Individual Differences in Working Memory and Reading

MEREDYTH DANEMAN AND PATRICIA A. CARPENTER

Carnegie–Mellon University

Individual differences in reading comprehension may reflect differences in working memory capacity, specifically in the trade-off between its processing and storage functions. A poor reader's processes may be inefficient, so that they lessen the amount of additional information that can be maintained in working memory. A test with heavy processing and storage demands was devised to measure this trade-off. Subjects read aloud a series of sentences and then recalled the final word of each sentence. The reading span, the number of final words recalled, varied from two to five for 20 college students. This span correlated with three reading comprehension measures, including verbal SAT and tests involving fact retrieval and pronominal reference. Similar correlations were obtained with a listening span task, showing that the correlation is not specific to reading. These results were contrasted with traditional digit span and word span measures which do not correlate with comprehension.

Many theorists have suggested that working memory capacity plays a crucial role in reading comprehension (cf. Just & Carpenter, in press; Kintsch & van Dijk, 1978); however, traditional measures of short-term memory, like digit span and word span, are either not correlated or only weakly correlated with reading ability (cf. Perfetti & Lesgold, 1977). The current paper proposes an alternative measure of working memory span that does correlate well with reading comprehension performance. The paper first discusses the nature of individual differences in working memory and presents the span test that was used to assess working memory capacity. Then it discusses how working memory capacity might influence two specific components of reading comprehension, retrieving facts and computing pronominal references. The first experiment shows that these two components correlate with the span measure. A second experiment shows that similar correlations are obtained when the span is assessed with a listening task.

While short-term memory traditionally has been conceived of as a passive storage buffer, the term working memory developed as a way to refer to a more active part of the human processing system (Newell, 1973). Working memory is assumed to have processing as well as storage functions; it serves as the site for executing processes and for storing the products of these processes (Baddeley & Hitch, 1974; LaBerge & Samuels, 1974). For instance, in reading comprehension, the reader must store pragmatic, semantic, and syntactic information from the preceding text and use it in disambiguating, parsing, and integrating the subsequent text. Information can become part of working memory through several routes; it may be perceptually encoded from the text; it may be sufficiently activated so that it is retrieved from long-term memory; finally, it may be the output of a comprehension process (Carpenter & Daneman, Note 1). Information can also be lost from working memory, since its capacity is assumed to be limited (Miller, 1956; Simon, 1974). Information may be lost through decay or displacement. Decay occurs if the activation of information subsides to a subthreshold level with time

(Collins & Loftus, 1975; Hitch, 1978; Reitman, 1974). Displacement occurs if additional structures are encoded, activated, or constructed until the capacity is exceeded. An important aspect of information loss is the assumption that processes and structures compete for a shared limited capacity (Baddeley & Hitch, 1974; Case, 1978; Lesgold & Perfetti, 1978). Consequently, a task that has heavy processing requirements should decrease the amount of additional information that can be maintained. One way this could occur is if the execution of more demanding processes required more attention and hence consumed a larger proportion of the capacity otherwise available for storage. Another way is if the processes in the demanding task generated intermediate products that displaced the additional information.

The trade-off between processing and storage seems like a potential source of individual differences in reading comprehension (cf. also Perfetti & Lesgold, 1977). The better reader might have more efficient processes so that he/she effectively would have more capacity for storing and maintaining information. "More efficient processes" could have several interpretations. One minimal hypothesis is that the processes of good and poor readers differ only in some quantitative way. For example, a good reader may require fewer processes than a poor reader to perform exactly the same computation; in good readers, the intermediate steps might be eliminated in some or all of the stages such as decoding, lexical accessing, parsing, inferencing, and integrating. Such efficiency would imply that the good reader would have fewer computational demands on working memory; hence, he would have more capacity for storing the necessary intermediate and final products of the reading process. More efficient processes would also be functionally faster and indeed, better readers are faster at reading-related tasks (Hunt, Frost, & Lunneborg, 1973; Hunt, Lunneborg, & Lewis, 1975; Jackson & McClelland, 1979;

Perfetti & Lesgold, 1977). A speed advantage could interact with the decay of information from working memory since less of the preceding information would decay simply because of the passage of time. Hence, the more efficient processes of the good reader could be functionally equivalent to a larger storage capacity.

Contrary to this theory, the evidence so far suggests that working memory capacity may not differentiate good and poor readers. Studies using the standard digit span test (Guyer & Friedman, 1975; Hunt et al., 1973) or a probe digit span test (Perfetti & Goldman, 1976) have found no systematic differences between good and poor readers who were classified on the basis of a general reading comprehension test. Studies using letter strings (Farnham-Diggory & Gregg, 1975; Rizzo, 1939) or similar sounding words (Valtin, 1973) as predictors of reading comprehension have been only slightly more successful. One explanation for the lack of correlation may be that digit span and word span tests do not sufficiently tax the processing component of working memory. The word span test, for instance, involves relatively simple processes such as rehearsal and access of common lexical items. A task with heavier processing demands might be needed to obtain a more marked trade-off between processing and storage. If the processing interfered with storage, the poor reader's less efficient processes would appear as equivalent to a smaller storage capacity.

