Individual Differences in Working Memory for Comprehension and Following Directions

RANDALL W. ENGLE
JULIE J. CARULLO
KATHRYN W. COLLINS
University of South Carolina

ABSTRACT The extent to which individual differences in working memory capacity were related to differences in comprehension and following direction was studied in first, third, and sixth graders. Working memory was tested by a simple word span and Daneman and Carpenter's (1980) reading span. A reading comprehension test and a following directions test, including directions heard in a classroom, were also given. The number of words recalled in both the word span and the reading span predicted comprehension for all grades. This relationship showed no age differences. However, the relationship between the span measures and following directions indicated that the role of working memory in following directions may increase with age. As the complexity of directions increased, low-span subjects in each grade had more difficulty performing directions than high-span subjects did. Implications for giving directions to school children are discussed.

The limited capacity working memory has been considered important to the development of general cognitive and social skills in children. For example, Brownell (1988) found that children's ability to combine imitative acts in social and motor play improved with age. That is, as children aged, they could remember and perform more behaviors as a sequence. Brownell found similar results across various task domains and suggested that one mediator of this increase in ability to combine behaviors may be growth in whatever mediates growth in memory span.

Memory span for digits doubles from age 5 to adulthood (Chi, 1977), but the cause of this increase is open to debate. The increase could be caused by increases in capacity (Pascual-Leone, 1970), increases in use of strategies (Chi, 1976; Dempster, 1978), or increases in the efficiency of processing information (Case, Kurland, & Goldberg, 1982). Manifest short-term memory capacity, as reflected by span, increases, and this increase in functional capacity has been an important factor in the development of cognitive and social skills in children (Brownell, 1988; Chao, Knight, & Dubro, 1986; Knight, Berning, Wilson, & Chao, 1987).

Thinking about the development of a limited-capacity processing and storage center dates back at least to Piaget (1928) with his notion of "field of centration." However, Pascual-Leone (1970) formalized thinking along Piagetian lines by proposing that there is a genuine increase in capacity of what he calls the M-operator over the early life span. The theory argues that for every 2 years during childhood, there is an increase of 1 in the number of chunks, subjective units, knowledge structures, or schemes that can be kept active in memory at one time. The number of items that we are able to recall clearly increases during childhood even if chunking and grouping are held constant (Burtis, 1982; Engle & Marshall, 1983). But it is not clear that the increase in recall is a result of an increase in the size of the working memory or M-space. For example, Case, Kurland, and Goldberg (1982) have argued that capacity remains invariant over age, but the processes used to encode and retrieve information during recall tasks are faster and more efficient during childhood. Consequently, these encoding and retrieval operations require less attention, leaving more of the flexibly divided limited-capacity attention to be allocated to the maintenance of the stored items.

These two different explanations to the functional increase in measures of short-term or working memory can also be used to explain why span performance would differ across individuals at a given age or stage of develop-

Address correspondence and requests for reprints to Randall W. Engle, Department of Psychology, University of South Carolina, Columbia, SC 29208.
This research was supported by AFOSR Grant 870069.

ment. In other words, the concepts of developmental theories can be borrowed to explain why individuals at a given developmental level might differ on measures of short-term or working memory. Thus, Daneman and Carpenter's (1980) theory of individual differences in adults borrows heavily from Case's theory.

Short-term memory has commonly been assumed to play a large role in the cognitive functioning of the individual. For example, the comprehension of language seems to rely on our ability to hold particular labels and concepts in memory and to process incoming information so that we might integrate that information over time. Short-term memory capacity has been thought of as a constraint and a source of individual differences in language comprehension (Kintsch & van Dijk, 1978) and problem solving (Newell & Simon, 1972). Thus, one might assume that a smaller working memory capacity would lead to poorer performance in tasks of higher cognitive functioning, such as reading.

From the outset of modern work on short-term memory, memory span and its variants have been assumed to be the prototypical measure of short-term memory capacity. Memory span had been assumed to measure something important about memory and its role in higher level cognition long before current thinking about short-term or working memory began (Brener, 1940; Wechsler, 1944). Therefore, it follows that measures of memory span should be related to measures of higher cognitive functioning such as language comprehension and reading. But span, at least digit span, has an inconsistent pattern of correlations with language comprehension (Dempster, 1985).

However, complex memory span measures that require processing as well as storage of information do correlate consistently with language comprehension and reading. Daneman and Carpenter (1980) used the same argument about individual variation in working memory capacity that Case, Kurland, and Goldberg (1982) used to explain developmental differences. Daneman and Carpenter assumed that there is a trade-off between processing and storage such that an individual with more efficient processes on a task like reading could give more attention to the information to be stored. An individual with efficient reading processes would have a larger storage component than would an individual with inefficient reading processes. For a span task to predict performance on a higher level cognitive task, however, the span task must incorporate both processing and storage demands, with the processing being similar to that required during the higher level task. In other words, the memory span would predict reading performance only if the span task itself included some component that also required reading.

Daneman and Carpenter (1980) used a complex span task that required subjects to read sets of increasing numbers of sentences aloud and to recall the last word of each sentence after each set. The reading span score predicted performance on tests of reading and listening comprehension and the Verbal Scholastic Aptitude test. Daneman and Carpenter argued that good readers do better on the reading span because they have more efficient reading skills and thus more residual capacity to remember the last words.

