

Children's Working-Memory Processes: A Response-Timing Analysis

Nelson Cowan
University of Missouri—Columbia

John N. Towse
Lancaster University

Zoë Hamilton
Royal Holloway University

J. Scott Sauls, Emily M. Elliott, Jebby F. Lacey,
and Matthew V. Moreno
University of Missouri—Columbia

Graham J. Hitch
University of York

Recall response durations were used to clarify processing in working-memory tasks. Experiment 1 examined children's performance in reading span, a task in which sentences were processed and the final word of each sentence was retained for subsequent recall. Experiment 2 examined the development of listening-, counting-, and digit-span task performance. Responses were much longer in the reading- and listening-span tasks than in the other span tasks, suggesting that participants in sentence-based span tasks take time to retrieve the semantic or linguistic structure as cues to recall of the sentence-final words. Response durations in working-memory tasks helped to predict academic skills and achievement, largely separate from the contributions of the memory spans themselves. Response durations thus are important in the interpretation of span task performance.

In working-memory (WM) span tasks, participants carry out sets of problems requiring processing and then recall target items that accompanied the problems, with one target item for each problem. WM-span measures are based on how many items can be recalled in order, after the corresponding problems have been correctly completed. No matter whether the processing task involves sentences (Daneman & Carpenter, 1980), counting (Case, Kurland, & Goldberg, 1982), arithmetic (Turner & Engle, 1989), or spatial cognition (Shah & Miyake, 1996), WM span has been shown to correlate rather highly with various cognitive abilities as measured by standardized tests in adults and children, typically better than traditional short-term memory (STM) span tasks that involve the storage and retrieval of lists of items but no supplementary processing (Daneman & Merikle, 1996). Given this correlation, the

field stands to benefit if the demands of the WM-span tasks can be clarified. This clarification, however, has proved to be difficult to obtain on the basis of span measures alone.

The present article uses measures of the timing of recall in reading-span, listening-span, counting-span, and standard digit-span tasks in children and adults to clarify the processes involved in task performance. In order to explain what this investigation can contribute, we introduce in turn four specific topics that shaped our investigation: (a) WM span in children and its development, (b) the need for multiple WM measures, (c) the timing of responses in STM and WM-span tasks, and (d) the relation of WM and timing to complex skills.

WM Span in Children and Its Development

Children make up an interesting population for an analysis of WM-span performance because of the considerable practical implications related to education and development. Recent studies point to relations between impairments on various tests of WM and assorted types of learning disabilities (e.g., Gathercole & Baddeley, 1990; Henry, 2001; Jarrold, Baddeley, & Hewes, 1999; McLean & Hitch, 1999; Swanson & Sachse-Lee, 2001a, 2001b) as well as correlations with abilities tests among normal children (e.g., Gathercole, Willis, Emslie, & Baddeley, 1992; Hitch, Towse, & Hutton, 2001). There is recent interest, within the field of neuroimaging, in children's developing frontal lobes and WM (Nelson, 1995; Yeo, Hill, Campbell, Vigil, & Brooks, 2000) and, within the field of cognitive psychology, in mechanisms of WM in children that cut across stimulus domains (e.g., verbal and spatial domains) and those that are specific to one of these stimulus domains (Hitch et al., 2001; Swanson & Sachse-Lee, 2001b). The present study with children is relevant to these concerns as it

Nelson Cowan, J. Scott Sauls, Emily M. Elliott, Jebby F. Lacey, and Matthew V. Moreno, Department of Psychology, University of Missouri—Columbia; John N. Towse, Department of Psychology, Lancaster University, Lancaster, United Kingdom; Zoë Hamilton, Department of Psychology, Royal Holloway University, Egham, United Kingdom; Graham J. Hitch, Department of Psychology, University of York, York, United Kingdom.

This research was supported by National Institutes of Health Grant R01 HD-21338 in the United States and by Economic and Social Research Council Grant R000222789 in England.

Correspondence concerning this article should be addressed to Nelson Cowan, Department of Psychology, University of Missouri—Columbia, 210 McAlester Hall, Columbia, Missouri 65211, or to John N. Towse, Department of Psychology, Fylde College, University of Lancaster, Lancaster LA1 4YF, United Kingdom. E-mail: CowanN@missouri.edu or j.towse@lancaster.ac.uk

contributes to an understanding of the processes taking place in WM tasks. Moreover, an investigation of the childhood development of these WM processes (in Experiment 2) can touch on possible implications of age changes in speeds of processing (Cowan et al., 1998; Kail & Salthouse, 1994) for the mechanisms of WM-task performance. The present study explores a new methodology, timing of recall in WM-span tasks, that is applicable to almost any population.

Need for Multiple WM Measures

Multifaceted Nature of WM

Multiple measures of WM performance are needed if, as various investigators have claimed, WM is multifaceted in nature. Investigators differ in their definitions of WM (see Miyake & Shah, 1999). However, all views seem to hold in common (sometimes explicitly, other times implicitly) that the WM-span tasks measure something that is critical for success in various tests of applied cognitive abilities. In particular, WM capacity presumably reflects the amount of information that can be retained, while processing is carried out, for short-term recall. The application would be that various cognitive abilities (sentence comprehension, arithmetic ability, etc.) depend on holding data during processing. A linguistic example is the information given early in a sentence to be commented upon by the speaker later in the sentence, and an arithmetic example is an interim result derived on the way toward solution of a problem.

Understanding WM-span tasks has been difficult in part because performance depends on both specific skills that differ by domain and general skills that cross domains. Several studies examining individual differences in adults have suggested that there may be a general component of ability that is common across different WM tasks, as well as unique components of specific tasks (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Shah & Miyake, 1996). For example, using a latent variable approach, Engle et al. found that variance in common between WM-span tasks and standard memory-span tasks did not predict *g* on the basis of intelligence tests, whereas variance unique to WM tasks did predict *g*. Conway et al. (2002) replicated that finding and further found that the variance in common between WM and standard memory-span tasks was related to processing speed as reflected in several separate measures. Several studies (Hitch et al., 2001; Towse, Hitch, & Hutton, 1998, 2000) have shown that prolonging the time that information has to be held during processing in a WM-span task results in poorer performance, and a faster speed of processing potentially could minimize the delay.

Skill Profiles and Multiple Measures

Given that WM-task performance appears to depend on multiple traits (e.g., ability in the domain of the processing portion of the WM test, *g*, and the speed of various processes), it is possible that apparently equivalent WM-span performance in two individuals can be based on different profiles of skills. For example, one individual might achieve a relatively high reading span because of a high *g* factor score, whereas another individual might achieve the same span with a lower *g* factor score but good linguistic skills.

Still another individual might obtain the same reading span because of exceptionally fast processing.

Given that WM-span scores can be ambiguous, it may be useful to obtain multiple indices of performance within an individual. The predominant method of doing so in the field of WM has been to administer multiple tasks and correlate them. However, there have been a few attempts to examine multiple measures of processing within a WM-span task, including measures of the speed of item processing in the task (e.g., Engle, Cantor, & Carullo, 1992; Hitch et al., 2001). The present research with children extends this within-task measurement approach further by examining the times taken to complete spoken recall. The basic promise of using timing measures is that the pattern of recall timing can provide clues to the processes leading to correct recall. Such timing measures have already been applied usefully in standard span tasks; the next section reviews this evidence in order to make clear the potential importance of these measures for an interpretation of WM span.

Timing of Responses in STM and WM Span Tasks

The present studies use the timing of recall to provide clues to strategies used in WM-span and standard span tests. The basic notion is that the timing of processing and recall is determined by more than simply a global speed of processing (e.g., Kail & Salthouse, 1994), and that if different strategies are used, they can result in different specific patterns of response times.

Cowan (1992) described how one could use response timing at two levels of analysis in the interpretation of serial recall task performance. At a macroscopic level, for all correct list repetitions, one can simply examine the entire amount of time from the end of the stimulus list to the end of the response. At a microscopic level, one can distinguish between the durations of preparatory intervals, from the end of the list to be recalled to the beginning of the spoken response, word durations, and interword pauses between words in the response. Serial recall timing is likely to be affected by a number of processes including, for example, rehearsal, response planning, memory search, and reintegration (Cowan, 1992; Cowan et al., 1994, 1998; Hulme, Newton, Cowan, Stuart, & Brown, 1999; Sternberg, Monsell, Knoll, & Wright, 1978; Sternberg, Wright, Knoll, & Monsell, 1980). As we explain shortly, the different segments of the response appear to be differentially sensitive to different processes in recall. Thus, the microscopic analysis of response timing may be of considerable importance in understanding WM performance.

Nonetheless, we recognize that analysis at the macroscopic level has its place also. Insofar as a general speed of processing (Kail & Salthouse, 1994) may affect all response segments, the total response duration may be the most sensitive measure of this processing speed. Furthermore, participants in an experiment have options as to how and when to engage in processing. For example, they could assemble the complete sequence during the preparatory interval followed by rapid verbalization, or adopt an iterative retrieval process before each response word. Consequently, the overall response duration provides a simpler index of recall processes.

To appreciate how response timing could be of use, it is helpful to review what has been learned from measurements of response timing in STM-span tasks and then to consider these findings in relation to theoretical notions regarding WM span. In this past

research, the emphasis has been on the microscopic level of analysis.

Response Timing in STM Span Tasks

Response timing could be measured simply if the memory task required keypresses. Yet this yields data at a course level only; it does not distinguish between response planning and execution. In fact, Cowan (1992) found that, when spoken responses were analyzed directly, preparatory intervals, word durations, and interword pauses within correctly recalled lists yielded complementary information. Several subsequent studies have gone on to sketch out additional details of spoken response timing (Cowan, 1999; Cowan et al., 1994, 1998; Doshier & Ma, 1998; Hulme et al., 1999; Jarrold, Hewes, & Baddeley, 2000; Tehan & Lalor, 2000).

Recent studies of the timing of recall in STM-span tasks led to several conclusions that enhance our understanding of these tasks and are of potential relevance to WM tasks:

1. Interword pauses may reflect memory-search and retrieval operations in STM tasks. Cowan et al. (1998) found that pauses, but not the preparatory intervals, increase as list length increased (Cowan et al., 1998). The proposal has been that there is a memory search through the entire list repeatedly during interword pauses, in order to select each item to be recalled in turn. (The search might go on to some extent also during the pronunciation of words.) This search process may be the basis of the individual differences in pauses. The processing taking place during preparatory intervals is thought to be more complex, potentially including rehearsal of the list and motor programming as well as memory search.

Reinforcing the notion that a search process occurs during pauses in STM tasks, the influences on pauses parallel findings in memory search using probed recall (Chase, 1977; Clifton & Tash, 1973; Sternberg, 1975) or rapid pronunciation of lists (Jarrold et al., 2000; Sternberg et al., 1978, 1980). As in these other procedures, the memory-search process does not appear to be one that relies on verbal rehearsal, inasmuch as the pauses last no longer for lists composed of multisyllabic words than for lists composed of monosyllables (Cowan et al., 1994). The interword pauses and word durations do not change much as a function of serial position in the list except for sometimes shortening at the last serial position, so it seems unlikely that response alternatives can be eliminated from the search after they are used (Cowan et al., 1998). Interword pauses are much longer for lists composed of nonwords than for lists composed of English words, suggesting that lexical knowledge is used in the retrieval process (Hulme et al., 1999; see also Tehan & Lalor, 2000).

2. Interword pauses are sensitive to age differences as well as to individual differences within an age group. For lists of a particular length, preparatory intervals, word durations, and interword pauses in the memory responses all decrease with age. Word durations and interword pauses,

and especially the pauses, have been shown to be shorter within children of a particular age who have higher memory spans (Cowan, 1992; Cowan et al., 1998).