One purpose of the present study was to devise a measure that taxed both the processing and storage functions of working memory. The processing and storage components of the test, which will be called the reading span test, involved the usual demands of sentence comprehension. An additional storage component required subjects to maintain and retrieve the final words of sentences. The format of the reading span test was somewhat similar to that of the traditional digit span and word span tests. The subject was given a set of

sentences to read; at the end of the set, he attempted to recall the final word of each sentence. The number of sentences in a set was incremented from trial to trial and the subject's reading span was the maximum number of sentences he could read while maintaining perfect recall of the final words. If good readers use less processing capacity in comprehending the sentences, they should be able to produce more sentence final words than poor readers.

The span task was constructed so that its demands were compatible with the characteristics of working memory. First, the presentation times were short. The subject saw each sentence only for as long as it took to read it at a normal pace—approximately 5 seconds. As soon as one sentence was finished, the next was presented so that subjects were not able to overtly rehearse the words. Second, the amount of information that the reader had to retain roughly matched the capacity attributed to working memory. For example, while reading the third sentence in a set, a reader would have to have sufficient processing and storage capacity for that sentence while retaining some representation that would allow him to retrieve the last words of the prior two sentences. This amount of information is consonant with reading models that assume working memory can accommodate several propositions from prior sentences while processing a subsequent sentence. Thus, the subject's reading span was taken as an index of his working memory capacity.

Relating working memory capacity to comprehension requires not only a measure of working memory span, but also an appropriate measure of comprehension. Traditional assessment techniques have relied on standardized reading comprehension tests, such as the Verbal Scholastic Aptitude Test (SAT). (Scores from SATs were included in the present research also.) However, such global measures of comprehension are deficient from a theoretical point of view. Because such scores reflect a variety of subtasks, they are difficult to re-late to any particular process. For that reason, the current research included two tests that tapped more specific components of comprehension: one required the reader to store and retrieve facts and the other required the reader to compute pronominal reference.

In the current experiment, the ability to encode, store, and retrieve facts was assessed by asking the reader questions about simple facts in a short passage that had just been read. This kind of question answering is a frequent component of reading comprehension tests (Carroll, 1972; Davis, 1944, 1968). One reason is that this task taps a skill that has both educational and practical importance. A second reason for such a test is that retrieving facts is a component of other comprehension processes. For example, to make an inference to relate some current information to a fact that was mentioned earlier in the text, the reader must retrieve the earlier information, as well as compute the relation. Thus, a test that requires simple fact retrieval after reading a short paragraph makes demands on memory that may be comparable to the demands that are made during reading itself.

Both the initial encoding of facts and their subsequent retrieval involve working memory and could differentiate good and poor readers. Working memory capacity could influence both the duration that a fact remains in working memory and the probability that it is consolidated in long-term memory. In both cases, the better reader would have an advantage. A fact might persist longer in working memory for the better reader because his processing does not consume all of the available capacity. The fact will not be displaced as quickly. The good reader also might have an advantage in consolidating the fact in long-term memory. A larger processing capacity might allow more opportunities for integrating a particular fact into the general representation. The fact would be available during more of the subsequent processing so that

later information could be related to it. Consequently, the integration process also would provide more retrieval routes for later accessing the fact. Finally, a reader with more efficient processes might have additional capacity to devote to rehearsal and consolidation, while the poorer reader would require all his processing capacity to perform the minimal computations. In summary, fact retrieval is one aspect of reading comprehension that could reflect differences in processing capacity.

A second component of comprehension that was measured required readers to compute pronominal references. This process may be related to fact retrieval, but it has special properties that seem closely linked to working memory capacity. When a writer uses a pronoun he is assuming that the referential concept is currently active in the reader's working memory or "foregrounded" (Chafe, 1972). Chafe compared the foregrounding of a concept to an actor who is introduced on stage during a play. To remain foregrounded, the concept must take some part in the action. If it does not, it has some probability of retreating to the wings, that is, of fading from the reader's working memory. Once the foregrounding of the concept is attenuated, the writer can no longer use the pronoun but must replace it with its referent noun. Chafe suggested that foregrounding might be attenuated after two sentences, although he admitted that this boundary is arbitrary and could be increased to an unspecified limit. The preceding analysis of working memory capacity suggests that the boundary might vary for individual readers. The distance over which a pronoun's referent could be computed might be partially dependent on the reader's working memory capacity. In particular, a larger capacity might mean a higher probability that the referent noun is still active. Hence, good readers might find such computations easier than poor readers. If the referent is no longer in working memory, the reader would have to institute a search of long-term memory. As in fact

retrieval, the search might be more successful for better readers because they would have been more likely to have initially stored the original referent.

In summary, the central issue is whether the reading span test, as a measure of working memory capacity, correlates well with reading comprehension performance. Reading comprehension was assessed through a global score, the Verbal SAT, and two more specific comprehension tests. The first required readers to answer questions about facts given in narrative passages. The second required readers to answer questions that required pronominal reference; the task was one of identifying a pronoun's referent in passages that had increasing numbers of sentences intervening between the pronoun and its referent noun.

EXPERIMENT 1

Method

The subjects were given four tests: (1) a reading span test to measure the span of working memory, (2) a reading comprehension test that asked questions about facts and pronominal references, and (3) a traditional word span test.

Reading span test. Subjects had to read a series of sentences aloud at their own pace and recall the last word of each sentence. The test was constructed with 60 unrelated sentences, 13 to 16 words in length. Each sentence ended in a different word. Two examples are: *When at last his eyes opened, there was no gleam of triumph, no shade of anger. The taxi turned up Michigan Avenue where they had a clear view of the lake.* Each sentence was typed on a single line across the center of an 8 × 5-in. index card. The cards were arranged in three sets each of two, three, four, five, and six sentences. Blank cards were inserted to mark the beginning and end of each set.