Recently, Turner and Engle (1989) demonstrated that the relationship between the complex span score and reading comprehension does not require that the processing component of the complex span be similar to the task being predicted, in this case, reading comprehension. The authors had subjects decide whether an arithmetic string like $(15/3) - 4 = 5$ was correct. Immediately after the decision (yes or no), the subject was given a word that was to be recalled at the end of the set of operations. Counter to the predictions of the Daneman and Carpenter (1980) explanation of individual variation in working memory capacity, this operation-word span score predicted reading comprehension as well as did a reading span score. the result suggests that the complex span tasks reflect a difference in working memory capacity that is general to at least verbal tasks requiring the maintenance of information over time. The significant correlation between span and reading comprehension does not require that the processing component also be reading in nature. Turner and Engle argued that the complex span significantly predicted reading comprehension, because it measured the number of items that could be kept active in memory without the complicating factors of rehearsal and grouping. Rehearsal and grouping are strategies that increase performance in the simple span task but are probably irrelevant to reading comprehension and other higher level cognitive tasks.

The research presented here has several goals. One goal was to determine the extent to which individual differences in working memory capacity predicted performance on higher level cognitive tasks in children, and whether this relationship was markedly different from that typically found with college-aged subjects. Daneman and Blennerhasset (1984) studied working memory capacity and language comprehension in 3- to 5-year-old children. They found that, whereas simple span significantly predicted listening comprehension, the correlation between comprehension and the complex measure of working memory was higher. One problem with the study, however, was that the complex span required the child to recall the entire sentence and not just the last word. Recall of sentences likely would be affected by the extent to which the child understood the sentence, in which case the span score would be another measure of comprehension and, obviously, one measure of comprehension predicting another measure of comprehension is hardly interesting.

A second goal was to determine whether the capacity reflected by this complex span is important for higher level tasks other than reading and listening comprehension. In addition to listening and reading comprehension, we studied following directions. Being able to do something we have been told or shown how to do is a pervasive instance of everyday cognition; working memory capacity should be a limiting factor in our ability to do so. However, little work has been done on following directions in general, or in particular, the extent to which working memory is involved in the ability to follow directions. The developmental work that has been done on following directions has involved simple actions repeatedly presented to the subject for recall (Foellinger & Trabasso, 1977).

Early studies of following directions referred to this skill as "memory for commissions," and they generally investigated its relationship to intelligence. Binet and Simon (1912, cited in Kaplan & White, 1980) and Thorndike (1927) both noted the importance of the ability to follow directions by including such items on intelligence tests. Brener (1940) studied the ability to follow written directions as a measure of memory span. He found positive correlations between direction span scores and other span measures. Those studies suggest that working memory capacity may be a source of individual variation in direction following. In our study, we examined following directions as a higher level cognitive task and the role of working memory capacity in following directions.

Kaplan and White (1980), in one of the few analytical studies of following directions, studied typical classroom directions and what types of directions children at different ages could follow, The authors collected and analyzed the content of classroom directions and, on the basis of their analysis, developed a test of direction-following ability. Recognizing the need to consider both linguistic complexity and the processing load of directions, Kaplan and White described directions in terms of the number of behaviors and qualifiers that a direction contained. Accordingly, directions could include varying numbers of qualifiers and behaviors, and difficulty of a single direction would be indexed by the number of these components. Kaplan and White's results indicated that children are able to follow more complex directions as they get older, although the improvement continues only through second grade. However, the conclusion probably resulted from a ceiling effect in their measure.

We used Kaplan and White's (1980) procedure to measure following directions in the present study. We suspected that increases in ability to follow more complex directions with age are related to increases in memory span. Individuals with lower span scores, hence smaller working memory capacities, should have more trouble with complex directions than would individuals with higher span scores. We expected that the complex span measure, and possibly the simple span measure, would correlate the performance on a direction-following task in much the same way that such measures would correlate with reading comprehension.

Basically, the design of the study was as follows: subjects from first, third, and sixth grades were given simple word-span and reading-span tasks and then evaluated on tests of listening (first graders) or reading comprehension (all other groups) and following oral directions. All the subjects were given two working memory tests consisting of a simple word-span test and the reading-span test. On the second day of testing, all the children were tested on a grade-appropriate comprehension test and a following-directions task adapted from the test given by Kaplan and White (1980).

Method

Subjects

The participants were 120 children randomly selected from two school districts in Columbia, South Carolina. The subjects included 40 first graders (17 boys and 23 girls), 40 third graders (21 boys and 19 girls), and 40 sixth graders (18 boys and 22 girls) with mean ages of, respectively, of 7.07 (SD = 0.45), 8.96 (SD = 0.54), and 11.87 (SD = 0.59) years. Two experimenters tested the subjects, with each running half of the subjects in each grade. The study was approved by university and school district review boards, and appropriate written informed consent was obtained from parents and children.

Design and Materials

Subjects completed five tasks: the Wepman Auditory Discrimination test, a word-span task, a listening or reading comprehension test, and a following directions test. The subjects were tested individually in two 45-min sessions, separated by at least one day.