3. Response timing is a composite of separate processes and individual differences reflect this heterogeneity. Cowan et al. (1998) found that although interword pauses in responses and speeded-speaking estimates of rehearsal times (cf. Baddeley, 1986) were both correlated with span, they did not correlate with one another. Together, pauses and rehearsal speeds accounted for a large proportion of the variance in memory span. In a further analysis of the digit-span task of Cowan et al. (1998), Cowan (1999) showed that the two types of speed measures produced different patterns of correlations across ages. Rapid-speaking rates predicted span in first-grade children but not in older children, whereas interword pause durations predicted span in fifth-grade children but not in younger children. These two different timing measures appear to be influenced by STM rehearsal and retrieval operations, respectively (Cowan et al., 1998).
4. Although higher span individuals do recall sequences at a faster rate, when comparing the longest sequence that can be managed their responses take longer than those of lower span individuals (Cowan, 1992; Cowan et al., 1998; Tehan & Lalor, 2000). It is not the case that the advantage of high-span people rests solely on squeezing more responses into the same temporally limited window of opportunity.

Response Timing and WM Span Performance

A key empirical contribution of the present article is to adapt measures of the timing of responses developed in STM-span tasks to examine WM-span tasks. We examined spoken recall response timing in a reading-span task within a group of children from a narrow age range (Experiment 1) and in several other span tasks using children of two age groups and adults (Experiment 2). We also obtained the results of standardized tests of reading ability, arithmetic ability, and nonverbal intelligence in children and several academic tests and credentials in college students.

Both STM and WM spans involve recall of a list of items (typically words). In the case of verbal-STM span, these items presumably must be reconstructed on the basis of a residual phonological memory representation of the words (Baddeley, 1986) along with lexical knowledge that can help in an interpretation of degraded memories (Gathercole & Hitch, 1993; Hulme, Maughan, & Brown, 1991; Schweickert, 1993). If similar patterns of recall timing in WM-span and STM-span tasks are obtained, this would suggest that similar sets of processes may be used in the two situations. However, this may not occur because the demands of WM-span tasks differ from those of standard STM-span tasks in several ways. In reading span, for example, there is much more time and opportunity for interference intervening between the initial presentation of a word and the time when it has to be recalled. There are also, logically speaking, multiple possible routes of recall. The phonological memory representation of a word could be kept active, perhaps through rehearsal, throughout

the sentence-processing phase of the trial. Alternatively, the participant might reinstate the word at the time of recall from remembered information about the sentence. The response timing is of theoretical interest because interword pauses equivalent to those found in STM-span tasks would suggest that the words tended to remain activated, whereas much longer pauses would suggest that they were allowed to become inactive and that a relatively lengthy reactivation process was necessary.

Relation of WM and Timing to Complex Skills

Given that WM is related to complex cognitive task performance, correlations between recall durations and the criterion academic measures could help to clarify the nature of those relations by indicating how important retrieval speed may be within this complex task performance. In standard STM-span tasks, the interword pauses in the recall period are indicative of the child's efficiency of retrieval and are correlated with span (Cowan et al., 1998). Given that some studies have suggested that the duration of memory retention is important for recall in WM tasks (e.g., Towse et al., 1998, 2000), the speed of retrieval may be an important factor affecting recall; faster recall allows less time for forgetting of the not-yet-recalled portion of the memoranda. Differences in the speed of recall between children with better WM spans and those with poorer WM spans, and the relation of recall speed to scholastic abilities, would help to determine whether the retention-duration hypothesis is correct.

In sum, a new type of measure for WM tasks, recall timing, may provide evidence about several aspects of the nature of processing. Of particular relevance are (a) the timing of recall compared with STM-span tasks, which could indicate similarities as well as differences, such as more extensive reactivation processes in WM-span tasks, and (b) what measures of the timing of recall can contribute to the prediction of scholastic skills, both in terms of accounting for variance previously attributed to WM span and in terms of accounting for additional variance beyond what can be predicted by WM span alone. Both of these can contribute to an understanding of the processes involved in WM-span performance.

Experiment 1

In Experiment 1, children were tested with a reading-span task for an initial investigation of recall response timing and its relation to span and to scholastic skills.

Method

Participants

Data on reading span were collected for 62 children from schools in southwest England. The mean age of children was 101 months (i.e., 8 years, 5 months), ranging from 94 to 113 months ($SD = 4.32$).

Apparatus, Stimuli, and Procedure

Reading span. Task instructions were given on laminated practice cards, then the reading-span test was administered individually on an Apple Powerbook 5300c. The experimenter pressed a key on the keyboard to begin the experimental trial. The child was to read aloud a sentence that was visible on the screen. The final word of the sentence was missing (represented by an empty underline symbol), and children were instructed

to think of, and pronounce, an appropriate word. The computer selected sentences at random, without replacement, from a pool of 88 items of medium length from the corpus described in Towse, Hamilton, Hitch, and Hutton (2000). The sentences were constructed so that the missing target word had a high probability of being produced, for example, *Mary got home and unlocked the (door), Ben laughed and then clapped his (hands)*, with the expected but not presented completion word shown in parentheses in these examples.

The experimenter recorded the child's response on the computer either by entering a designated keystroke, in the case of expected response words, or by entering unexpected response words in full. The use of an external keyboard permitted unobtrusive data entry with the keyboard in the experimenter's lap. The computer presented the next experimental event on the screen immediately after the experimenter's first keystroke (even when the experimenter typed the full word). After words had been supplied for each sentence in a trial, children were asked to recall these items in the presented serial order. The computer provided feedback to the child on item-by-item recall success and sentence-processing times on each trial. (A screen appeared with a box for each serial position. A green-colored box indicated a correct answer; a red-colored box, an incorrect answer. For incorrect answers, the originally generated and recalled answers were repeated in order to emphasize the mismatch.)

Children received three trials at each reading-span test length. Each child began with two sentences to read and therefore two words to remember. Provided that the child recalled at least one trial correctly, the sequence length was increased by a single item and three further trials were administered. When the child successfully recalled at least two lists of a particular length correctly, a visual and auditory message of congratulations was presented.

Scholastic attainment. Children completed three subscales of the British Abilities Scale II tests (BAS; Elliot, Smith, & McCulloch, 1997). Word Reading and Matrices were tested individually; some orally administered Number Skills items were given individually, whereas the written questions were completed separately in a group class setting. The Number Skills test emphasizes arithmetic computation skills (including number reading, number identification, and various arithmetic problems, some of which require the carryover of partial results). The Word Reading test involves graded single-word reading, and the Matrices test is also a graded assessment that has similarities with Raven's Progressive Matrices. Because of an error in publication, the items for Question 14 were not included in the Matrices test book, an error that was not noticed until testing commenced. Consequently, tests had to be administered without this item, lowering the overall values but making little difference in the relative distribution of scores.

Order of tasks. The separate, written Number Skills assessment was taken in one pre-session. After that, on a separate day, another session included the BAS scales followed by a reading-span test.

Dependent Measures

Scholastic attainment. The raw scores from the BAS tests were converted into ability measures according to standard procedure (Elliot et al., 1997).

WM span. Two different measures of WM span were used. One measure (which we term *maximal span*) equaled the number of sentences in the longest stimulus set resulting in correct recall of the self-generated words, in the correct order. This measure indicated the list lengths available for recall-response timing, given that the timing was to be restricted to trials in which recall was correct.

A second, more sensitive measure (which we term *aggregate span* or *Span-A*) was obtained in which performance across trials was considered. First, a base span was taken as the highest list length at which the responses for all three trials were correct. Next, a partial score of 0.33 was added to this base span for every list of a higher length that was correctly recalled (given that three lists of each length were presented). For example, if a

participant correctly recalled 3 two-word lists, 2 three-word lists, and 1 four-word list, a Span-A of $2 + .066 + .033 = 3.00$ would be awarded. If no list length produced three correct list recalls, a Span-A of 1.0 was awarded.

Response-timing measures. The durations of segments of the recall response were recorded on audiotape and later analyzed using a speech waveform editor on a Macintosh computer. As in STM studies of recall timing (e.g., Cowan, 1992), each preparatory interval, word duration, and interword pause in every response was measured by computer, using both the sound and the oscillographic display of each segment.

A second rater measured the responses of 27 randomly selected participants; correlations between participant means based on the two raters were calculated. These correlations were, for the total response duration, $r = 1.0$; for preparatory intervals, $r = .99$; for the first word, $r = .69$; for the following interword pause, $r = .99$; and for the second word, $r = .63$. The lower correlations for words than for silent periods occurred because the absolute magnitudes of interrater discrepancies were comparable in the two cases, even though word durations were much shorter than pause durations, making the discrepancies proportionally larger for words.

Results

The subsections of the results focus on (a) the means and variability in all tasks, (b) the overall pattern of the timing of recall in comparison to the STM literature, (c) individual differences in WM span and recall timing, and (d) the contribution of recall to the prediction of scholastic skills. Points b and d are of greatest theoretical interest, but Points a and c are necessary to appreciate Points b and d.

Means and Variability

Table 1 shows the means and standard deviations on span and BAS variables for the entire sample ($N = 62$, left column) and these plus timing on two-item trials for the children with timing measures on multi-item responses ($N = 53$, right column). The children's maximal spans included 7 with a span of one item, 34 with a span of two, 17 with a span of three, and 4 with a span of four. One child with a span of two and 1 with a span of three produced inaudible tapes that did not allow timing of responses. Of course, children with a span of one had no correct multi-item responses to be timed. Thus, timing measures will concentrate on children with spans of two and three.

Because of outliers in the data, we set a boundary of 20 s for the total response duration through the second word or 25 s through the third word and, except where otherwise noted, we eliminated all trials with responses longer than those. This procedure eliminated six trials, from 6 different participants, at List Length 2 and none at List Length 3.

Pattern of Timing of Recall in Comparison to STM Studies

Segment durations. Although the word durations were very comparable with what has been found in responses to STM-span tasks (e.g., Cowan et al., 1994, 1998), the preparatory intervals shown in Table 1 were several times larger, and the interword pauses were an order of magnitude larger. By way of comparison, Cowan et al. (1998) found that 9- to 10-year-old children, slightly older than those in the present sample, produced correct responses to two- and three-digit lists with mean preparatory intervals of 0.81

Table 1
Means of Key Measures in Experiment 1

Measure	$N = 62$		$N = 53$	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Span				
Maximum	2.29	0.76	2.45	0.64
Aggregate	1.85	0.52	1.97	0.45
Span response timing for two-item lists (in s)				
Preparatory interval			3.31	2.12
<i>M</i> word duration			0.46	0.11
Interword pause duration			2.38	1.57
Total response duration			6.61	3.42
British Abilities Scales				
Word Reading	129.50	27.17	134.25	24.47
Matrices	86.15	17.70	88.68	16.80
Number	104.48	12.68	106.57	11.17

Note. $N = 62$ includes the entire sample; $N = 53$ includes only children who had usable timing data for two-item lists.

and 0.68 s, mean word (digit) durations of 0.49 and 0.52 s, and mean pause durations of 0.15 and 0.20 s. This discrepancy between STM-span and reading-span response durations is not attributable to the use of digits as stimuli, given that Cowan et al. (1994) obtained similar results with word stimuli (although the results were presented in a more complex fashion in that study, not by absolute list lengths). The much longer silent intervals within responses in the present reading-span task confirm that much more processing occurred at retrieval than is the case for STM-span tasks.

List-length effects. As has been found with STM span in the past (e.g., Cowan, 1992; Cowan et al., 1994, 1998; Tehan & Lalor, 2000), segments were not squeezed together into a constant period in the recall of longer lists. The rate of correct recall instead tended to slow down across list lengths. As in these STM-span studies, interword pauses in reading span increased as a function of list lengths within individuals. We examined 20 children who had correct trials for two- and three-item lists and found that the first interword pause was shorter for two-item trials ($M = 2.04$ s) than for three-item trials ($M = 4.11$ s), $F(1, 19) = 9.59$, $MSE = 4.46$, $p < .01$, $\omega^2 = .18$. (Effect sizes are described by partial omega squared, which we denote as ω^2 . This statistic is unaffected by which other factors happen to be included in the analysis and follows the recommendation of Keppel, 1991.) For three-item trials, the second pause duration ($M = 3.64$ s) was not significantly different from the first. The list-length effect suggests that in reading span, as in STM span, memory search through the entire list (or a proportion of it) occurs during interword pauses.

A comparable analysis on preparatory intervals or on the duration of the first word did not approach significance. (As in Cowan et al., 1998, the duration of the final word tends to be shorter than other words in the list.) Thus, the relevant effect of list length on segment durations in this task appears to be restricted to interword pauses.