The experimenter showed one card at a time to the subject. The subject was required to read the sentence aloud. As soon as the sentence was read, a second card

was placed on top of the first and the subject read the new sentence. The procedure was repeated until a blank card signaled that a trial had ended and that he was to recall the last word of each of the sentences in the order in which they had occurred. Subjects were given several practice items at the two sentence level before the test began. They were warned to expect the number of sentences per set to increase during the course of the test. The span test contained three sets each of two, three, four, five, and six sentences. Subjects were presented increasingly longer sets of sentences until they failed all three sets at a particular level. Testing was terminated at that point. The level at which a subject was correct on two out of three sets was taken as a measure of the subject's reading span. No subjects correctly recalled any set of cards at a higher level than their defined spans.

Reading comprehension tests. The subjects were given a series of passages to read and then at the end of each passage they were asked two questions; the first interrogated the referent of a pronoun mentioned in the last sentence and the second probed some other fact from the passage. Subjects were instructed to read each passage silently at a comfortable pace but to be prepared to answer questions about it. The passages were typed on separate sheets of paper and presented in random order. Subjects were given a sheet of cardboard to cover the lines of text as they completed reading them to prevent them from reexamining the text before answering the questions. After this task, they were questioned about their strategies; no subjects realized that they were being asked about pronominal reference after each paragraph.

There were 12 narrative passages of approximately 140 words in length. The fact question interrogated a simple fact with no restrictions as to the fact's location in the passage. The pronoun question always interrogated the referent of a pronoun that occurred in the last sentence. The passages

were constructed so that the distance between the pronoun and its referent was systematically varied. The final sentence in each paragraph contained a pronoun, *she, her, he, him,* or *it*. The antecedent noun occurred two, three, four, five, six, or seven sentences prior to the final pronoun sentence. Varying the position of the antecedent noun prevented readers from adopting a strategy of attending to a particular sentence. Each of the six places was represented by two paragraphs. In each pair of paragraphs one referent noun was a common noun and one was a proper noun. In all cases, the referent of the pronoun was logically unambiguous. However, additional nouns of a similar class occurred in the sentences preceding the sentence containing the referent noun to make the task nontrivial. There was an attempt to make the referents relatively homogeneous in thematic importance since this factor may play a role in determining how long an item is foregrounded. Two sample passages and the fact and pronoun reference questions are presented in Table 1. In the Teenager Passage, the pronoun and its referent were separated by a distance of six sentences; in the Animal Passage, the distance was two sentences.

Word span test. Subjects were required to recall sets of individual words. While the presentation was auditory, previous research has shown that auditory word span and visual word span correlate almost perfectly (Lyon, 1977). The test was constructed with 81 one-syllable common nouns that were as semantically and phonetically unrelated as possible. The procedure was similar to that used in the reading span test. The words were grouped in three sets each of two, three, four, five, six, and seven words. The experimenter presented the word sets orally to the subject at a rate of one word per second. Subjects were required to recall all of the words of a set in the exact order of presentation. They were warned to expect the number of words per set to increase during the course

TABLE 1
SAMPLE PASSAGES AND QUESTIONS FROM THE COMPREHENSION TESTS

Teenager Passage (Distance 6)

Sitting with Richie, Archie, Walter and the rest of my gang in the Grill yesterday, I began to feel uneasy. Robbie had put a dime in the juke box. It was blaring one of the latest "Rock and Roll" favorites. I was studying, in horror, the reactions of my friends to the music. I was especially perturbed by the expression on my best friend's face. Wayne looked intense and was pounding the table furiously to the beat. Now, I like most of the things other teenage boys like. I like girls with soft blonde hair, girls with dark curly hair, in fact all girls. I like milkshakes, football games and beach parties. I like denim jeans, fancy T-shirts and sneakers. It is not that I dislike rock music but I think it is supposed to be fun and not taken too seriously. And here he was, "all shook up" and serious over the crazy music.

Questions

Pronoun	(1) Who was "all shook up" and serious over the music?
Fact	(2) Where was the gang sitting?
Fact	(3) Who put money in the juke box?[a]
Theme	(4) Provide a title for the passage that captures its theme.[a]

Animal Passage (Distance 2)

It was midnight and the jungle was very still. Suddenly the cry of a wolf pierced the air. This anguished note was followed by a flurry of activity. All the beasts of the jungle recognized that an urgent meeting had been summoned by the lion, their king. Representatives from each species made rapid preparations to get to the river clearing. This was where all such emergency assemblies were held. The elephant and tiger were the first to arrive. Next came the gorilla, panther and snake. They were followed by the owl and the crocodile. The proceedings were delayed because the leopard had not shown up yet. There was much speculation as to the reasons for the midnight alarm. Finally he arrived and the meeting could commence.

Questions

Pronoun	(1) Who finally arrived?
Fact	(2) Where were these emergency meetings held?
Fact	(3) What broke the stillness of the night?[a]
Theme	(4) Provide a title for the passage that captures its theme.[a]

[a] These questions were included only in Experiment 2.

of the test. Sets of increasing length were administered until a level was reached at which a subject failed all three sets. Phonetic confusions were very rare, but when a subject produced a word with one differing feature (such as *gate* for *date*) it was scored as correct, since it could have reflected an acoustic confusion. The level at which the subject was correct on two out of three sets was taken as a measure of the subject's word span.