The Wepman Auditory Discrimination test consists of 40 one-syllable word pairs, with the words in each pair differing by one phoneme. The Wepman test was used as a precautionary measure to screen subjects for possible hearing or speech discrimination difficulties.

The span tasks were constructed from two lists of 75 words each. All the words were common, one- and two-syllable, concrete nouns. Stimuli for both span tasks were typed in 14-point black letters and presented on 4 × 6 index cards. The sentences used in the sentence-word-span task were constructed so that each sentence ended in one of the words from the two lists of words. In addition, sentences varied for each of the three grades. First-grade sentences were 4 to 6 words long and consisted of subject, verb, and object (e.g., "Jake climbed the pole."). The sentences were then modified for the 3rd and 6th graders. Third-grade sentences were 6 to 9 words long and were constructed by adding adjectives (e.g., "Jake climbed the tall slippery pole."). Sixth-grade sentences were 9 to 12 words long and were constructed by adding adjectives, adverbs, and negative, passive voice (e.g., "Jake was unable to climb the tall greased pole.").

The reading comprehension test for each of the three grades was composed of passages and questions taken from the Analytical Reading Inventory and the Burns/Roe Reading Inventory. Passages in both inventories are normed to grade level. Seven passages were used at each of the three graded levels (i.e., first, third, and sixth grades). Passages were altered, if necessary, to include a pronoun reference. The passages were typed, double-spaced, on standard paper. Questions taken from the inventories tested either for main idea, detail, cause and effect, or inference. An additional question testing pronoun reference was added to the questions for each passage.

The following directions test was constructed on the basis of a direction game developed by Kaplan and White (1980). Their test consisted of classroom directions that were classified according to the number of behaviors and qualifiers that an individual was required to follow simultaneously. For example, 'Point to the picture at the top of page three and copy it twice'' is a 2 + 3 direction (i.e., a direction consisting of two behaviors and three qualifiers), where ''point'' and ''copy'' are the behaviors and ''at the top,'' ''page 3,'' and ''twice'' are the three qualifiers. Directions consisted of pencil-and-paper tasks as well as action-oriented tasks such as ''Sit on the floor Indian style.'' In the present study, 45 directions similar to those used by Kaplan and White were given. The number and type of directions given were as follows: four each of 1 + 0; two each of 1 + 1, 1 + 2, 1+ 3, 2 + 0, 2 + 1, 2 + 3, 2 + 4, 3 + 0, 3 + 1, 3 + 2, 3 + 3, 3 + 4, 4 + 1, 4 + 2, 4 + 3, and 4 + 4. We added one each of 5 + 2, 5 + 3, 5 + 4, 5 + 5, 6 + 4, 6 + 5, and 6 + 6 directions to ensure against a possible ceiling effect. Difficulty level was not considered in the order of the directions. The 45 directions were arranged randomly, with the constraint that those directions requiring the same setting or materials were grouped together. Further, within those groups of directions requiring the same setting, the presentation of directions was random.

Procedure

All 120 subjects were given the Wepman test and the two span tasks during the first testing session. The Wepman test was always administered first. The span tasks were then given, with the order of the tasks as well as the order of the two word lists counterbalanced across subjects.

The reading comprehension test and the following directions test were given during the second testing session, with the order of the tasks counterbalanced across subjects. Half of the subjects receiving the word-span task first received the following directions test first, and half received the reading comprehension test first. Half of the subjects receiving the sentence-word-span task first received the reading comprehension first, and half received the following directions test first.

The Wepman Auditory Discrimination test. The test was administered according to the instructions provided by the author. The experimenter read the word pairs aloud, one pair at a time, to the subject who was faced away from the experimenter. The subjects were told to respond ''same'' or ''different'' to the pairs of words, and the experimenter recorded their responses.

Word-span task. The subjects were shown words on index cards at approximately a one-word per s rate. Each of the subjects received a different random order of words that was achieved by shuffling the cards on which the words were printed before each subject was tested. The words were groups from two to five words per list, with five trials at each of the list lengths (e.g., five trials of two words per list, five trials of three words per list, etc.). Lists of words were separated by a blank index card. The word-span task was administered in all three grades in the same manner, except that 1st graders listened as the experimenter read the words aloud, whereas the 3rd and 6th graders read the words aloud themselves. The subjects were instructed that when a blank card was shown they were to tell the experimenter the words that they could remember from the list. The students were allowed to recall the words in any order and were given as much time as they needed to recall the words. Recall was oral, and the experimenter recorded the subjects' responses to prevent difficulty with poor writers. The subjects were given two or three practice trials of two or three words per list. The practice words were not recorded or analyzed.

Sentence-span task. The administration of the sentence-span task was similar to the word-span task. Sentences were shown one at a time on index cards, and each of the subjects received a random order of sentences. The 1st graders listened as the experimenter read sentences aloud, whereas the 3rd and 6th graders read sentences aloud at their own pace. Like the word-span task, the sentences were grouped from two to five sentences with five trials at each of the group levels. Groups of sentences were separated by a blank index card. Subjects were instructed either to listen as the experimenter read the sentences or to read the sentences aloud. The students were told to recall the last word of each of the sentences in a group when they saw a blank card and to recall the critical words in any order that they wished. Again, recall was unpaced and oral, and the experimenter recorded the responses. Two or three practice trials with one or two sentences per group were given. The practice trials were not recorded or analyzed.