Individual Differences in WM Span and Recall Timing

In STM-response timing, interword pauses within correctly recalled lists of a particular length are shorter for participants with

higher spans (e.g., Cowan et al., 1998). The pattern of results basically confirmed that interword pauses were shorter for children with a span of 3 ($N = 16$) than for children with a span of 2 ($N = 33$). All trials in which the response was correct (including outliers) were included. Means (and *SDs*) for the durations of the preparatory interval, first word, interword pause, and second word, respectively, were, for Span 2 children, 4.88 s (4.82), 0.53 s (0.19), 2.78 s (1.51), and 0.40 s (0.10); and for Span 3 children, 2.89 s (1.50), 0.47 s (0.11), 1.84 s (0.98), and 0.42 s (0.10). Separate one-way analyses of variance (ANOVAs) on the three kinds of segment (preparatory intervals, words, and interword pauses) were conducted because of the very different magnitudes and variances of the three kinds of segments in the response. The difference between children with high and low spans was significant only for the interword pauses, $F(1, 47) = 5.10$, $MSE = 1.85$, $p < .03$, $\omega^2 = .08$, and not for preparatory intervals or word durations.

With outliers in response times omitted, the ability-level difference discussed above was no longer significant. The change occurred because 5 of the 6 children who made these long responses had both a Span-A and a maximal span of only 2.0 (i.e., fairly low). Thus, the long responses appear to reflect actual characteristics of the children. It was also characteristic of the standard span data (e.g., Cowan et al., 1998) that ability differences were seen most strongly in some unusually long response times in less able children, although that point was not emphasized in previous reports. From here on, data will be reported with the outlying trials eliminated.

The issue of ability differences can also be examined among the children with higher memory scores by carrying out a tripartite split on the basis of Span-A among 20 individuals who correctly recalled at least 1 three-item list. They were split along natural lines into 6 children with a Span-A of 2.0 or lower, 7 with a span equal to 2.33, and 7 with a span of 2.66 or more. The result can be observed in Figure 1. Separate analyses for each type of segment resulted in no effect of span group for preparatory intervals, $F(2, 17) < 1$, $MSE = 2.98$, or words, $F(2, 17) < 1$, $MSE = 0.02$, but there was such an effect for interword pauses, $F(2, 17) = 6.14$, $MSE = 5.13$, $p < .01$, $\omega^2 = .20$. It was clear that the lowest span group paused longer than the other two groups, which did not appear to differ.

Relation Between WM Span, Timing Measures, and Scholastic Skills

Correlations between various measures are shown in Table 2. The correlations below the diagonal include only children with timing measures, whereas the correlations above the diagonal include all children in the sample. As expected, span measures were correlated with all of the BAS skills. The correlation with Matrices was significant for the full sample, but not for the sample with response timing because it distinguished primarily between children with a span of one item versus more advanced children.

Table 2 shows correlations between response timing measures and BAS skills. Note that, in contrast to the prediction of span, for which it was specifically the pauses that seemed to matter, in the prediction of scholastic skills the correlations were rather general across response segments; preparatory intervals, words, and pauses

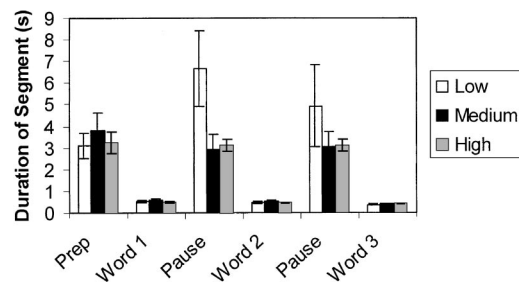


Figure 1. Experiment 1: Mean durations (in seconds) of segments within correct responses to three-item lists in children with the lowest span (white bars), medium span (black bars), or highest spans (gray bars) among those who were correct on at least 1 three-item list. Error bars are standard errors.

all predicted BAS Word Reading, and preparatory intervals predicted BAS Number skills better than pauses. Therefore, it seems simplest to rely on the total response durations in examining correlations with other measures. Note that the total response duration was significantly correlated not only with BAS Word Reading, $r = -.51$, but also with BAS Number, $r = -.30$.

Given the correlations between WM span, response durations in the span task, and BAS skills, we examined how much of the variance in span and response durations was related to domain-specific skills and how much was related to skills that cross domains (specifically, skills used in both Word Reading and Number tasks). To examine this question, in a set of regression analyses with Word Reading as the dependent variable, we used Span-A as the representative of span and the total response duration as the representative of memory response timing. We also included BAS Number skills in order to examine the common processes between Number and Word Reading tasks. (Because Matrices scores were not correlated with other measures in this restricted sample of children who had timing measures, these scores were excluded from the regressions.) A set of six stepwise regression analyses was conducted, as summarized in Table 3. These regressions can be used to construct a diagram of all of the shared and unique sources of variance among the independent variables in the prediction of Word Reading, in the manner described by Chuah and Maybery (1999). Such an analysis is shown in Figure 2. Although both span and response durations picked up comparable amounts of cross-domain variance (.11 vs. .12, respectively), the response durations contributed more variance that was specific to the Word Reading task (.14) than did span (.07). Notice also that span and response timing shared no variance other than general variance that was also shared with the Number task (.05 of the variance, shown in the center section of Figure 2).

Table 2 shows that it was only the preparatory intervals, not the pauses or word durations, that were significantly correlated with both Word Reading and Number tasks. A set of regressions using preparatory intervals instead of the total response durations (shown in parentheses in Table 3) resulted in a figure very similar to what is shown in Figure 2, with all variance estimates identical or close to the ones shown in the figure (changes being that the unique Number variance and that shared by span and duration both increased by .01, that shared by Number and duration decreased by .01, and unique variance of duration decreased by .04, to .10).

Table 2
Correlations Between Key Measures in Experiment 1

Measure	MSp	SpA	PrI	WdD	PsD	RsD	WdR	Mat	Num
Span									
Maximal span (MSp)	—	.86*					.46*	.30*	.37*
Span-A (SpA)	.78*	—					.55*	.35*	.45*
Response timing									
Preparatory interval (PrI)	-.17	-.11	—						
Word durations (WdD)	-.01	-.09	-.15	—					
Pause (PsD)	-.17	-.11	.66*	.24	—				
Response duration (RsD)	-.18	-.12	.93*	.27*	.88*	—			
BAS criterion									
Word Reading (WdR)	.29*	.42*	-.46*	-.32*	-.45*	-.51*	—	.37*	.56*
Matrices (Mat)	.10	.18	.02	-.20	.18	.06	.24	—	.26*
Number (Num)	.21	.34*	-.29*	-.10	-.25	-.30*	.50*	.07	—

Note. Correlations below the diagonal reflect the 53 participants who had timing measures for two-item lists. Correlations above the diagonal reflect all 62 participants. BAS = British Abilities Scales.
* $p < .05$.

Interword pauses performed in a similar manner but with slightly lower correlations.

Contributing to the impression that response durations in reading span pick up variance related to reading skills, stepwise regressions with BAS Number skills as the dependent variable showed that the response durations contributed no variance that was not already contributed by span and BAS Word Reading ability, and that entering response durations first did not completely eliminate the contributions of these other variables.

Finally, to investigate the low correlations with the BAS Matrices task, another multiple regression on all 62 children (including those without two-item responses to be timed) was conducted to predict Word Reading using Matrices as well as Number and WM span. Matrices accounted for a significant .14 of Word Reading if entered first. It added a significant .06 if entered second after Number, and a nonsignificant .04 if added second after WM Span. There was no significant contribution of Matrices among the 53 children with List Length 2 timing data. Thus, the Matrices

task discriminates only between children at the low end of the WM scale, which includes those who have no two-item lists to be timed.

Discussion

This experiment has established several basic points. First, the silent intervals within the responses in reading span were 4 to 10 times longer than those found within responses in STM span procedures (e.g., Cowan, 1992; Cowan et al., 1994, 1998; Hulme et al., 1999). This suggests that the processes involved in recall are much more elaborate for the present reading-span task. One likely interpretation of that finding is that children did not really hold the list-final words in a ready form but, rather, allowed the words to become inactive in memory during sentence processing and then had to recall the sentences covertly as reminders regarding the list-final words. This would be one version of the attention-switching hypothesis proposed by Hitch et al. (2001).

Second, despite these differences, there are important similarities between the patterns of recall phenomena reported previously in STM span tasks and found with reading span here. Specifically, pauses tended to increase within an individual as list length increased, and children with lower spans showed longer interword

Table 3
Regressions (N = 53) for BAS Word Reading in Experiment 1

Variable	ΔR^2	Variable	ΔR^2
Analysis 1		Analysis 4	
Span	.18* (.18*)	Number	.25* (.25*)
Number	.14* (.14*)	Duration	.14* (.11*)
Duration	.14* (.11*)	Span	.07* (.07*)
Analysis 2		Analysis 5	
Span	.18* (.18*)	Duration	.26* (.21*)
Duration	.21* (.17*)	Span	.13* (.14*)
Number	.07* (.08*)	Number	.07* (.08*)
Analysis 3		Analysis 6	
Number	.25* (.25*)	Duration	.26* (.21*)
Span	.07* (.07*)	Number	.13* (.15*)
Duration	.14* (.11*)	Span	.07* (.07*)

Note. Span is Span-A; Number is the BAS Number Skills score; Duration is the total response duration for correctly recalled two-item lists (in parentheses: Duration = preparatory intervals). BAS = British Abilities Scales.
* $p < .05$.

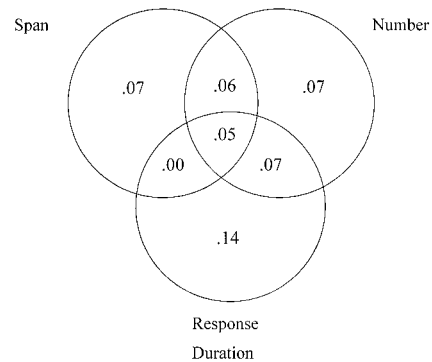


Figure 2. Experiment 1: Diagram of shared and unique sources of variance in the prediction of the Word Reading test from the British Abilities Scales.

pauses for lists of a particular length. These findings are consistent with the notion that the pauses were used at least partly for some sort of memory search process, as Cowan et al. (1994, 1998) have suggested for STM tasks, and that this process was accomplished more efficiently in children with a higher span.

Third, analysis has shown that individual differences in response timing are related to children's skills on other tasks. As shown in Figure 2, response timing in the reading-span task predicted linguistic skill in a manner that did not reduce the correlation between span and scholastic skills. Instead, it predicted considerable additional domain-specific linguistic ability (as well as some further variance, in common with the Number task, in the prediction of Word Reading). It did not similarly add to the prediction of the Number task performance beyond what span predicted, however. All of this suggests that response durations in the reading-span task may reflect the processing and retention of language processing per se, in a manner not fully captured by WM-span scores.

Although memory-search processes seemed to affect the interword pauses, the correlations with scholastic skills were not limited to those pauses and seemed more general across segments of the response. Correlations were slightly stronger for preparatory intervals than they were for interword pauses. Given that preparatory interval durations were in a range similar to interword pauses but were not sensitive to list length, it appears that memory-search processes are not the processes driving the correlations with scholastic skills. Instead, a more general speed of processing (e.g., Kail & Salthouse, 1994) or increases in retrieval efficiency may drive the correlations. The generality of these findings is explored across age groups, with three other span tasks, in Experiment 2.

Experiment 2

Experiment 2 provided the opportunity to clarify further the interpretation of response timing in WM tasks. It focuses on the cause of the long recall times described above, and considers further the correlations between response timing and scholastic skills. These questions were addressed through an examination of a set of two WM tasks and an STM task, carried out by participants from a much wider age range than in Experiment 1. Also, rapid-speaking tests were included as a way to test the specificity of findings based on recall durations.