The subject went through all four tests in the same order in a session lasting approximately 40 minutes. They were then asked their Verbal SAT score as a global measure of reading ability.

Subjects. The subjects were 20 Carnegie–Mellon University undergraduates who were enrolled in an introductory course in psychology. They were all native speakers of English.

Results and Discussion

As shown in Table 2, the span test was correlated with the traditional assessment of comprehension, Verbal SAT scores, $r(18) = .59$, $p < .01$. The readers' SATs ranged from 400 to 710 with a mean of 570 ($SD = 79.8$). The span test was even more closely related to performance on the two specific tests of comprehension, the fact questions and the pronominal reference questions; the correlations were $r(18) = .72$ and $.90$, respectively, and $p < .01$ for both. The reading span for the 20 readers varied from 2 to 5 with a mean of 3.15 ($SD = .93$).

TABLE 2
CORRELATIONS BETWEEN SPANS AND READING
COMPREHENSION IN EXPERIMENT 1

	Reading comprehension measures		
	Fact questions	Pronoun reference questions	Verbal SAT
Reading span test	.72*	.90*	.59*
Word span test	.37	.33	.35

* $r(18), p < .01$.

Readers with smaller spans performed much worse than readers with larger spans on both tests. For example, the five readers with span size 2 correctly answered only 8.2 out of the 12 fact questions and 5.4 out of the 12 pronoun reference questions. By contrast, the six readers with span sizes 4 and 5 answered 11 out of the 12 fact questions and 9.7 of the 12 pronoun reference questions. The mean performance overall was 9.4 for the fact questions and 7.4 for the pronoun reference questions.

The traditional word span measure in this study, as in many previous studies, was not significantly correlated with any comprehension measure, whether Verbal SAT, $r(18) = .35$, fact questions, $r(18) = .37$, or pronominal reference, $r(18) = .33$, as shown in Table 2. Moreover, the correlations between word span and the two specific comprehension tests were significantly lower than the correlations between reading span and these tests, $t(19) = 2.99, p < .01$, for the fact questions and $t(19) = 5.92, p < .01$, for the pronominal reference questions. Scores on the traditional word span test for the 20 subjects varied from 4 to 6; the mean was 5.15 ($SD = .67$), higher than the measure of working memory span. The two span tests were moderately correlated themselves, $r(18) = .55, p < .01$.

The results for the pronoun reference questions showed that there was a close correspondence between the reader's span and the distance over which he/she could correctly answer the question. The results for readers with different spans for small, medium, and large pronoun-referent noun

distances is shown in Figure 1. There were no significant differences in computing a common versus proper referent noun ($t(19) < 1$) so the two types of nouns were not treated separately in the analyses. An analysis of variance with four spans and six distances showed that readers with larger reading spans performed better, $F(3,16) = 27.56, p < .01$, and readers tended to do better on passages with smaller distances, $F(3,80) = 37.10, p < .01$. In addition, span interacted with the distance over which a reader could make a pronominal inference, $F(15,80) = 2.10, p < .05$. As shown in Figure 1, readers with a span of 2 or 3 had difficulty with noun−pronoun distances greater than 2 or 3. Readers with a span of 4 made numerous errors only when the distance was greater than 5 sentences. Span 5
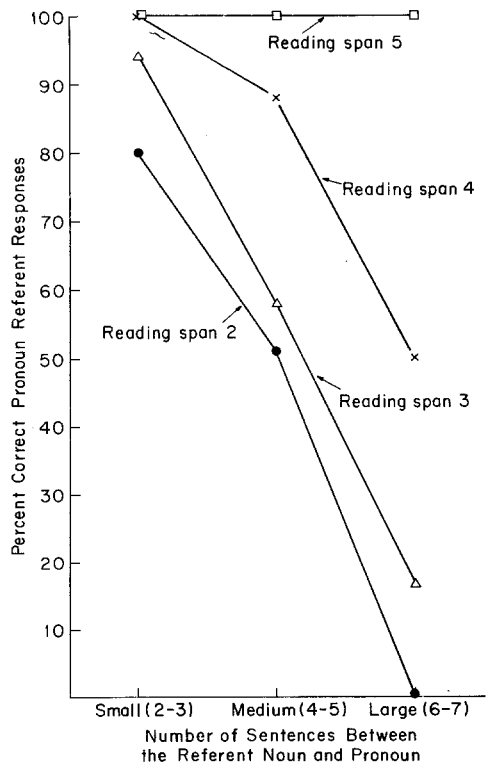

FIG. 1. The percentage of correct responses to the pronoun reference questions as a function of the distance between the pronoun and the referent noun. The curves represent the various reading spans in Experiment 1.

subjects had no errors even for distances of 6 and 7. The analysis of variance used a least-squares formula to handle the unequal n's in the various spans. Post hoc individual comparisons indicated that readers with different spans were significantly different from each other except that the span 2 and span 3 readers' performances did not differ significantly.