Reading comprehension test. The subjects were given four passages randomly selected from the seven for the appropriate grade level. The subjects were handed passages, one at a time. The 1st graders followed along as the experimenter read the passages aloud. The 3rd and 6th graders read passages silently. After each passage was read, the passage was removed, and the experimenter asked the subjects questions about the passage. The subjects answered the questions aloud and the experimenter recorded their answers.

Following directions test. The subjects were told that they were playing a game, the object of which was to see how well they could follow directions. The subjects were instructed to listen carefully to each direction read by the experimenter and then try to follow it as best as they could. In addition, subjects were told a direction could not be repeated and that they should try to follow the direction in the order it was given. They were then encouraged to follow any part of a direction that they could remember even if they could not remember the order of the direction. All material (a booklet, paper, pens, pencils, crayons, and paper shapes) needed for subjects to successfully complete each direction were provided. As subjects performed the directions, the experimenter recorded the behaviors and qualifiers that the subjects carried out by marking a corresponding score sheet with + or − for parts of the direction followed.

Scoring

A total of 17 scores for each of the 120 subjects were derived from the Wepman test, the two span tasks, the reading comprehension test, and the following directions test.

The Wepman test score was the number out of 40 word pairs that subjects responded to positively (39 pairs were scored because, in our part of the world, almost no one can say or hear pin and pen as different). Of the 120 subjects tested, only one subject had trouble with the test. When the subject was retested by the school speech teacher, the subject's score fell within the normal range and the subject's data were included. Because this test was included only as a precautionary measure, and all subjects scores fell within the normal range, we will not discuss the test further.

Four different scoring methods were used for both the word-span and sentence-span tasks, resulting in eight span scores for each subject. The results included the span score, the perfect span score, the absolute span score, and the total span score. The *span score* was derived by determining the level at which subjects recalled all words correctly on three of the five trials. The subjects were given partial credit if they correctly recalled all words in one or two trials at the next trial level, and no credit was given for any correct trials after the three-out-of-five criterion was met. For example, if a subject correctly recalled all the words on the *two* trials and only got three of the *three* trials correct, the span score would equal three. The *four* trials would then be examined. If three of those trials were correct, then the score would be increased to four. However, if only two of the *four* trials were perfect, the score would be 3.50, and if only one of the *four* trials was perfect, then the span score would be 3.25. If fewer than three of the *four* trials were perfect, the *five* trials would not be considered. The highest possible score using this method was five.

The *perfect span score* was based on Broadbent's (1971) admonition that it is the correct way to measure the capacity of short-term memory. The score was determined by giving subjects credit only if they correctly recalled the words from all five trials. No partial credit was given if four or fewer trials at a level were correct. For example, if a subject recalled all five trials of two words correctly and failed to recall all five trials at three words correctly, the perfect span score would be 2. The *four* and *five* trial would not be considered because the subject did not recall the *three* trials perfectly. The maximum score was 5. The *absolute span score* was calculated by summing the words recalled correctly from all of the perfect trials. The latter score is similar to the span score because it takes into account only perfect trials. However, the absolute span score also includes trials beyond the three out of five criterion in which recall of a trial was perfect. For example, if a subject recalled three out of five trials correct at the *two* and *three* trials and then only recalled one perfect *four* trial, the absolute score would be 19 [$2 \times 3) + (3 \times 3) + (4 \times 1)$]. The highest possible score was 70. The *total span score* was derived by summing the total number of words recalled correctly regardless of whether the whole trial was correct. The maximum total score was 70.

The overall reading comprehension score was determined by dividing the number of a subject's correct responses on questions from all four passages by the number of possible correct responses. Individual scores for each of the five types of questions (i.e., pronoun reference, fact, main idea, inference, and cause and effect) were derived in the same manner as the overall reading score.

The total directions score was the total number of behaviors and qualifiers performed, regardless of order, out of 231 possible. The perfect directions score was the number of directions out of 45 that subjects performed perfectly, regardless of order.

Results

Reliability Analysis

A relatively crude "split part" analysis of the reliability of the span scores and the scores from the directions test was performed. Five different trials were given at each list length for words and sentences. One trial at each list length was randomly chosen without replacement, and these were aggregated into a single score. Thus, each subject had five such scores, which were analyzed using Cronbach's alpha. The reliability of the directions task was analyzed by taking those directions for which there were two directions at a given difficulty level, as determined by the number of behaviors plus the number of qualifiers. A total of 18 directions contributed to the analysis. All reliability estimates are shown in Table 1.

The reliabilities for the span scores are all in the mid-to-high 80s, with one exception, and there appears to be little difference across age or between the simple word and sentence-span scores. The directions task appears to be somewhat less reliable. It should be kept in mind, however, that the number of directions contributing to the analysis was small.

Descriptive statistics. All of the analyses reported below were performed on all the different scoring procedures described above. There were no substantial differences in the conclusions drawn from the different scoring procedures for the word and sentence spans, so we have confined the reporting to the analyses of the absolute span score. Our method consisted of summing the number of words from all the trials that led to perfect recall, and the possible range for that score is 0 to 70.