The approach toward individual differences was slightly different than in Experiment 1. Rather than providing a relatively large sample within a single age group, smaller samples of three age groups were used (children of two age groups and adults). For examinations of the pattern of span timing responses, normal development was used as the basis of individual differences, providing a quasi-experimental manipulation of these individual differences. For examinations of scholastic measures, given that children and adults had different measures, within-group correlations were examined. With this purpose in mind, the adult group was about as large as the two child groups combined.

What Is the Cause of Long Recall Durations?

One possible explanation of the long recall times in Experiment 1 is that they result from the inclusion of a processing component within the memory task. If that is the case, then long recall times should occur also in any other such WM task. In

Experiment 2 we examine listening and counting WM spans, and digit span as a control STM task. A second possibility is that the long recall times are specific to tasks involving reading processes. In that case they should not be obtained in any of the tasks administered in Experiment 2.

A third possibility is that it is the content of the processing portion of the reading-span task (e.g., semantic or syntactic content, or both) and its role in the retrieval of sentence structure that causes the long recall times. If so, then these long recall times should occur for listening span as in reading span, the main difference between them being the written versus spoken modality of the presentation. Meaningful sentences are used in listening comprehension as in reading, and both allow reconstruction of the sentence on the basis of its meaning or structural coherence. In contrast, by this view, much shorter recall times should be found for counting span, in which participants count the number of dots on each screen and remember all of the reported sums. That task provides very little unique context on every screen that would identify its sum. Moreover, in contrast to the use of each sentence only once in the listening-span task, in counting span particular sums are repeated on different trials. Given the considerable interference in memory between one screen and the next, retrieval of the processing context for each sum should be impractical and participants should have to maintain the sums in memory as a list of numbers, as in standard span.

What Is the Basis of Correlations Between Recall Durations and Scholastic Skills?

In Experiment 1, it was suggested that recall durations correlated with scholastic skills perhaps because they carried information about reading ability. Thus, recall durations contributed considerable unique variance to Word Reading, rather than variance in common with reading span, and did not contribute uniquely to the Number Skills task. However, it could have been certain linguistic processing abilities generally, rather than just reading abilities, underlying these correlations. If so, similar correlations would also be expected with a listening-span task (and perhaps with STM tasks in the language domain). An alternative possibility is that recall durations reflect something general about WM performance and that any WM task would yield similar correlations with scholastic measures. In short, there are multiple interpretations of the correlations observed in Experiment 1 and additional experimentation with multiple WM measures should help to discriminate among them.

Changes in Method

To address these issues, Experiment 2 involved time coding of several spans in an experiment that differed from the first in several ways:

1. Performance was examined in three age groups (9–10 years, 11–12 years, and adults). If interword pauses for lists of a fixed length are consistently smaller for individuals with higher spans, they should be smaller for more mature individuals (cf. Cowan et al., 1998, using digit spans). Development is, however, a particularly

powerful, quasi-experimental manipulation of ability level.

2. Different span tasks were examined. These included a listening-span task, a counting-span task, and a digit-span task to replicate prior studies of response timing in STM span (e.g., Cowan et al., 1998) and potentially to provide a contrast with other spans. Spoken responses were automatically recorded (on tape for counting and listening span, digitally for the other tasks) and were not hand-recorded until after the trial ended.
3. Different criterion scholastic tasks were examined. For many of the children, the Cognitive Abilities Test (CAT) was available from the schools. This test serves roughly the same purpose as the BAS, with roughly similar subtests (Verbal, Mathematical, and Nonverbal). For college students, the American College Test (ACT) was available, including English, Mathematics, Reading, and Science subtests. This test is one of two commonly used college entrance examinations that is intended to predict success in college. The high-school grades percentile, based on classroom grades awarded by high-school teachers for all aspects of course performance, was also available. We asked what recall timing would add to the prediction of these scholastic measures.
4. Rapid-speaking tests were carried out. These are considered to be estimates of the ability to carry out rapid covert rehearsal (Baddeley, 1986; Landauer, 1962). Cowan et al. (1998) found that retrieval times (estimated by digit-span recall pauses) and rehearsal times (estimated by rapid-speaking tests) did not correlate with one another and picked up different portions of the variance in memory span. Therefore, we asked whether these types of measures in WM-span tasks would correlate differently with scholastic abilities or whether a global speed of processing (Kail & Salthouse, 1994) could account for all of the timing results.

Method

Design

Children in two age groups and adults received various tasks for other purposes, and the present analyses focus on the ones deemed relevant: a counting-span task and a listening-span task from one test session and a digit-span task and several rapid-speaking tasks from a second session held on a different day. Spans and response timing were measured as in Experiment 1.

Participants

The participants were children and adults from the Columbia, Missouri, area. The final sample of participants with all measures included 25 third graders (15 girls, 10 boys) ranging in age from 97 to 121 months, with a mean age of 105.96 months, ($SD = 5.51$; 8- and 9-year-olds); 25 fifth graders (17 girls, 8 boys) ranging in age from 119 to 136 months, with a mean age of 128.16 months ($SD = 5.21$; 10- and 11-year-olds); and 51 adults (33 women, 18 men) ranging in age from 217 to 333 months, with a mean age of 232.80 months ($SD = 17.54$). An additional 14 children (4

female and 5 male third graders and 2 female and 3 male fifth graders) who carried out all tasks but could not be used for span response timing, because they did not correctly and audibly repeat at least 1 two-item list in each span task, were included in some correlations not involving timing, to derive estimates of population values.

All of the participants reported normal or corrected-to-normal vision and normal hearing. Children were recruited from the Columbia Public Schools system and received either \$5 and a book for their participation or \$10. Adults were recruited from the department's participant pool and received course credit. Testing required approximately 1.5 hr in one session and 1 hr in a second session. There were multiple opportunities for breaks throughout both experimental sessions. The children were also rewarded with stickers at several points.

Apparatus, Stimuli, and Procedures

The experimental sessions took place inside of sound-attenuated booths. A total of seven tasks were administered over the two experimental sessions on separate days. However, the timing analyses to be reported are restricted to four types of tasks. Intervening ones, which are shown in brackets, will not be described in detail. Session 1 included [running span], counting span, listening span, and [visual array memory], in that order. Session 2 included digit span and a set of rapid-speaking tasks, in that order. The counting- and listening-span tasks were conducted using MEL Version 2.0 by Psychology Software Tools (Schneider, 1988), whereas the other tasks were conducted using programs developed for Power Macintosh computers with SuperCard software. Listening span was presented in a female voice, whereas digits were presented in a male voice. A description of the relevant tasks follows.

Counting span. Counting span was adapted from Conway, Bottoms, Nysse, Haegerich, and Davis (2000; modeled on Case et al., 1982). Arrays of targets (dark blue circles) and distractors (red squares and circles) were presented on the screen, with the targets to be counted. There were three to nine targets on a screen and no repetition of the same sum more than once within a trial. There were one to five circular distractors and one to nine square distractors, which varied independently. After several such displays were counted, a cue to recall the separate sums for all screens was presented: the printed word "RECALL" accompanied by a 1000-Hz tone measured at 73 dB(A).

The displays required that participants orally pronounce the number of dark blue circles (targets) that appeared on each screen, while ignoring the red items. Participants moved through the program by pressing the spacebar, determining the duration of presentation for each screen; the experimenter recorded the accuracy of the sum as well as the recall. Upon completion of counting target dots within all screens in the trial, a cue to recall the sums for each screen separately, in the serial order in which screens had been presented, was given. Three practice trials of List Length 2 (i.e., two screens in a trial) were completed before the test trials began. For the children, 3 trials were presented for lists of each length: two, three, four, and five screens, one trial per list length in that order, repeated for three runs for a total of 12 trials. The adults followed the same procedure except that they proceeded to List Length 6 in each run, for a total of 15 lists. Using six-item trials was necessary to allow more sensitivity to discriminate among adults, but it was found to be discouraging to many children, and so was omitted. When data from children and adults were compared or combined, spans were recalculated with List Length 6 omitted so that all participants' scores were based on the same trial types.

Listening span. The listening-span task followed the specific procedure of Kail and Hall (1999; modeled on Daneman & Carpenter, 1980). Sentences were presented through speakers at 66–68 dB(A), and the same recall cue as on the counting-span task was used. Participants were instructed to listen to each sentence and determine whether it was true or not. They were to respond "yes" or "no" and then to repeat the final word of the sentence because they would need to remember it for later. For instance, in

the practice example, the sentence was *A fox can drive a truck* and the correct response was “no, truck.” That sentence is typical in difficulty level (e.g., *a chicken lays eggs; you wear pants on your arms*), and no sentence was used more than once. Three practice lists of two sentences each were completed first to ensure that participants understood the task instructions. In the test, participants pressed the spacebar to move to the next sentence when they were ready. When they heard the recall cue and saw the word RECALL printed on the screen (as in counting span), they were to repeat the sentence-final words in the order in which they had been presented. Then the trials were presented to the children with 1 trial of each list length (i.e., 2–5 sentences in a trial in that order), repeated for three runs for a total of 12 trials, in a manner comparable with the counting span. Also as in the counting-span task, adults proceeded to List Length 6 in each run, for a total of 15 sentences. The experimenter recorded the responses by hand and later entered them into the computer.

Digit span. On each trial, lists of digits were presented through headphones at 68–70 dB(A) at a rate of one per second. Participants were asked to listen to the digits and wait for the simultaneous visual and auditory recall cues (a tone, as above, and a change in box color) before beginning their spoken response, recalling the digits in the presented order. The experimenter recorded their responses and then entered them into the computer program. Four lists at List Length 2 were used as practice, and then the test trials within each span run began at List Length 2 and continued to a maximum of List Length 9. The computer program presented four trials at each list length and continued until the point at which the participant did not get any lists correct at a given list length. Two span runs were completed, and the second span run did not include any practice trials.

Rapid speaking. Stimuli were presented through headphones at 68–70 dB(A). Two seconds after the end of the stimulus presentation (if any), a ready cue appeared for 2 s, consisting of a yellow box with the word *ready* in it. After the ready cue disappeared, the response cue (a 100-ms, 440-Hz triangular tone) occurred after a random delay of between 1 and 2 s. (A random delay was used because a predictable cue onset might result in some participants beginning to speak even before the cue.) Participants were instructed to wait for the tone and then to speak their response as quickly as possible. They were corrected if they made a mistake, and were told to maintain clarity of pronunciation if necessary.

In the list-repetition task, two practice trials preceded each stimulus set and were followed by four test trials with that set. On each trial, the set of three random numbers was aurally presented, followed by the tone. Participants were to repeat the three numbers as quickly as possible after the tone. This procedure was carried out sequentially for three different three-digit sets (259, 386, and 741). Additional rapid-speaking tasks carried out subsequently included counting from 1–10 as quickly as possible after a tone signal and reciting the alphabet from A–Z as quickly as possible after a tone signal. In each case, two practice trials were followed by four test trials. These responses could be analyzed for both preparatory intervals and pronunciation intervals but interword pauses were generally too short to measure.

Results and Discussion

The results are organized in the same manner as in Experiment 1 with subsections on the means and variability, the pattern of span response timing in different tasks, individual differences in this timing, and the relation of timing and other measures to scholastic abilities. The main difference is that the investigation focuses on differences between tasks and age groups rather than on differences in ability level within a single age group on a single task.

Means and Variability

The same criteria for outliers in the timing data were used as in Experiment 1 (trials with response durations totaling more than

20 s through the first two words or more than 25 s through the first three words). This stable criterion was maintained to allow comparisons across data types and to ensure that age differences in response times could not be attributed to differential treatment of the data. The criterion eliminated only 1 two-item trial in a fifth grader and 2 three-item trials, one from a third grader and one from a fifth grader, in the listening-span task. No other tasks were affected and adults were unaffected.

The means and standard deviations for various key measures are presented in Table 4 for the final sample of participants in each age group. Regarding spans, Table 4 suggests that there were developmental improvements in all sorts of spans and decreases in duration for all sorts of timing measures. An Age Group \times Memory Task ANOVA on Span-A produced not only main effects of the age group, $F(2, 98) = 47.62$, $MSE = 1.19$, $p < .001$, $\omega^2 = .24$, and memory task, $F(2, 196) = 624.11$, $MSE = 0.36$, $p < .001$, $\omega^2 = .80$, but also an interaction of these factors, $F(4, 196) = 12.45$, $MSE = 0.36$, $p < .001$, $\omega^2 = .13$. Post hoc Newman–Keuls tests showed that each of the three age group’s span differed from the other two groups and that each of the three span tasks differed in performance levels, at $p < .05$ or lower. As the means in Table 4 indicate, the difference between the young children and the adults was 1.86 units for digit span, 1.53 units for listening span, and 0.92 units for counting span. The difference between tasks is not a range effect or artifact from combining different measures, given that the same interaction was significant when z scores for each task were used in the analysis.