In summary, several aspects of the results support the hypothesis that the reading span task is related to working memory capacity. In particular, the limits of performance in the span task are consistent with a limited working memory. It is interesting to note that a span of 5 sentence final words is the best performance observed in this experiment or in the experiments that have been run subsequently (with approximately 100 subjects in all). Moreover, subjects' retrospective reports about how they performed the task suggest that they actively tried to retain the sentence final words in working memory. Subjects sometimes reported trying to rehearse the words mentally while reading the next sentence aloud. Also, subjects sometimes reported forming associations among sentence final words. Finally, it was apparent during some of the recall trials in the span test that subjects sometimes used reconstructive processes. When the subject was unable to produce the sentence final word, he/she might attempt to reconstruct the sentence on the basis of whatever "gist" had been retained. In these reconstructions, subjects sometimes made systematic errors by choosing a word that had been present in the sentence but was not the final word. These informal observations and the subjects' retrospections suggest that the reading span task was successful in taxing the processing and storage capacity of working memory. Of primary importance, of course, is the fact that the span task correlates with three measures of reading comprehension. The theory proposes that working memory capacity is the source of the correlation and hence, an important source of individual differences in reading.

EXPERIMENT 2

While the role of working memory in comprehension has been described in the context of reading, the same processes could occur in listening comprehension as well; no specifically visual processes are central to the present argument. Experiment 2 examined working memory capacity for spoken as well as written verbal material to determine the relation between reading span and listening span and their correlations with reading and listening comprehension.

To assess listening span, the test was modified so that it was suitable for both reading and listening. The modified span test consisted of sentences to be individually verified as true or false. Then, at a signal, the subject recalled the last word in each sentence. This span measure was used with silent reading, oral reading, and listening. These span measures were then compared to silent reading comprehension and listening comprehension measures. The verification component was included in the span measure to assure that subjects processed the entire sentence and did not concentrate only on the sentence final word.

The oral component of Experiment 2 also ruled out a reading speed interpretation of Experiment 1, that is, that the good readers are superior to bad readers solely because they are faster and therefore had shorter retention intervals between the reading of a critical word and recall. In the present experiment, the listening span test and listening comprehension test were presented at a constant rate and still revealed systematic individual differences.

This experiment also examined evidence for qualitative differences in the errors made by readers with different spans. The major analysis classified readers' erroneous responses to fact questions and pronominal reference questions according to how consistent they were with the gist of the passage. The issue was whether readers with large and small spans made different kinds of errors.

Method

The subjects were given three working memory span tests: (1) an oral reading span test, (2) a silent reading span test, and (3) a listening span test. These were followed by two comprehension tests: one involved silent reading and the other involved listening.

The reading span and listening span tests. Subjects had to read or listen to a sentence and answer true or false. Sentences were presented in a set and at the end of a set, the reader had to recall the last word of each sentence. The true–false component was included to ensure that subjects processed the entire sentence and did not just concentrate on the final words, a possible strategy if subjects had to only listen or silently read a sentence. While this is not a possible strategy when reading orally, the true–false component was also included in the oral reading span test to keep all three span tests comparable.

The span tests were constructed from a set of 220 different sentences. Each was 9 to 16 words in length, and ended in a different word. The sentences were taken from general knowledge quiz books and covered a wide number of knowledge domains such as the biological and physical sciences, literature, geography, history, and current affairs. Half were true and half were false. Care was taken to select statements that were of moderate difficulty, for example, (1) *You can trace the languages English and German back to the same roots.* (2) *The Supreme Court of the United States has eleven justices.*

The subjects read or listened to the sentence and then had 1.5 seconds in which to answer true or false before the next sentence was presented. In the reading versions, the end of a trial was signaled by a blank card, in the listening version, by a tone. Subjects were given several practice items at the two sentence level before the test began. If a subject had difficulty verifying the sentence in 1.5 seconds, he was urged to try to answer correctly and if he did not know the answer, to give the most

plausible answer. All the subjects mastered the procedure by the end of the practice session. Whether or not subjects verified the sentences correctly was ignored. At the end of the experiment subjects were questioned about their strategies; they all believed the true–false aspect was relevant.

The silent and oral reading span tests contained three sets each of two, three, four, five, and six sentences. As in Experiment 1, subjects were presented increasingly longer sets of sentences until they failed to recall the sentence final words of all three sets at a particular level. Testing was terminated at this point. The level at which a subject was correct on two out of three sets was taken as a measure of the subject's reading span. If the subject was correct on only one set at a particular level he was given a credit of .5. For example, if a subject was correct on two out of the three four-sentence sets he was assigned a span of 4. If he was correct on only one of the three sets he was assigned a span of 3.5. Because the test proved to be so difficult, the subject was given credit for any set for which he recalled all sentence final words, irrespective of the order of recall. In the listening span test there were five sets each of two, three, four, five, and six sentences. Subjects were defined as being at the span level at which they were correct on three out of the five sets. If they were correct on two out of the five sets they were given a credit of .5.

The comprehension tests. The materials and procedure were similar to those in Experiment 1. The only change was that subjects were asked four rather than two questions at the completion of each passage. The first question required the identification of the pronominal referent. The second and third questions interrogated other facts from the passage. The fourth question asked for a title for the passage.

There were 24 passages which were randomly divided into two groups of 12 each, with the constraint that two out of the four passages at each pronoun–noun distance level was assigned to a group. Half the

TABLE 3
CORRELATIONS BETWEEN SPANS AND COMPREHENSION IN EXPERIMENT 2

	Fact questions	Pronoun reference questions	Verbal SAT
		Reading comprehension measures	
Oral reading span test	.81**	.84**	.55**
Silent reading span test	.74**	.86**	.49**
Listening span test	.67**	.72**	.53**
		Listening comprehension measures	
Oral reading span test	.42*	.78**	
Silent reading span test	.43*	.71**	
Listening span test	.47*	.85**	

* $r(19), p < .05$.
** $r(19), p < .01$.

subjects read the first group of passages and listened to the second group; the other subjects read the second group and listened to the first. The listening comprehension paragraphs and questions were tape recorded by a male speaker at the speed of normal speech.