For the reading comprehension test, we discuss the results of the overall score, which was the percentage of all questions answered correctly, but, where appropriate, we report certain analyses of the specific question types. For the following directions task, we report on the analysis of the total directions score. This measure was the total number of qualifiers and behaviors performed correctly, as determined by free scoring. The highest possible score was 231. The descriptive statistics for the span measures, reading task and directions task, are shown in Table 2.

Several points should be noted about the descriptive statistics shown in Table 2. We would expect certain scores to increase over age but other scores should be relatively stable. Word span and following directions all should show an increase over age because all the subjects received the same task. However, the reading comprehension scores should be relatively similar if the passages were chosen correctly for the different groups of subjects. Because the sentences used in the reading span test were adjusted for complexity for the different groups, one might expect the number of words recalled for the three groups of children to be similar. Case et al. (1982) and Daneman and Carpenter (1980) would argue that, if one chose sentences for this task that were perfectly matched in difficulty for the different ages, the number of words recalled should be equal. The test results in Table 2 conform to those expectations, with one notable exception. For all four tasks, the 3rd graders' score was somewhat lower than one might have expected.

There was a significant increase in the simple word span from first to sixth grade, but the 3rd graders were no different than the 1st graders, $F(2, 117) = 13.4, p < .001$. The 1st and 6th graders performed equally on the sentence span test, but the 3rd graders were significantly lower, $F(2, 117) = 6.6, p < .01$. Overall reading scores for the 1st and 6th graders were not significantly different, but the third-grade scores were significantly lower than those for the sixth grade, $F(2, 117) = 4.4, p < .05$. The results suggest that the passages given to the 1st and 6th graders were appropriate to graded level, but those given to the 3rd graders may have been relatively more difficult. Performance on the directions task showed the expected significant improvement over age, $F(2, 117) = 33.7, p < .001$.

Table 1.—Reliability (Cronbach's Alpha) for Span and Directions

Grade	Word span	Sentence span	Directions
1	.83	.87	.76
3	.75	.81	.71
6	.84	.87	.61

Table 2.—Descriptive Statistics for Span Measures, Reading Comprehension, and Directions

	Grade	*M*	*SD*	Min.	Max.
Word span	1	37.15	10.89	10	60
	3	37.93	10.44	19	65
	6	49.15	13.26	19	70
Sentence span	1	18.66	9.38	2	46
	3	12.18	6.57	4	34
	6	18.18	10.34	2	56
Reading comprehension	1	58.93	15.08	20	88
	3	55.75	17.24	16	94
	6	66.25	16.28	33	91
Directions	1	183.93	18.01	119	211
	3	200.85	11.95	174	224
	6	209.23	11.80	166	229

All of our conclusions were based on ANOVAs and Tukey HSD post hoc comparisons on the appropriate scores, with significance established at the .05 level.

The 1st graders listened to all the stimuli for all their tasks, but the 3rd and 6th graders had to read the stimuli for all their tasks. The reading requirement for the 3rd graders may have been sufficiently demanding, even with grade-appropriate materials, such that all the tasks for the 3rd graders were functionally harder than those for the 1st and 6th graders.

Correlational Analysis

Listening/reading comprehension. The principle mission of this study was to measure the relationship between the two word-span measures and the two higher level cognitive tasks. Comprehension (listening/reading) and following directions, and to determine whether the relationship differed across age. Table 3 shows the correlations between the two span measures in the top panel, the simple word span and comprehension in the middle panel, and between-sentence word span and comprehension in the bottom panel. The top panel clearly indicates that the two span scores were closely correlated, and the magnitude of the correlation was similar for the three different groups of children.

The real punchline of this study can be seen in the first row of the next two panels—the row containing the correlations between the simple and complex word spans and overall comprehension for each group of subjects. The correlations between simple word span and comprehension were all significant and not statistically different from one another. Looking at subtests of the comprehension measure, one can identify only four significant correlations with simple word span and not much of a pattern to the results.

The bottom panel of Table 3 shows the correlations between the complex sentence span and comprehension. The first row shows the results for the overall comprehension score and, again, all three grades showed significant correlations with comprehension and were similar and not statistically different from one another. Further, the set of correlations between sentence span and comprehension were not significantly different from those between simple word span and comprehension (true for all three age groups). A look at the correlations between the various subtests of the comprehension measure and the sentence span indicates that 9 of the 15 possible subtests correlated significantly with sentence span. The one developmental trend apparent in these data was that the two groups of older subjects showed more correlations between the sentence span and the subtests (8 of the 10 possible) than did the 1st graders (1 of the 5 possible).

Table 3.—Correlations Among Word Span, Sentence Span, and Reading Scores

	Grade		
	1	3	6
Word span			
Sentence span	.61**	.59**	.66**
Overall comprehension	.42*	.36*	.41*
Pronoun	.16	.29	.28
Main idea	.22	.23	.19
Fact	.42**	.24	.39*
Inference	.18	.35*	.16
Cause and effect	.12	.10	.31*
Sentence span			
Overall comprehension	.54**	.53**	.44**
Pronoun	.24	.46*	.43*
Main idea	.24	.47*	.19
Fact	.56**	.47*	.38*
Inference	.28	.22	.33*
Cause and effect	.18	.37*	.36*

*$p < .05$; **$p < .005$.