Pattern of Timing of Recall in Different Tasks

Segment durations. A key finding in Experiment 1 was that preparatory and pause intervals in reading span were much longer than had been obtained in the past in STM span. To compare segment durations across tasks in Experiment 2, first an ANOVA of the total response durations in all span tasks for two-item lists was conducted (i.e., 3 age groups \times 3 span tasks). This analysis produced main effects of age group, $F(2, 98) = 21.60$, $MSE = 2.39$, $p < .001$, $\omega^2 = .12$, and span task, $F(2, 196) = 80.65$, $MSE = 1.90$, $p < .001$, $\omega^2 = .34$, as well as an interaction of these two factors, $F(4, 196) = 11.58$, $MSE = 1.90$, $p < .001$, $\omega^2 = .12$. The basis of these effects is shown in Table 4 and also in Figure 3. Newman–Keuls tests showed that all three age groups differed from one another and that all three span tasks differed from one another. As Figure 3 shows, total response times were especially long for the listening-span task, and this was especially so in the youngest age group. The same Age Group \times Span Task interaction was significant when z scores for each task were used in the analysis.

It is interesting that the pattern of timing shows that not all WM-span tasks involve comparable processing. In the listening-span task, as in the reading-span task of Experiment 1, it would be possible to retrieve the sentence-final words by recalling the sentence context in which those words appeared. In contrast, in the counting-span task, counting each set of dots does not provide a distinctive context that could help in recalling the count totals. In the counting-span task, as in digit-span tasks, it may be more consistently necessary to retain the list of items in short-term storage rather than reconstructing it from long-term memory, which, we suggest, is a possible mechanism of recall in the

Table 4
Means and Standard Deviations for Key Measures in Experiment 2

Measure	Grade 3 (<i>N</i> = 25)		Grade 5 (<i>N</i> = 25)		Adult (<i>N</i> = 51)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Span-A						
Standard digit span	5.12	0.87	5.47	0.78	6.97	1.00
Listening span	2.22	0.59	2.74	0.75	3.75	0.88
Counting span	2.82	0.82	3.58	0.63	3.74	0.61
List Length 2 response timing						
Standard span						
Preparatory	0.58	0.15	0.60	0.16	0.56	0.14
Interval						
Word duration	0.43	0.06	0.44	0.06	0.38	0.06
Pause duration	0.14	0.13	0.16	0.15	0.09	0.11
Total duration	1.57	0.28	1.64	0.30	1.40	0.27
Listening span						
Preparatory	3.46	2.90	2.01	1.43	1.33	0.65
Interval						
Word duration	0.60	0.12	0.58	0.15	0.48	0.07
Pause duration	1.08	1.47	0.61	0.95	0.18	0.23
Total duration	5.74	4.12	3.73	2.41	2.48	0.78
Counting span						
Preparatory	1.37	0.66	1.01	0.28	0.83	0.24
Interval						
Word duration	0.42	0.11	0.40	0.08	0.35	0.07
Pause duration	0.22	0.29	0.13	0.21	0.06	0.10
Total duration	2.43	0.73	1.94	0.41	1.58	0.32
Separate rapid speech						
Repeat three digits						
Preparatory	0.44	0.13	0.35	0.17	0.25	0.10
Interval						
Speaking duration	0.70	0.13	0.68	0.11	0.67	0.12
Total duration	1.15	0.16	1.03	0.24	0.91	0.18
Recite alphabet						
Preparatory	0.51	0.16	0.42	0.18	0.35	0.12
Interval						
Speaking duration	5.01	1.26	4.16	0.67	3.72	0.66
Total duration	5.52	1.33	4.58	0.76	4.07	0.73
Count 1–10						
Preparatory	0.46	0.16	0.43	0.21	0.30	0.10
Interval						
Speaking duration	1.90	0.28	1.72	0.34	1.55	0.28
Total duration	2.36	0.38	2.15	0.45	1.86	0.34
Children's Grade 2 criterion (Cognitive Abilities Test)						
Verbal	112.12	12.72	111.96	14.02		
Quantitative	113.64	14.75	114.12	11.42		
Nonverbal	115.00	13.60	113.48	16.01		
Composite (%)	116.04	14.36	114.36	12.71		
Adults' criteria						
High-school grades percentile					76.80	19.10
American College Test						
English					25.12	4.43
Math					23.94	4.94
Reading					26.67	4.66
Science					24.06	4.19
Composite					25.12	3.79

Note. Only adults received a List Length 6 in listening and counting span. Spans (with *SDs*) with List Length 6 included were 3.94 (1.05) and 3.92 (0.77), respectively. Children's criterion measures are age-adjusted standard scores.

listening-span task. This type of consideration can explain why response durations were much longer in the reading- and listening-span tasks than in the counting- and digit-span tasks.

Separate one-way ANOVAs were conducted to confirm the statistical significance of age group effects on response timing in

each type of span. On the basis of past research we expected an age group effect for response durations in the digit-span task (Cowan et al., 1998) and that was obtained, $F(2, 98) = 7.13$, $MSE = 0.08$, $p < .01$, $\omega^2 = .11$. The same was true for counting span, $F(2, 98) = 27.16$, $MSE = 0.22$, $p < .001$, $\omega^2 = .34$, and listening span, $F(2,$

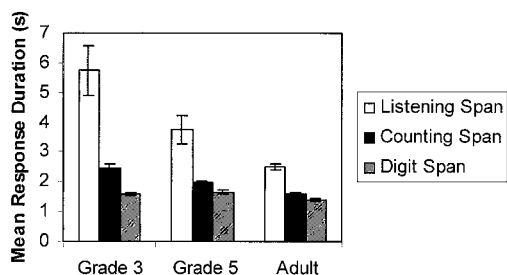


Figure 3. Experiment 2: Mean durations (in seconds) of recall within correct responses in each age group to two-item lists in listening-span (white bars), counting-span (black bars), and digit-span (gray bars) tasks. Error bars are standard errors.

98) = 15.12, $MSE = 5.89$, $p < .001$, $\omega^2 = .22$. In all tasks, younger children produced longer responses for correctly recalled two-item lists.

Age Group \times Span Task type analyses also were carried out separately for the preparatory intervals, word durations, and interword pauses. These were all similar to the analyses of total response durations and therefore will not be presented separately. In all three analyses, the main effects of age group and of span task were significant. In the analyses of preparatory intervals and interword pauses, the interaction was also significant but it was only marginal for word durations. Table 4 shows that the pattern of means was generally similar in each case. This general age trend conforms to previous reports (Cowan et al., 1998).

List-length effects. Different types of span task proved to have basically similar patterns of response timing, although the strength of the pattern varied across tasks. The digit-span task, in which all participants in the main sample had usable timing for two-, three-, four-, and five-item lists, replicated the pattern expected from previous work (Cowan et al., 1998), which included effects of list length on interword pause durations in children. These effects are shown in Figure 4. All children in the final sample ($N = 101$) had timing for two-, three-, four-, and five-digit lists. An Age Group \times List Length ANOVA of the first interword pause in a list produced significant main effects of age group, $F(1, 98) = 6.08$, $MSE = 0.07$, $p < .01$, $\omega^2 = .05$, and list length, $F(3, 294) = 6.82$, $MSE = 0.003$, $p < .001$, $\omega^2 = .08$, as well as an interaction of these factors, $F(6, 294) = 3.87$, $MSE = 0.003$, $p < .001$, $\omega^2 = .08$.

As Figure 4 shows, for the children, there was an increase in interword pauses across list lengths but that was not the case for adults. The absence of list-length effects in adults, despite a much larger sample size than in the children, brings up the possibility that such effects occur only when the list lengths are close to span length, as they were, for example, in a previous study of spoken response timing of STM-span using adult participants (Hulme et al., 1999).

For the counting and listening spans, it was only possible to examine two- and three-item lists because of the paucity of correctly repeated lists at longer lengths. Counting span could be analyzed with 17, 25, and 50 participants in the three age groups, respectively, and listening span, with 16, 22, and 49 participants, respectively. These analyses did not produce consistent effects of list length across age groups for any response segment. However, there were effects on the interword pauses for third-grade children

(approximately the same age participants as in Experiment 1, who showed list-length effects in reading span). In the counting-span task, the third graders' mean pauses went from 0.15 s for two-item lists to 0.46 s for three-item lists; the other two groups declined slightly across list lengths and, consequently, the Age Group \times List Length interaction was significant, $F(2, 89) = 5.14$, $MSE = 0.07$, $p < .01$, $\omega^2 = .04$. In the listening-span task, none of the effects were significant but there was again a trend toward an effect of list length for interword pauses in third graders, with 0.66 s ($SD = 0.15$) for two-item lists and 0.85 s ($SD = 0.15$) for three-item lists, but not in older participants.

In sum, the analyses have demonstrated both similarities and differences between the span measures. They are basically similar in the pattern of effects of list length. The main differences, which pertain to mean response interval lengths, do not divide digit span neatly from the WM spans. Instead, reading and listening spans are similar in that they both produced very long response intervals, unlike either digit span or counting span. Thus, the tasks with linguistic and semantic information in the processing component are the ones showing the long intervals.

Individual Differences in Span and Response Timing

In the digit-span task, the results basically confirm findings from previous studies of STM span (Cowan, 1999; Cowan et al., 1998). Children with higher spans repeated lists of a particular length with shorter silent intervals. Also, in agreement with previous research (Cowan, 1999), the effect on interword pauses was age-dependent. In both cases it was fifth-grade children who showed the effect. In the 25 fifth-grade children of Experiment 2, the correlations between digit span and interword pauses in that same task were, for List Length 2, 3, 4, and 5, respectively, $r = -.40$, $p < .05$; $r = -.31$, ns ; $r = -.43$, $p < .05$; and $r = -.46$, $p < .05$. None of these correlations with interword pauses approached significance in the third graders (replicating Cowan, 1999) or in adults (replicating Hulme et al., 1999).

In Experiment 2 an additional within-age effect was found, which had not been examined in the same way before. Specifically, in third-grade children, children with higher spans had shorter preparatory intervals within correct list repetitions. For List Length 2, 3, 4, and 5, respectively, $r = .32$, ns ; $r = -.46$, $p < .05$; $r = -.50$, $p < .01$; and $r = -.57$, $p < .01$. The preparatory intervals were not related to span in either of the older groups. Thus, there is simply a shift across ages in the location within the response at which differences in the speed in processing emerge.

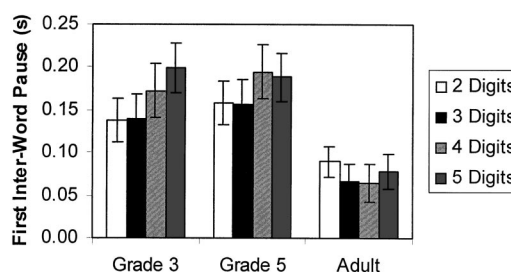


Figure 4. Experiment 2: First interword pauses (in seconds) in responses within the digit-span task for lists of two, three, four, and five digits (graph parameter) in three age groups (X axis). Error bars are standard errors.

These age-specific individual-difference effects on digit-span response timing are illustrated in Figure 5.

The investigation within listening and counting spans was limited to List Length 2 because only that list length produced timing data in all 101 participants within the main sample. For the listening spans, only the third-grade children showed effects of span on timing and these were specific to the preparatory intervals (as in the digit spans). The results were not significant in counting spans. Thus, in sum, the durations of response segments seem to show task- and age-specific patterns.