All subjects were given the three span tests followed by the two comprehension tests in a session lasting approximately 1 hour. To control for order effects, half the subjects were given the reading versions of the span and comprehension tests before the listening ones, while the remaining subjects received the tests in the opposite order. Finally, the subjects' Verbal SAT scores were requested as an additional measure of reading ability.

Subjects. The subjects were 21 Carnegie–Mellon University undergraduates who were enrolled in an introductory psychology course. They were all native speakers of English.

Results and Discussion

The results of Experiment 2 corroborate the findings of Experiment 1 by demonstrating high correlations between the measures of working memory span and the measures of reading comprehension. As shown in Table 3, readers with large oral reading spans were better at fact questions and pronoun reference questions, $r(19) =$ .81 and .84, $p < .01$, respectively. These correlations are similar to the .72 and .90 found in Experiment 1 with an oral reading span task that did not require verification. Similarly, readers with large silent reading spans were better at fact questions and pronoun reference questions, $r(19) = .74$ and .86, $p < .01$, respectively. Moreover, both spans show significant correlations with Verbal SAT scores, $r(19) = .55, p < .01$, $r(19) = .49, p < .05$, compared to $r(18) =$ .59 in Experiment 1. The subjects' SATs ranged from 410 to 760 with a mean of 600 $(SD = 75)$.

Listening span measures are almost as good at predicting reading comprehension as reading span measures. Subjects with larger listening spans were better at answering questions about facts and pronominal references that had been read, $r(19) = .67$ and $.72, p < .01$, respectively. While the correlations are somewhat lower than those obtained with reading spans, they are still impressive. Finally, the listening span measure correlated as well with the Verbal SATs as did the other spans, $r(19) = .53, p < .01$.

Not surprisingly, the three span measures were highly correlated; the oral reading span correlated .88 with the silent reading span and .80 with the listening span, and the silent reading span correlated .75 with the listening span, all $p < .01$. The

absolute range of the spans was also similar. The oral reading span scores of the 21 subjects varied from 1.5 to 5, with a mean of 2.76 (SD = .80). The silent span varied from 1.5 to 4, with a mean of 2.38 (SD = .70). Finally, the listening span scores varied from 2 to 4.5, with a mean of 2.95 (SD = .72).

All three span measures did reasonably well in accounting for listening comprehension, although the listening span measure was slightly better. As shown in Table 3, the correlations for listening fact questions ranged from .42 to .47, $p < .05$. These correlations are lower than those obtained for reading fact questions. This appeared to be due to three subjects who performed better than would be predicted by their span task, rather than due to some question that was systematically easier in listening than reading. The correlations obtained for listening pronoun inference questions are much more like those obtained in reading, ranging from .71 to .85, $p < .01$.

The relatively high correlations between reading and listening performance suggest that a significant amount of individual difference is common to both reading and listening. Particularly for adult readers, such as the present university population, visual encoding processes may not be the bottleneck in reading. Rather, comprehension processes that tax working memory may be responsible for difficulties that subjects have in either reading or listening. This result is somewhat similar to that of Sticht (1972) for a group of army personnel. He found that subjects who had problems reading more difficult material also had problems when it was presented orally. Both sets of results suggest that an individual differences theory that stresses visual processes, such as the time to access letter codes from visual material, may not capture a significant source of variance that is common to both reading and listening comprehension.

Performance on the pronominal reference questions is shown in Figures 2 and 3

for the reading and listening conditions. The same analyses of variance as in Experiment 1 were applied to the data. The overall results and the patterns of performance were similar for both the reading and listening conditions. As in Experiment 1, subjects with large silent reading and listening spans performed better, $F(4,16)$ = 12.14, $p < .01$, and $F(5,15) = 8.33, p < .01$, respectively. Subjects tended to do better on the written and spoken passages with shorter distance, $F(2,32) = 11.25, p < .01$, and $F(2,32) = 9.16, p < .01$, respectively. Unlike Experiment 1, the interactions between span and distance for both the reading and listening data were not significant. In both Experiments 1 and 2, the subjects with the largest working memory spans performed at asymptote. The main difference was between the slopes for the remaining spans in Experiment 1 (Figure 1) and those in Experiment 2 (Figures 2 and 3). In Experiment 1, even the small span subjects made relatively few errors on the short distance passages. However, in Experiment 2, all subjects with silent reading or listening spans less than 3.5 made errors at the shortest distances, with the percentage of errors increasing inversely with span size. In other words, in Experiment 2, subjects with large spans had an advantage over low-span subjects at retrieving referents at all distances, not just when the distance was large. Furthermore, in Experiment 1 the performance of subjects with different spans began dropping at different rates; in Experiment 2 performance dropped at a parallel and more gradual rate.