Table 4.—Correlations of Various Tasks With the Total Directions Score

	Grade		
	1	3	6
Total directions score			
Overall comprehension	.46**	.67**	.52**
Word span	.19	.48**	.51**
Sentence span	.34*	.27	.48**
Composite span (Word + sentence scores)			
Overall comprehension	.53**	.50**	.54**
Total directions	.30	.42**	.47**

*$p < .05$; **$p < .005$.

One finding that was consistent for both the simple span and the sentence span across the three groups was that they predicted recall of facts better than performance on the other subtests.

Because there was no developmental difference in the correlations between simple word and sentence span with the overall comprehension score, we did an analysis of this relationship on all 120 subjects (collapsed over age). The one important conclusion from the analysis was that the relationship between sentence span and comprehension ($r = .50$, $p < .001$) did not differ statistically from the relationship between simple word span and comprehension ($r = .45$, $p < .001$). This conclusion, of course, is counter to previous reports and will be discussed below.

Following directions. The correlations of the various tasks with the total directions score are shown in Table 4. Most important is that performance on the directions task showed significant and sizable correlations with performance on the comprehension task for all three age groups. For the span results of the 1st graders, simple word span did not predict the directions score, but sentence span did significantly predict this score. However, the reverse was true for 3rd graders; simple span predicted performance on the directions task, and sentence span did not. For the 6th graders, the two different span measures both showed significant and similar correlations with directions following. Apparently, there may be a developmental difference in the relationship between working memory capacity and following directions with working memory taking on increased importance over age.

The composite of the two span scores perhaps gives a clearer picture of the relationship with age. Table 3 shows that the correlations between the two span tasks was high and about the same for all three age groups (.61 to .66). Therefore, the two scores were converted to Z scores and added together for each subject to form a composite score.

The last row of Table 4 contains the correlations between the composite span scores and the total direction scores. The correlations range from a nonsignificant .30 for the 1st graders to a .47 for the 6th graders. This should be compared with the relationship between the composite score and the comprehension task, in which the correlations were all above .5 and virtually identical for the three age groups. Whatever is measured by the span tasks is important to listening and reading comprehension and is equally important across the age range studied here. However, following directions is different, because there is an apparent age effect in the importance of whatever is measured by the span tasks.

The directions varied in difficulty, as defined by the number of behaviors and qualifiers specifying the direction. Working memory should not be so important to the following of relatively simple directions. However, as the difficulty and complexity of the direction increase, the importance of working memory capacity should also increase. To determine whether this occurred, we analyzed the directions score of high- and low-span subjects as a function of the difficulty of the direction. We chose high-span subjects based on the upper quartile of the absolute score on the sentence-span task and low-span subjects based on the lower quartile. As the directions became more difficult, the older subjects stayed at ceiling longer than the younger subjects did. But eventually, all three groups showed a divergence of high- and low-span subjects, with the low-span subjects suffering more from the difficult directions (see Figure 1).

The purpose of this study was to determine whether the relationship that has been reported between measures of working memory capacity and higher level tasks like listening/reading comprehension with adults would extend to children. Our results show clearly that (a) the number of words recalled in a simple word-span task or as part of a sentence-span task significantly predicted comprehension, (b) there were few or no age differences in this relationship, and (c) the simple word span and sentence span were not different in their ability to predict listening/reading comprehension.

The role of working memory in following directions seemed to increase over the grades tested. By the sixth grade, both simple and complex-span tasks predicted directions following well. When the subjects were divided into high and low spans, all three grades showed that the more complex directions hurt low-span subjects more than they damaged the high-span subjects.

Discussion

This study was not directed at answering the question of whether the manifest increase in short-term memory capacity found over the early years is a result of true increases in capacity, increases in speed, or improvement in rehearsal and grouping strategies. The increase in functional capacity was demonstrated in our study by the increase in the word span from the first and third to sixth grades. However, the sentence-span task, in a sense,