Relation Between WM Span, Timing Measures, and Scholastic Skills

Given that children and adults had different measures of scholastic performance, of necessity they were examined separately. The partial correlations between all measures in children, with age group partialled out, are shown in Table 5. (The restricted sample is shown below the diagonal, and results including children who did not have all of the timing measures are shown above the diagonal.) These correlations include only the overall response durations; below we will clarify which response segments drove the correlations.

In the final sample of 51 adults, the span response durations for two-item lists did not correlate with any other variables. We noticed, though, that 48 of these adults also had response timing measures for three-item lists in all span tasks. These recall durations for three-item lists proved to have higher correlations with

other measures. Table 6 shows the correlations between measures in these 48 adults, using the response timing to three-item lists.

Correlations between spans and test scores. Tables 5 and 6 both show that spans correlated with scholastic tests rather well. A curious aspect of the findings was that standard digit span generally correlated with scholastic tests almost as well as listening span and better than counting span (except for the science portion of the test for the college students), in contrast to the general conclusion (e.g., Daneman & Merikle, 1996) that WM spans correlate more highly because they require processing and storage at the same time. Daneman and Merikle showed that the usual difference in sizes of correlations cannot be explained on the basis of differences in task reliability (see also Engle et al., 1999). In the present case, as well, reliability was not the issue. In the final sample of 51 adults, we estimated the reliability of the composite ACT score on the basis of its four subtests and found Cronbach's standardized $\alpha = .84$; for the counting and listening spans, on the basis of three test runs, respectively, $\alpha_s = .69$ and $.88$; and for the digit span, on the basis of the correlation between two runs, the reliability was $.86$. Whereas the raw correlations between the ACT composite and the three spans (counting, listening, and digit spans) were $.33$, $.52$, and $.42$, respectively, the correlations corrected for attenuation due to imperfect reliability were $.43$, $.60$, and $.49$, respectively. Thus, in the corrected correlations, digit span still produced a higher correlation than counting span, although not as high as listening span. In children, the reliability of the CAT composite score, on the basis of its three subtests, was $\alpha = .85$, and the reliabilities of the counting, listening, and digit spans were $.76$, $.80$, and $.83$, respectively (calculated as above). Whereas the partial correlations between the CAT composite score and the counting, listening, and digit spans (with age partialled out) calculated from uncorrected raw correlations were $r_p = .05$, $.39$, and $.42$, respectively, the partial correlations on the basis of first-order correlations that had been corrected for attenuation were $r_p = .03$, $.49$, and $.50$, respectively. In this case, digit span produced correlations at least as high as did WM spans, even when corrected for attenuation.

Correlations and regressions with response durations in children. Table 5 shows that, in children, the response durations for two-item lists in span tasks did not correlate particularly well with scholastic measures. However, even the correlations with age partialled out cannot capture the fact that there were different age-specific patterns. In the 25 third-grade children, the digit-span response durations correlated with the verbal CAT score, $r = -.40$, $p < .05$, and with the quantitative CAT score, $r = -.45$, $p < .05$. In contrast, in the 25 fifth graders, it was the listening-span rather than the digit-span durations that yielded correlations (although these were only marginal for total response durations). In this group, the listening-span preparatory intervals correlated with the verbal CAT score, $r = -.47$, $p < .05$, and with the quantitative CAT score, $r = -.47$, $p < .05$, and the listening-span word durations correlated with the verbal CAT score, $r = -.56$, $p < .01$; the quantitative CAT score, $r = -.48$, $p < .05$; and the nonverbal CAT scores, $r = -.40$, $p < .05$.

Multiple regressions did not turn up a unique contribution of the response timing for third graders, but they did for fifth graders when using the specific measures that showed correlations with the CAT. When all three spans were entered first into a regression, they accounted for the composite CAT score fairly well, $R^2 = .31$, but the relevant timing measures from the listening-span task,

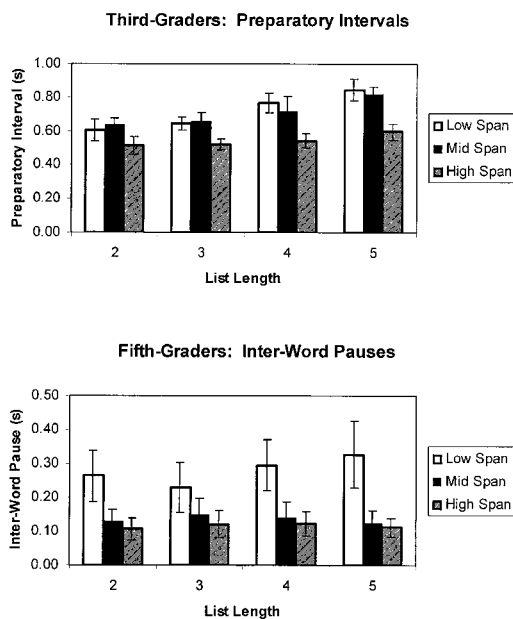


Figure 5. Experiment 2: Mean duration of specific segments within correct responses to lists of different lengths in the digit-span task in children with the lowest span (white bars), medium span (black bars), or highest span (gray bars). Top panel: Preparatory intervals (in seconds) in third-grade children. Bottom panel: Interword pauses (in seconds) in fifth-grade children. It was these segments that showed span effects in the digit-span task. Error bars are standard errors.

Table 5
Partial Correlations Between Measures in Children in Experiment 2, Controlling for Age Group

Measure	CS	LS	DS	CD	LD	DD	3D	AL	10	VA	QA	NA
Working memory spans												
Counting span (CS)	—	.48*	.28*				-.01	-.03	.07	.16	.19	.11
Listening span (LS)	.42*	—	.43*				-.21	-.09	-.20	.49*	.39*	.29*
Digit span (DS)	.27	.41*	—				-.31*	-.22	-.27*	.45*	.47*	.38*
Response duration (2-Item totals)												
Counting span (CD)	.03	-.34*	-.29*	—								
Listening span (LD)	-.36*	-.31*	-.25*	.34*	—							
Digit span (DD)	.13	-.19	-.20	.21	.05	—						
Rapid speaking												
Three digits (3D)	-.05	-.26	-.30*	.24	.09	.41*	—	.58*	.73*	-.32*	-.39*	-.31*
Alphabet (AL)	-.13	-.13	-.22	.09	.23	.28*	.57*	—	.51*	-.26*	-.47*	-.20
Count 1-10 (10)	-.11	-.11	-.18	.00	.11	.20	.77*	.65*	—	-.18	-.36*	-.23
Cognitive abilities subtests												
Verbal (VA)	.11	.48*	.44*	-.13	-.12	-.11	-.32*	-.17	-.16	—	.66*	.45*
Quantitative (QA)	.13	.37*	.38*	-.28*	-.07	-.33*	-.41*	-.47*	-.34*	.64*	—	.61*
Nonverbal (NA)	.02	.36*	.33*	-.18	-.01	-.33*	-.41*	-.31*	-.31*	.55*	.58*	—

Note. Below the diagonal, $N = 50$, 25 children per age group who have span task response timing measures for two-item sets in each span task. Above the diagonal, $N = 64$, the entire sample of children who have carried out all of the tasks.

* $p < .05$.

entered subsequently, added a substantial and significant amount, $\Delta R^2 = .28$, raising the total R^2 to .59. Although these particular timing variables could not be identified a priori, this analysis still highlights the promise of using timing to increase the predictive capability of WM-span tasks.

Correlations and regressions with response durations in adults. In the adults, the duration of responses to three-item lists did not account well for scholastic test results. It is interesting, though, as shown in Table 6, that the duration of responses to three-item lists in the counting- and listening-span tasks provided significant predictors of a percentile rating based on high-school grades. Except for the subtests of the ACT, there was no other significant predic-

tor of high-school grades; memory spans did not significantly predict these grades. Perhaps the durations of responses in difficult WM-span trials reflect a linguistic skill that is akin to what students need to study and do well on academic tests, a skill that apparently is not captured by WM spans. A microscopic analysis indicated that the preparatory intervals drove the correlations between WM-span response durations and high-school grades. The preparatory intervals ($N = 48$) correlated significantly with grades both for counting-span responses, $r = -.41$, and for listening-span responses, $r = -.39$.

Using this information about the usefulness of preparatory intervals in WM tasks, two sets of stepwise regressions on high-

Table 6
Correlations Between Measures in Adults in Experiment 2

Measure	CS	LS	DS	CD	LD	DD	3D	AL	10	GP	EN	MA	RD
Working memory span													
Counting span (CS)	—												
Listening span (LS)	.44*	—											
Digit span (DS)	.49*	.59*	—										
Response timing (3-item sets)													
Counting span (CD)	-.07	-.15	-.10	—									
Listening span (LD)	-.24	-.44*	-.24	.63*	—								
Digit span (DD)	-.08	-.07	-.14	.33*	.12	—							
Rapid speaking													
Three digits (3D)	-.38*	-.27	-.27	-.21	-.10	.01	—						
Alphabet (AL)	-.31*	-.24	-.33*	-.18	-.16	.03	.75*	—					
Count 1-10 (10)	-.30*	-.22	-.25	-.25	-.16	.02	.89*	.81*	—				
High-school grades % (GP)													
	.24	.20	.21	-.34*	-.38*	-.02	.05	.06	.00	—			
ACT subtests													
English (EN)	.29*	.36*	.38*	-.33*	-.24	-.17	-.01	-.14	-.04	.46*	—		
Math (MA)	.35*	.55*	.45*	-.21	-.24	-.13	-.12	-.15	-.07	.44*	.67*	—	
Reading (RD)	.21	.44*	.37*	-.18	-.19	.06	.06	-.06	.00	.46*	.49*	.38*	—
Science	.29*	.47*	.19	-.15	-.24	-.09	.00	-.13	-.05	.51*	.57*	.68*	.60*

Note. $N = 48$ participants who have span task response timing measures for 3-item sets. ACT = American College Test.

* $p < .05$.

school grades were conducted using 48 adults who had timing for three-item trials in all three span tasks, one set based on total response durations and a second set based on preparatory intervals (in parentheses in Table 7). The ACT scores were entered into the analysis also, to determine how much of the variance was shared between different types of indicators of academic success. Table 7 shows that both sets of analyses were similar but that preparatory intervals accounted for somewhat more variance. A diagram depicting the different shared variances, based on the set of regressions using preparatory intervals, is shown in Figure 6. It is clear from Figure 6 that preparatory intervals, unlike spans, shared a great deal of variance with high-school grades. The preparatory intervals accounted for .19 of the variance in high-school grades and, of this, .08 was shared with ACT scores.

Relations between different kinds of timing measures. Last, it is worth noting that the rapid-speaking task measures did not show the same pattern of correlations as did the response durations in span tasks. For example, they did not correlate with high-school grades in adults as did the response durations in WM-span tasks (see Table 6). This difference supports the view that individual differences include multiple, noninterchangeable speeds of processing (Ackerman, Beier, & Perdue, 2002; Cowan et al., 1998).

On the other hand, there were common components of the timing measures. Given the suggestion that there are multiple processing components indexed by speeded-speech measures (Jarrod et al., 2000), we examined the correlations between different segments of the recall responses (preparatory intervals, word durations, and interword pauses) and different segments of the speeded-speech responses (preparatory intervals and speaking durations). This was done using all 101 participants in partial correlations controlling for age. The most regular finding was that the preparatory intervals in all three of the speeded measures were correlated with the preparatory intervals for counting span (for three digits, $r_p = .36$; for counting from 1 to 10, $r_p = .31$; and for reciting the alphabet, $r_p = .28$, all $ps < .05$) and for digit span ($r_p = .29$, for each of the three). For listening span, the correlation was significant only for reciting the alphabet, ($r_p = .24$). Also, the time

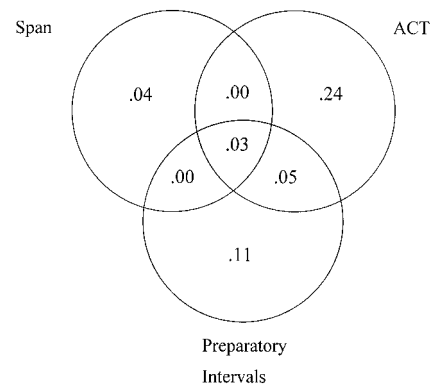


Figure 6. Experiment 2: Diagram of shared and unique sources of variance in the prediction of high-school grades percentile in college students. ACT = American College Test.

taken to recite three digits (speaking duration) was correlated with the word pronunciation time for digits in the digit-span task ($r_p = .27$, $p < .05$). Thus, despite task differences, there is some commonality among the response durations (preparatory intervals in span with preparatory intervals in speeded-speech tasks, and word durations in span with speeded-speaking durations).