In order to examine possible qualitative differences between good and poor readers an analysis was made of their errors in the comprehension tests. The analysis was done on the four conditions: the reading fact questions, the reading pronoun reference questions, the listening fact questions, and the listening pronoun reference questions. For the reading analysis, subjects were classified into large and small silent reading spans. The cutoff point was desig-
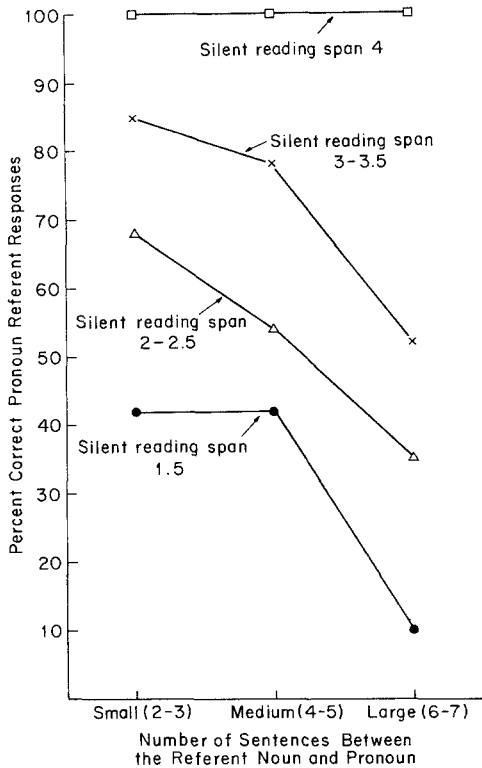
FIG. 2. The percentage of correct responses to the pronoun reference questions in the reading comprehension task as a function of the distance between the pronoun and its referent noun. The curves represent the various silent reading spans in Experiment 2.
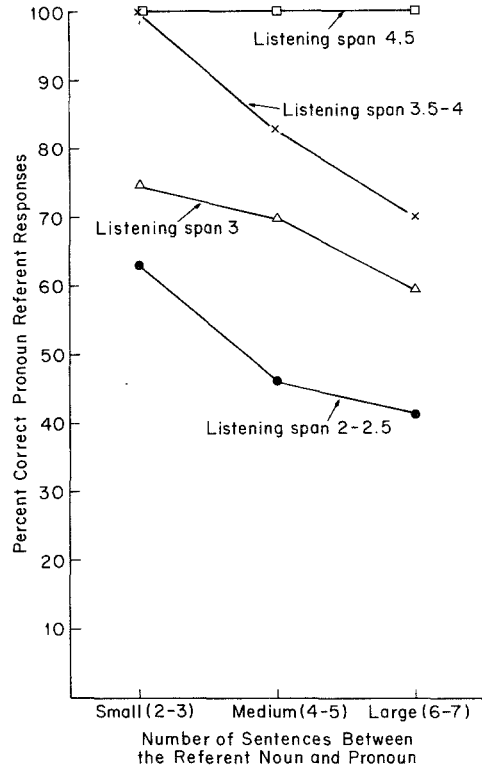
FIG. 3. The percentage of correct responses to the pronoun reference questions in the listening comprehension task as a function of the distance between the pronoun and its referent noun. The curves represent the various listening spans in Experiment 2.

nated as the span interval at which the largest decrement in performance occurred. This occurred between spans 2.5 and 3. Readers with larger spans made an average of 5.4 errors out of 24 fact questions and 2.5 out of the 12 pronoun reference questions. Subjects with smaller spans made an average of 11.3 errors out of the 24 fact questions and 6.2 of the 12 reference questions. For the listening task, the cutoff occurred between 3 and 3.5. Subjects with large listening spans made an average of 6.6 errors out of the 24 fact questions and 1.4 errors out of the 12 pronoun reference questions. Subjects with small listening spans made an average of 8.8 errors out of 24 fact questions and 5.1 out of 12 pronoun reference questions. The erroneous responses to fact and pronoun reference questions in the

reading and listening tasks were classified into several categories: reasonable substitutions, incomplete answers, foils from the passage, fabrications, confusions, and don't remembers.

The first category of errors were reasonable errors; these included reasonable substitutions and incomplete but partially correct answers. A reasonable substitution was a superset for the correct answer. An example from the Teenager Passage in Table 1 for the question, *Where was the gang sitting?*, would be the answer *the restaurant* as a substitution for *the Grill*. An incomplete answer was one in which subjects gave only part of an answer. An example from the Animal Passage in Table 1 for the question, *Where was the emergency meeting held?*, would be the answer

the river rather than *the river clearing*. The incomplete answers category was irrelevant for the pronoun reference questions since the response consisted of a single referent noun. Reasonable substitutions and incomplete answers were considered together for fact questions. The two types were collapsed because they are both very consistent with the gist of a passage and because they were distributed similarly within the small and large span groups.

Another category included foil errors and fabrications. A foil was an item that occurred in the passage and had features similar to the correct answer. An example in the Teenager Passage for the question *Who put the money in the juke box?*, would be the answer *Archie* instead of *Robbie*. Fabrications were items that did not occur in the passage although they resembled the correct answer in features such as number and gender. An example for the question about who put the money in the juke box would be the response *Tom* because *Tom* did not occur in the passage. Foils and fabrications were collapsed because they had similar distributions within small and large span groups.

The last two error types were "don't remembers" and confusions. Confusions indicated a fundamental misunderstanding. An example from the Teenager Passage for the question, *Who was "all shook up" and serious over music?*, would be the answer *the narrator;* such an error reflects that the subject failed to process the major distinction made by the passage about the narrator and his best friend.

