viewing time was a function of level of difficulty, but the number of words recalled was not. That is, subjects spent longer on the processing component in the difficult span task than in the easy span task, yet the number of words recalled in the two tasks was not statistically different. Furthermore, when we partialled viewing time out of the correlations between span and VSAT, the correlations remained virtually unchanged. Thus, we argue that the critical component of the task is the attention-switch itself, not the trade-off in resources, and not the time spent processing.

One potential problem with our procedure is that the mathematical ability of each subject was tested under strict time constraint whereas the operation span task is subject-paced. In the mathematical operations session, the subject was only allowed three seconds to answer each mathematical operation. In the subsequent operation span tasks, the subject read the operation aloud at his or her own pace. One may argue that the nature of the processing underlying these two tasks is quite different due to the differing time constraints. However, before the operation span tasks, we encouraged our subjects to perform the operations as quickly as possible without sacrificing accuracy. Also, if the subject appeared to be performing the operations slowly in the operation span tasks, the experimenter encouraged them to perform the operations more quickly. Therefore, we do not feel that this difference in procedure contaminated the outcome of the experiment.

In conclusion, we argue that working memory is a very general resource which plays a role in a wide variety of cognitive tasks. Furthermore, we hope the current article makes the point that it is not sufficient simply to identify a relationship between working memory and some aspect of cognition. We must go beyond the identification of the relationship and investigate exactly why the relationship occurs. Only then will we be able to understand fully the role of working memory in normal human information processing.

Manuscript received 29 August 1995
Manuscript accepted 8 January 1996

REFERENCES

Anderson, J.R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*, 451–474.

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

Cantor, J., & Engle, R.W. (1993). Working memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *5*, 1101–1114.

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

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

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.* New York: Academic Press.

Case, R., Kurland, M.D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386–404.

Conway, A.R.A., & Engle, R.W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, *123*, 354–373.

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.

Engle, R.W. (1995). Individual differences in memory and their implications for learning. In R. Sternberg (Ed.), *Encyclopedia of intelligence.* New York: Macmillan.

Engle, R.W., Cantor, J., & Carullo, J.J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 972–992.

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.

LaPointe, 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.

Schneider, W. (1988). Micro-Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, and Computers*, *20*, 206–217.

Towse, J.N., & Hitch, G.J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *The Quarterly Journal of Experimental Psychology*, *48*, 108–124.

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