General Discussion

The timing of recall has proven useful in the understanding of standard STM tasks (e.g., Cowan, 1992, 1999; Cowan et al., 1998; Hulme et al., 1999; Tehan & Lalor, 2000). The present studies are the first to examine the timing of recall in WM-span tasks in which a processing task component is combined with a memory task component. It is important to understand WM-span tasks inasmuch as they provide strong correlations with complex cognitive tasks and intelligence tests (e.g., Daneman & Merikle, 1996), which are related to individuals' potential in education, the workplace, and other intellectual and creative pursuits. A common approach for investigating WM tasks has been to show that high- versus low-span individuals respond differently to tasks with manipulations involving the use of attention (e.g., Conway, Cowan, & Bunting, 2001; Conway & Engle, 1994; Engle et al., 1999; Klein & Boals, 2001; Tuholski, Engle, & Baylis, 2001). Another strategy, however, is to use empirical and analytic techniques to investigate directly the processes involved in WM-span tasks. That type of strategy is exemplified by Case et al. (1982), Engle et al. (1992), Hitch et al. (2001), and Towse et al. (1998, 2000) and is developed further in the present article through the introduction of recall response timing analysis.

Two important conclusions can be drawn clearly from the present research. First, span response timing does not clearly differentiate the processes taking place in STM versus WM tasks. Instead, it differentiates the processes taking place in span tasks that do versus do not provide a context for retrieval (i.e., semantic or lexical context). The tasks that do provide such a context produce much longer response times. Second, span response timing can account for a considerable amount of variance in scholastic measures that is, for the most part, independent of the variance accounted for by the span measure itself. To the extent that the purpose of using WM-span tasks is to predict scholastic perfor-

Table 7
Regression (N = 48) for High-School Grades Percentile in Adults in Experiment 2

Variable	ΔR^2	Variable	ΔR^2
Analysis 1		Analysis 4	
Spans	.07 (.07)	ACT	.33* (.33*)
ACT	.29* (.29*)	Duration	.06 (.10*)
Duration	.09* (.11*)	Spans	.06 (.04)
Analysis 2		Analysis 5	
Spans	.07 (.07)	Duration	.16* (.20*)
Duration	.12 (.16*)	Spans	.03 (.03)
ACT	.26* (.24*)	ACT	.26* (.24*)
Analysis 3		Analysis 6	
ACT	.33* (.33*)	Duration	.16* (.20*)
Spans	.03 (.03)	ACT	.23* (.23*)
Duration	.09* (.11*)	Spans	.06 (.04)

Note. Spans are Span-A for counting- and listening-span tasks, Duration refers to total response durations for correctly recalled three-item lists in these two tasks (in parentheses: Duration = preparatory intervals), and ACT refers to all four subtests of the American College Test.

* $p < .05$.

mance, that purpose appears to be much better served if timing measures are used along with span measures.

Beyond these conclusions there also were some complex aspects of the results. Given that the WM-span tasks allow for alternative processing strategies, it is not surprising that the timing results differ across tasks and age groups. Nevertheless, the overall pattern of results can be conveyed simply, in terms of segment durations, list-length effects, individual differences in span and timing, and relations to scholastic measures. These topics will be addressed in turn.

Segment Durations

The finding here is simple and noteworthy. Reading- and listening-span tasks resulted in much longer response times than did counting- or digit-span tasks. This can be seen clearly in Figure 3 for the tasks in Experiment 2. Moreover, Table 1 shows that the mean response duration for reading span in Experiment 1 was just slightly higher than the mean for the youngest (most comparable) age group for listening span in Experiment 2.

This finding provides strong evidence against the notion that processing is equivalent in a variety of WM-span tasks. Instead, it suggests that the extra context provided by sentences in the reading- and listening-span tasks is used to help reconstruct the list of sentence-final words. This process was not necessarily beneficial for span performance, which was, in fact, slightly lower in listening span than in counting span, as Table 4 shows. Perhaps the attempt to reconstitute the list from sentential context in the linguistic spans has the unintended consequence of prolonging the recall episode long enough for some information to be lost from working memory (cf. Hitch et al., 2001; Towse et al., 1998, 2000). In any case, it seems likely that the ability to reconstitute information from a sentential context, or the decision as to whether to use that strategy, is likely to differ from other abilities underlying WM-span task performance. This research highlights the importance of understanding the processes involved in span tasks if they are to be used in the interpretation of individual differences in scholastic abilities.

The interpretation receives support also from a recent study by Copeland and Radvansky (2001). They found that phonological similarity among words to be recalled in a word-span test or an operation-span test was harmful to recall (i.e., the usual phonological similarity effect). In contrast, phonological similarity was helpful to recall in a reading-span test. Their interpretation, like ours, is that the rich semantic information available in the reading-span context provided the most important cue to recall, not the phonological memory that governed recall in the other span situations. Within that rich semantic context, rhymes between items were helpful cues.

List-Length Effects

List-length effects are informative because they provide a signature if interword pauses increase with list lengths, that listwide processing (e.g., memory search) occurs during these pauses. This effect from the developmental STM literature (e.g., Cowan, 1992; Cowan et al., 1998) was replicated closely for the digit-span measure in the present Experiment 2. Consistent with Hulme et al. (1999), it was somewhat weak in adults, appearing in the prepa-

ratory intervals at long list lengths but not in interword pauses. In WM tasks, the effect was obtained for reading span in 7- to 9-year-old children in Experiment 1 (e.g., Figure 1) and for counting span in the roughly comparable age group in Experiment 2 but not for these tasks in older children or adults. The developmental difference in this regard could occur because, for the list lengths we could examine across participants, more mature participants were not sufficiently challenged and the search times were small compared with any processes that do not depend upon list length (e.g., motor planning to make a response).

Individual Differences in Span and Timing

Younger children showed larger differences in timing between individuals with a lower versus a higher span. Moreover, Experiment 2, which examined children of two ages and adults on three span tasks, differed from previous research in that preparatory intervals, as well as interword pauses, showed large effects of ability level. The previous research, in which effects of interword pauses, but not preparatory intervals, varied as a function of span (Cowan, 1992, 1999; Cowan et al., 1994, 1998; and the present Experiment 1), had examined individuals from a restricted age range in a single task.

Ability-level differences showed up in preparatory intervals within one age group and in interword pauses within another age group (e.g., see Figure 5). This suggests that there may be different strategies of carrying out a span task. It is possible that processing of a certain type (such as memory search) can be carried out in the preparatory interval, deferred to the interword pauses, or carried out repeatedly in both intervals.

One resolution of these findings would be to suggest that both preparatory intervals and interword pauses include memory search operations but that preparatory intervals also include other processes (e.g., retrieval, rehearsal, response planning, and motor preparation), some of which may not be dependent on list length. Under certain circumstances, if processes independent of list length are long enough, they dominate the intervals and so mask list-length-dependent effects. These latter processes might become more measurable if participants learn to shorten list-length-independent processes. Although there has been some related theoretical work in the case of speeded pronunciation of lists following a start cue (Sternberg et al., 1978, 1980), additional work is needed to help interpret timing in span tasks. What is abundantly clear from all of the results is that longer-duration responses, and especially silent periods within those responses, serve as useful indexes of the difficulty or duration of processing in a span task.

Multiple types of processing are possible in span tasks. This point is underscored by the fact that different WM tasks can be carried out in different ways (e.g., with response times that depend on how easily the processing episodes can be used as retrieval cues) and by the fact that the processes appear to change with development (e.g., with much larger differences between listening- and counting-span response durations in younger children). Researchers tend to operate under the convenient assumption that various tasks that they wish to lump together for theoretical reasons operate similarly when, in fact, experimental participants are free to use any and all processes at their disposal to get a complex task done. This is certainly true for WM tasks. Whereas they have

sometimes been taken simply as a reflection of one processing factor, which has often been equated with a general processing ability (g). Plomin and Spinath (2002, p. 173) criticized this view as follows:

It is increasingly clear that various measures of working memory correlate with g near the reliability of the measure . . . This could mean that working memory is the Factor X that explains g . . . However, it seems more likely that working memory is just another name for g —tests of working memory look suspiciously like psychometric tests of g .

Given that g is based on a collection of skills, the same may easily be true of WM tasks. Bayliss, Jarrold, Gunn, and Baddeley (2003) recently presented evidence that processes, visuospatial storage, and verbal storage all play separate roles in WM-span tasks, and Ackerman et al. (2002) showed that g is separately related to processing speeds as well as WM capability. In light of this recent work, it makes sense that we find differences between the processing mechanisms that contribute to different WM tasks.

Relations to Scholastic Measures

This is the first set of studies to examine correlations between STM or WM recall timing and scholastic or intellectual ability tests. We conclude that recall timing in WM-span tasks picks up substantial variance that is important in intellectual tasks. Yet, it does not appear to be the same variance that the spans themselves pick up and it does not seem to be particularly general in nature. Instead, it appears that the recall durations, and especially the preparatory intervals, may convey information about specific skills.

It is thus reasonable to believe that using response times as well as accuracy measures will increase the ability to predict intellectual performance in other tasks. That was clearly the case in this study. Figures 2 and 6 show two dramatic examples in which the addition of response times in recall greatly increased the predictive power of span tests in comparison to the use of span accuracy measures alone.

Considering that accuracy and reaction time in psychological tasks are generally expected to trade off against one another, it can be viewed as surprising that accuracy and response times in WM-span tasks reflected primarily nonoverlapping pools of variance in accounting for scholastic ability measures. Yet, accuracy and reaction time sometimes do not trade off (e.g., Busemeyer, 1993). Our findings suggest that there may be processing factors that determine the speed at which WM-span responses are made that are not critical for good performance in the span task itself but that, nevertheless, are more important for at least some other scholastic tasks.

Conway et al. (2002) made clear a division between general ability, on one hand, and processing speed, on the other hand, by showing that variance unique to WM spans predicted a g factor in intelligence tests, whereas variance cutting across WM- and STM-span tasks did not do so, but did correlate with processing speed. The present article takes this point farther, distinguishing between two types of speed. Specifically, in Experiment 2, rapid-speaking speeds correlated with spans but not with academic tests or high-school grades. In contrast, the preparatory intervals in span tasks were related to high-school grades in a way that span tasks were

not. This pattern, like those observed by Cowan et al. (1998) and Ackerman et al. (2002), argues against the predominance of a single, global speed of processing (e.g., Kail & Salthouse, 1994). In order to interpret a processing speed, one must consider what processes are involved and what mechanisms speed up or slow down those processes. We are reminded of the story in which a psychologist claims that his research area is "reaction times" and another replies, "Mine is percent correct." One would not consider a global percent correct measure to reflect a specific processing mechanism, and a global reaction time is probably similar.

There are at least two ways to interpret this difference between accuracy and response duration measures. In one interpretation, response duration measures allow a finer gradation of information; response timing may reveal differences even where spans cannot. This interpretation easily accounts for why there is scholastic task variance unique to response timing, but it has more difficulty explaining why there is also considerable variance unique to span (as shown in Tables 2, 5, and 6). Alternatively, span and response durations may reflect different processes. For example, WM spans may reflect the ability to control (Engle et al., 1999) or switch (Hitch et al., 2001) attention, whereas response times may reflect retrieval speeds (e.g., Kail & Salthouse, 1994) and/or the efficiency of memory organization (Ericsson & Kintsch, 1995).

This type of account seems compatible with the particular measures best predicted by retrieval durations. In the Word Reading task modeled in Figure 2, retrieval speed in the reading-span task might have resulted from linguistic knowledge that is more critical for Word Reading performance than it is for WM-task retrieval (given that the processing component of the reading-span task is deliberately made relatively easy). For high-school grades, modeled in Figure 6, similarly, a good retrieval organization might be critical for effective test-taking or note-taking to an extent that is not as critical for the simple information presented in the WM-span tasks; yet, a good retrieval organization might quicken the pace of recall in the span tasks.