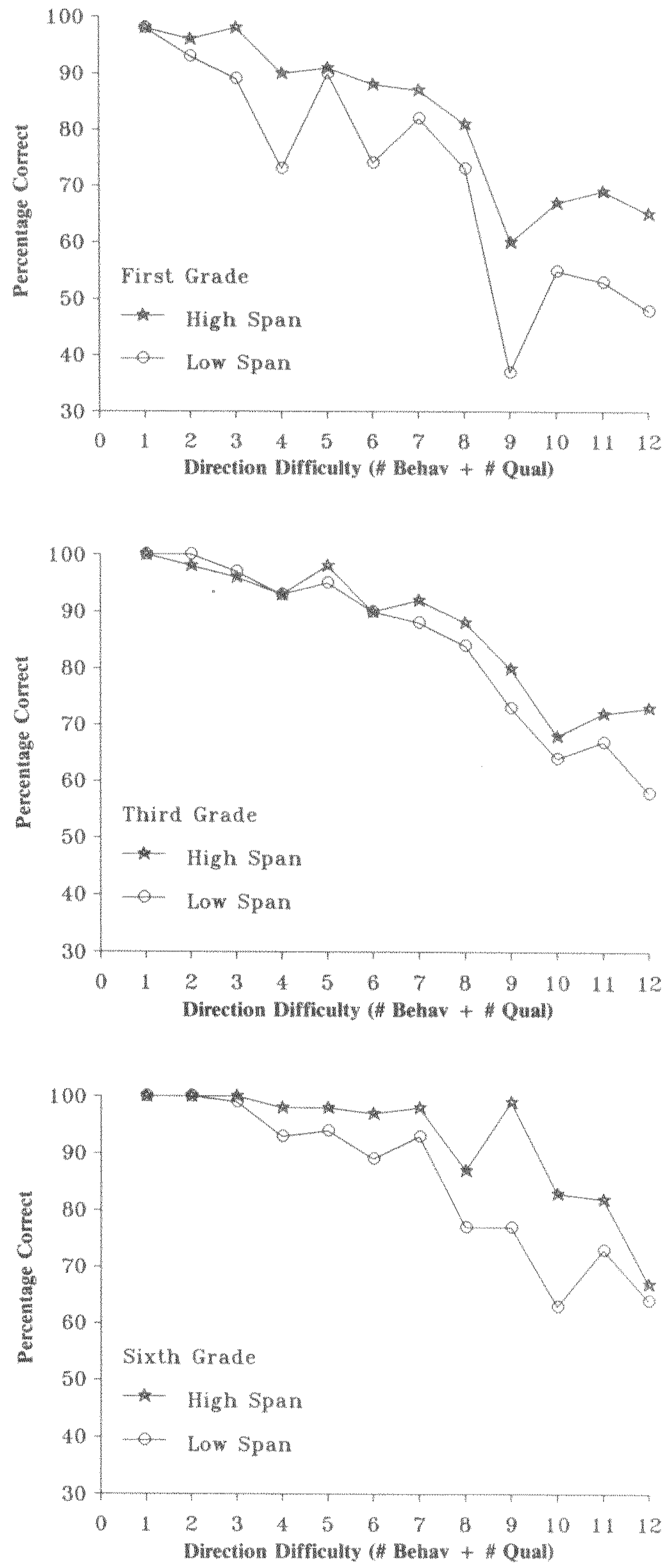

Figure 1. Mean Percentage Correct on the Directions Task for the Three Groups as a Function of the Difficulty of the Directions

equated the three groups of children for processing difficulty and showed a comparable level of recall for the 1st and 6th graders. One interpretation of our finding is that, when the processing component of working memory is not allowed to vary, the functional storage capacity of working memory does not differ over the ages studied here. The implication is that, even though children become more proficient procedurally, the storage component of working memory does not increase developmentally. Even with the groups equated for the complexity of the sentences, however, there was still within-group variation in the number of words that could be recalled in the sentence-span task. Further, variation in the task predicted variation in the comprehension and, for the older children, following directions task.

The within-group variation in span might, like the developmental differences, also be a result of differences in procedural proficiency, and not storage capacity. However, Turner and Engle (1989) showed that the significant relationship between span and comprehension in adults does not require that the background component of the span task be procedurally identical to the comprehension task. Therefore, we must conclude that the within-age variation in the span score is either (a) a result of individual variation in storage capacity or (b) a result of variation in some processing component common to tasks other than reading and listening comprehension. The component could be basic phonological coding of written and spoken language such as the articulatory loop proposed by Baddeley and Hitch (1974).

One of the questions posed by this research was whether the previously reported superiority of the sentence span over the simple span in predicting comprehension (Daneman & Carpenter, 1980; Turner & Engle, 1989) would be eliminated with children. Daneman and Carpenter (1980) compared a simple word span with the sentence span. They found that the sentence span was a better predictor of comprehension in college students than the word span was. The same result was obtained by Turner and Engle (1989). In the present study, the simple word span was as good a predictor for the higher level tasks of comprehension and following directions as was the sentence span. Our finding was novel and unexpected, but several points need to be made about it.

Subsequent research in our laboratory has shown that the significant correlation between simple word span and comprehension is the norm rather than the exception. Since collection of the data reported in Turner and Engle (1989), we have completed over 20 studies with college students that directly compare the simple word span as described here with a complex span task—either a sentence span like the one from this study or an operation span task like that used by Turner and Engle (e.g., Engle, Nations & Cantor, in press; La Pointe & Engle, in press). In nearly all of those studies, the simple word span significantly predicts performance on a measure of reading comprehension. Typically, the magnitude of the correlation between word span and comprehension is slightly smaller than that between the complex span and comprehension and, over studies, is more variable, but is nearly always significant.

Thus, we cannot view the findings obtained here as aberrant. The simple word span is not as consistent as the complex spans in predicting higher level cognitive tasks, but the difference between the two tasks is not as great as has been reported previously. The simple word task may just be less reliable and more prone to the recognition of idiosyncratic associations and patterns in the word stimuli than is the complex task. We would also like to call attention to the fact that the word span, as measured by us and Daneman and Carpenter (1980, 1983) is inherently different from the way that span has traditionally been measured. With our procedure, the subject does not get words repeated from list to list as was traditionally done with span tasks using words, digits, and letters. The difference may be important to the predictive value of span tasks, and our laboratory is investigating that possibility. Nevertheless, the span tasks used here appear to be measuring something that is important to higher level tasks, at least those involving verbal processing. Further, the tasks have as much predictive value for comprehension in first graders as for adults.