It also bears mention that the present data are not heavily consistent with the standard view (Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Engle et al., 1999) that STM span is consistently a poorer predictor of scholastic and intellectual task performance than is WM span. In Experiment 2, we found digit span to predict scholastic tests almost as well as listening span and better than counting span. The outcome may have something to do with the nature of the scholastic tasks against which digit span was being compared. At any rate, there are other studies that have found less-than-striking advantages for a WM-span task (e.g., Hutton & Towse, 2001) or advantages for an STM-span task over a WM-span task (e.g., Booth, MacWhinney, & Harasaki, 2000) in accounting for complex task performance. It may be that some scholastic tasks heavily weigh verbal-domain-specific processing abilities including rehearsal and phonological retention, which are emphasized in digit-span performance. In line with this reasoning, among our adults, digit span did well in accounting for ACT English, Math, and Reading subtests, but not Science, whereas counting span did relatively well in accounting for Science but not well in accounting for Reading. This does suggest that the domain of the memory task is an important factor that must be considered in evaluating the strength of association with scholastic tasks (e.g., Shah & Miyake, 1996), without denying the possibility that a

general, cross-domain factor will also be found (e.g., Engle et al., 1999).

Implications for Models of WM and Its Development

On the basis of a theoretical framework and some data, cognitive psychologists are often motivated to hope and believe that a particular task is a pure measure of a psychological process. This certainly is likely to apply to WM, in which one would like to have a measure of the ability to maintain information in temporary storage without contamination from domain-specific long-term knowledge (e.g., Ericsson & Kintsch, 1995) or general processing abilities (e.g., Kail & Salthouse, 1994) that could assist in the task of holding information temporarily. The hope that a task can be process-pure has encouraged researchers to proceed on the assumption that multiple WM tasks all measure the desired set of processes. One approach has been to suggest that it is the variance common to these different tasks that measures the desired processes (e.g., Conway et al., 2001; Engle et al., 1999). Although there is merit in this, it also seems likely that a more complete task analysis is essential and that response times as well as accuracy information will be important in carrying out this task analysis.

The finding that reading and listening spans produce much longer reaction times than counting span tends to weaken the assumption that WM tasks involve similar processes. Instead, the similarity between WM tasks may only be the fact that they require the coordination of multiple processes (e.g., Bayliss et al., 2003), the specific nature of which could be task-dependent. The processes could sometimes entail rehearsal of the items to be recalled later throughout the processing task, as Daneman and Carpenter (1980) assumed. That type of processing would account for the relatively rapid rate of recall in the counting-span task, rather comparable with simple digit span, in Experiment 2. The processes sometimes could entail the use of episodic information along with the prior knowledge base to reconstruct the items to be recalled from memory. That type of processing would account for the relatively slow rate of recall in the reading- and listening-span tasks of Experiments 1 and 2, respectively. It would be expected by Ericsson and Kintsch (1995) on the basis of their notion of long-term working memory, by Hitch et al. (2001) on the basis of their pattern of responses observed during the processing portion of WM tasks, and perhaps by Baddeley (2000) on the basis of his notion of an episodic buffer, or temporary storage of episodic information. Given the very different processes that may be involved in our WM tasks, it is also understandable that they produced different patterns of correlations with specific scholastic measures (e.g., Tables 5 and 6). If we are to predict complex performance, we must know as much as possible about processing in the WM-task predictors.

Ultimately, psychometric scholastic tests are considered useful only to the extent that they predict performance in school and on tasks that are important in life outside of school. Therefore, high-school grades percentile could be considered to constitute a more important measure than scholastic tests. From this viewpoint, it is a resounding success of the response time measures that they significantly predicted high-school grades, whereas span measures per se did not. Given that high-school grades are relatively easy to examine in a college-student population, considerable progress could be made in understanding higher level cognition by trying to

understand what it is about grades that distinguishes them from test scores, and what it is about preparatory intervals in WM tasks that resulted in their success in predicting high-school grades. There are many possibilities at this point (e.g., a possible role of alertness and motivation in school success and in response speeds; a role of response speeds in taking adequate notes in a classroom setting).

The present study may have implications for the development of WM in children and for the usefulness of development in understanding adult cognition. Regarding the first of these issues, response times may be especially informative regarding skills, other than those emphasized in span scores, that are critical for good scholastic performance early in development (e.g., lexical and linguistic knowledge). Thus, reading-span response durations were more useful than reading spans themselves in predicting BAS Word Reading performance (see Figure 2). In older children and adults, when basic skills become better learned and automatized, more of the predictive power may shift to spans. Regarding the relevance of development for adult cognition, Figure 3 illustrates that differences between span tasks in response durations were small in adults but much larger in young children (cf. Cowan et al., 1998; Kail & Salthouse, 1994). One reason that developmental data are pertinent to understanding cognitive processes in general is that larger effect sizes in children can make analyses more tractable.

In conclusion, it is unlikely that two experiments could capture in full all the relevant aspects of a new measure of WM performance. However, the present studies have documented that recording and analyzing spoken responses in span tasks can be worth the considerable effort. It has shown that span response timing depends on whether the processing portion of the task provides a semantic or lexical context for retrieval (in which case the act of retrieval lasts longer), and it has shown that individual differences in span response timing account for substantial variance in scholastic measures that is not accounted for by span itself. It is not yet clear whether these benefits can accrue if a keyboard response is used or whether a spoken response is necessary. Concentration on the timing of WM processes seems, in any case, to be a priority for the near future.

References

- Ackerman, P. L., Beier, M. E., & Perdue, M. B. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*, 567–589.
- Baddeley, A. D. (1986). *Working memory* (Oxford Psychology Series No. 11). Oxford, England: Clarendon Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
- Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*, 71–92.
- Booth, J. R., MacWhinney, B., & Harasaki, Y. (2000). Developmental differences in visual and auditory processing of complex sentences. *Child Development*, *71*, 981–1003.
- Busemeyer, J. R. (1993). Violations of the speed–accuracy tradeoff relation: Decreases in decision accuracy with increases in decision time. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgment and decision making* (pp. 181–193). New York: Plenum Press.
- Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency

- and the growth of short-term-memory span. *Journal of Experimental Child Psychology*, 33, 386–404.
- Chase, W. G. (1977). Does memory scanning involve implicit speech? In S. Dornic (Ed.), *Attention and performance VI* (pp. 607–628). Hillsdale, NJ: Erlbaum.
- Chuah, Y. M. L., & Maybery, M. T. (1999). Verbal and spatial short-term memory: Common sources of development change? *Journal of Experimental Child Psychology*, 73, 7–44.
- Clifton, C., & Tash, J. (1973). Effect of syllabic word length on memory-search rate. *Journal of Experimental Psychology*, 99, 231–235.
- Conway, A. R. A., Bottoms, B. L., Nysse, K. L., Haegerich, T. M., & Davis, S. L. (2000). *Working memory capacity and distractibility in children*. Manuscript submitted for publication.
- Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, 8, 331–335.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163–183.
- 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.
- Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition*, 29, 774–776.
- Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, 31, 668–684.
- Cowan, N. (1999). The differential maturation of two processing rates related to digit span. *Journal of Experimental Child Psychology*, 72, 193–209.
- Cowan, N., Keller, T., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language*, 33, 234–250.
- Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term-memory span. *Journal of Experimental Psychology: General*, 127, 141–160.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450–466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3, 422–433.
- Doshier, B. A., & Ma, J.-J. (1998). Output loss or rehearsal loop? Output time versus pronunciation time limits in immediate recall for forgetting-matched materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 316–335.
- Elliot, C. D., Smith, P., & McCulloch, K. (1997). *British Abilities Scale second edition (BAS II) technical manual*. Windsor, Berkshire, England: NFER-Nelson.
- 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, and Cognition*, 18, 972–992.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29, 336–360.
- Gathercole, S. E., & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 189–209). Hove, England: Erlbaum.
- Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28, 887–898.
- Henry, L. A. (2001). How does severity of a learning disability affect working memory performance? *Memory*, 9, 233–247.
- Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, 130, 184–198.
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, 30, 685–701.
- Hulme, C., Newton, P., Cowan, N., Stuart, G., & Brown, G. (1999). Think before you speak: Pause, memory search, and trace reintegration processes in verbal memory span. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 447–463.
- Hutton, U. M. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory*, 9, 383–394.
- Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1999). Genetically dissociated components of working memory: Evidence from Down's and Williams syndrome. *Neuropsychologia*, 37, 637–651.
- Jarrold, C., Hewes, A. K., & Baddeley, A. D. (2000). Do two separate speech measures constrain verbal short-term memory in children? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1626–1637.
- Kail, R., & Hall, L. K. (1999). Sources of developmental change in children's word-problem performance. *Journal of Educational Psychology*, 91, 660–668.
- Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, 86, 199–255.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed). Upper Saddle River, NJ: Prentice Hall.
- Klein, K., & Boals, A. (2001). The relationship of life event stress and working memory capacity. *Applied Cognitive Psychology*, 15, 565–579.
- Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, 15, 646.
- McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology*, 74, 240–260.
- Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge, England: Cambridge University Press.
- Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, 31, 723–738.
- Plomin, R., & Spinath, F. M. (2002). Genetics and general cognitive ability (g). *Trends in Cognitive Sciences*, 6, 169–176.
- Schneider, W. (1988). Micro-Experimental Laboratory: An integrated system for the IBM-PC compatibles. *Behavior Research Methods, Instrumentation, and Computers*, 20, 206–217.
- Schweickert, R. (1993). A multinomial processing tree model for degradation and reintegration in immediate recall. *Memory and Cognition*, 21, 168–175.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27.
- Sternberg, S. (1975). Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology*, 27, 1–32.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The

- latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 116–152). New York: Academic Press.
- Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 469–505). Hillsdale, NJ: Erlbaum.
- Swanson, H. L., & Sachse-Lee, C. (2001a). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, 79, 294–321.
- Swanson, H. L., & Sachse-Lee, C. (2001b). A subgroup analysis of working memory in children with reading disabilities: Domain-general or domain-specific deficiency? *Journal of Learning Disabilities*, 34, 249–263.
- Tehan, G., & Lalor, D. M. (2000). Individual differences in memory span: The contribution of rehearsal, access to lexical memory, and output speed. *Quarterly Journal of Experimental Psychology*, 53(A), 1012–1038.
- Towse, J. N., Hamilton, Z., Hitch, G. J., & Hutton, U. (2000). *Sentence completion norms among adults: A corpus of sentences differing in length* (Tech. Rep. CDRG 7). London, England: Royal Holloway, University of London.
- Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, 39, 195–217.
- Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, 28, 341–348.
- Tuholski, S. W., Engle, R. W., & Baylis, G. C. (2001). Individual differences in working memory capacity and enumeration. *Memory & Cognition*, 29, 484–492.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154.
- Yeo, R. A., Hill, D., Campbell, R., Vigil, J., & Brooks, W. M. (2000). Developmental instability and working memory ability in children: A magnetic resonance spectroscopy investigation. *Developmental Neuropsychology*, 17, 143–159.

Received June 1, 2001

Revision received October 30, 2002

Accepted October 30, 2002 ■

ORDER FORM

Start my 2003 subscription to the *Journal of Experimental Psychology: General*! ISSN: 0096-3445

\$33.00, APA MEMBER/AFFILIATE _____
 \$66.00, INDIVIDUAL NONMEMBER _____
 \$168.00, INSTITUTION _____
In DC add 5.75% / In MD add 5% sales tax _____
TOTAL AMOUNT ENCLOSED \$ _____

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO:
 American Psychological Association
 Subscriptions
 750 First Street, NE
 Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.
 TDD/TTY (202) 336-6123.
 For subscription information, e-mail:
 subscriptions@apa.org

Send me a FREE Sample Issue

Check enclosed (make payable to APA)

Charge my: VISA MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

Signature (Required for Charge)

BILLING ADDRESS: _____

City _____ State _____ Zip _____

Daytime Phone _____

SHIP TO:

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____ XGEA13