It is not clear why the predictive value of the span tasks for following directions increased from the 1st to 6th graders when the relationship with comprehension was comparable over grades. One possibility is that the factor that both span tasks measure that is important to comprehension is some form of speech or phonological coding. Possibly, even by the middle of the first grade, children are frequently tested on story or reading-like information in such a manner that encourages the retention of speech or phonological information. Active phonological coding is probably not encouraged for following directions, and it would thus develop later.

Do the results have implications for educational practice? To the extent that short-term memory and working memory refer to similar constructs, we know that working memory is limited in capacity, and the limitation is probably smaller than the 7 ± 2 proposed by Miller (1956). The limitation is probably on the order of 3 to 4 propositions or knowledge structures that can be kept active at one time. The developmental and individual differences research, like that presented here, suggests that the limitation will be smaller for some individuals and larger for others. The amount of information in working memory can be functionally increased by chunking or grouping the material. The procedure can be done by the learner, but studies teaching children to use grouping and chunking strategies have not shown much generalization from one task to another. The teacher can increase chunk size by building upon materials that the learner already knows. We know that individuals use larger chunks when working with material that they understand well. So prior knowledge should be used whenever possible, along with new knowledge. Thus, reading materials on topics known by the child should lead to larger chunks in working memory at one time. In spelling classes, more letters will be represented if the child has a good knowledge of spelling patterns than if representation is at the individual letter level.

When working with materials of a verbal nature, children tend to rely on the phonological code to represent that material in working memory. The reliance seems to be particularly important to the retention of surface-level knowledge and the retention of word order. Thus, the task should minimize sound-alike confusions within the same grouping, phrase, or constituent.

Teachers need to be acutely aware that the directions they provide must be stated in such a way that minimizes the burden on working memory. The directions should, if possible, be structured so that grouping can be used and should be packaged with no more than three or four propositions per direction. Also, the directions should not include many phonologically confusing words. Most important, the teacher must be aware that there are individual differences in working memory capacity and that this factor will affect the ability of the child to process information in many different tasks.

REFERENCES

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

Binet, A., & Simon, T. (1912). A method for measuring development of intelligence of young children. Lincoln, Illinois: The Courier Co.

Brener, R. (1940). An experimental investigationof memory span. *Journal of Experimental Psychology, 26*, 467–482.

Broadbent, D. E. (1971). The magic number seven after fifteen years. In R. A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory*. New York: Wiley.

Brownell, C. A. (1988). Combinatorial skills: Converging developments over the second year. *Child Development, 59*, 675–685.

Burns, P. C., & Roe, B. D. (1985). *Burns/Roe informal reading inventory: Preprimer to twelfth grade* (2nd ed.). Boston: Houghton Mifflin.

Burtis, P. J. (1982). Capacity increase and chunking in the development of short-term memory. *Journal of Experimental Child Psychology, 34*, 387–413.

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

Chao, C. C., Knight, G. P., & Dubro, A. F. (1986). Information processing and age differences in social decision-making. *Developmental Psychology, 22*, 500–508.

Chi, M. Y. H. (1977). Age differences in memory span. *Journal of Experimental Child Psychology, 23*, 226–281.

Chi, M. Y. H. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory and Cognition, 4*, 59–572.

Daneman, M., & Blennerhasset, A. (1984). How to assess the listening comprehension skills of prereaders. *Journal of Educational Psychology, 76*(19), 450–466.

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

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

Dempster, F. N. (1978). Memory span and short-term memory capacity: A developmental study. *Journal of Experimental Child Psychology, 26*, 419–431.

Dempster, F. N. (1985). Short-term memory development in childhood and adolescence. In C. J. Brainerd & M. Presseley (Eds.), *Basic processes in memory development: Progress in cognitive development research*. New York: Springer-Verlag.

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

Foellinger, D. B., & Trabasso, T. (1977). Seeing, hearing, and doing: A developmental study of memory for actions. *Child Development, 48*, 1482–1489.

Kaplan, C. H., & White, M. A. (1980). Children's direction-following behavior in grades K-5. *Journal of Educational Research, 74*, 43–48.

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

Knight, G. P., Berning, A. L., Wilson, S. L., & Chao, C. C. (1987). The effects of information-processing demands and social-situational factors on the social decision making of children. *Journal of Experimental Child Psychology, 43*, 244–259.

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

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Pascual-Leone, J. A. (1970). A mathematical model for transition in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.

Piaget, J. (1928). *Judgment and reasoning in the child*. New York: Harcourt, Brace & World.

Thorndike, E. L. (1927). *The measurement of intelligence*. New York: Teachers College, Columbia University.

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

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

Wepman, J. M. (1958). *Auditory discrimination test*. Los Angeles: Western Psychological Services.

Woods, M. L. & Moe, A. L. (1985). *Analytical reading inventory* (3rd ed.). Columbus: Charles E. Merrill.

Reprinted from **The Journal of Educational Research**, published by HELDREF PUBLICATIONS, 4000 Albemarle Street, N.W., Washington, D.C. 20016.