The error analysis showed that subjects with larger spans made less serious errors. The distributions for subjects with large and small silent reading spans and large and small listening spans is shown in Table 4. For both the reading and listening fact questions, subjects with larger spans made different sorts or errors than subjects with smaller spans, $X^2(3) = 13.20$, $p < .01$ and $X^2(3) = 13.70$, $p < .01$, respectively. The subjects with larger spans made a larger percentage of the reasonable errors (reasonable substitutions and incompletes); subjects with smaller spans made a larger percentage of the types of errors that indicated more serious problems (the confusions and don't remembers). The same trend was present for the pronoun reference questions in the reading and listening conditions, $X^2(3) = 10.30$, $p < .01$, and $X^2(3) = 37.2$, $p < .01$, respectively.

Not only do good readers make fewer errors than poor readers but their errors are more likely to reflect at least some understanding of the passages. On the other hand, when poor readers or listeners fail to

TABLE 4
PERCENTAGES OF EACH TYPE OF ERROR

	Large silent reading spans	Small silent reading spans	Large listening spans	Small listening spans
	Reading fact questions		Listening fact questions	
Reasonable substitutions and incomplete answers	53	31	49	27
Foils and fabrications	19	27	12	29
Don't remember	14	17	18	20
Confusions	14	25	21	24
	Reading pronoun questions		Listening pronoun questions	
Reasonable substitutions	20	11	30	1
Foils and fabrications	45	42	42	52
Don't remember	25	21	0	18
Confusions	10	26	28	29

retrieve a fact or a pronoun's referent, their errors frequently reflect a more fundamental misunderstanding of the passage. Their errors might contradict the whole theme of the passage. This was particularly evident in the confusion errors for pronoun reference questions. A typical example in the Teenager Passage was the answer to the question *Who was "all shook up" and serious over the music?* The correct answer was *Wayne (the narrator's best friend)*; but some subjects answered *the narrator*. While none of the subjects with large spans made this error, six subjects with small spans did. Such an error suggests that the reader or listener had misunderstood the central theme, which was the narrator's scorn for the seriousness with which his friends reacted to the music.

Another indication of a qualitative difference among subjects appeared in the themes that were given for the passages. The titles that subjects provided for each passage in the reading and listening comprehension tests were scored from 0 to 2 according to how well they expressed the main idea. A score of 0 was assigned to a title that represented an unimportant detail of the passage or that denoted an incorrect interpretation of the passage; examples of titles from the Teenager Passage in Table 1 were *The Boys' Night Out* and *The Malt Shop*. A score of 1 was assigned to a title that captured the theme of a passage at a very general level or that represented an incomplete theme; an example from the same passage was the title *Teenage Boys*. A score of 2 was given to a title that was a more detailed account of the passage or that succinctly expressed the author's intentions or moral; examples were *My Objections to Rock Music* and *Taking Things in Their Proper Perspective*.

Subjects with larger spans were better at abstracting a theme from a written or spoken narrative passage. In the reading comprehension task, the mean rating for subjects with large silent reading spans was 16 out of 24; for subjects with small spans the

mean was 11.8 out of 24. In the listening task, the mean rating for subjects with large listening spans was 16.7 out of 24; for subjects with small spans the mean was 12.1 out of 24. The ability to abstract a theme from written narratives correlated significantly with silent reading span, $r(19) = .71$, $p < .01$. Similarly, the ability to abstract a theme from spoken narratives correlated significantly with listening span, $r (19) = .82$, $p < .01$. Again, abstracting themes from written and from spoken verbal material seemed to share common skills: silent reading span correlated with listening theme abstraction, $r(19) = .57, p < .05$; listening span correlated with reading theme abstraction, $r(19) = .69, p < .01$.

GENERAL DISCUSSION

The high correlations between reading span and the various comprehension measures are striking; the reading span task succeeds where previous short-term memory measures have failed. The argument has been that the span task reflects working memory capacity and that this capacity is a crucial source of individual differences in language comprehension. While more promising than previous individual differences analyses, the present studies are still correlational and more evidence will be needed to show precisely how and when working memory capacity limits the comprehension process.

One useful experimental approach might expand on the design of the pronominal reference task. To successfully compute a pronominal referent, the reader must retrieve prior information, in this case, the referent, and link it to the pronoun. All integrative computations involve relating some current information to previous information. The previous information might have been explicitly stated or it might have to be inferred. It seems reasonable that the computation will be easier if the required prior information is active in working memory and harder if it must be searched for and retrieved from long-term memory. The

present theory predicts that readers with smaller spans are less likely to have the prior information active in working memory. In the current study, this lowered probability manifested itself in higher error rates. However, it might also manifest itself in longer response times, assuming that the poorer reader attempts to search long-term memory, and in overt patterns of eye fixations including pauses and regressions. These are issues that we are currently investigating.

The present paper has stressed quantitative differences in working memory capacity as the source of individual differences. The most parsimonious assumption seems to be that readers are making the same computations, but that they differ in the speed or efficiency with which they can make those computations. However, the analysis of readers' errors also indicated some qualitative differences among readers; high- and low-span readers do not make the same kinds of errors. One theoretical explanation might be that the processes of good and poor readers are qualitatively different. Poor readers may be doing fundamentally different things than good readers and the different processes also may be less efficient. An alternative explanation, and one to be pursued here, is that qualitative differences can emerge from quantitative differences.

One example of how differences in capacity could result in qualitative differences in processing is in the chunking process. This process recodes concepts and relations into higher order units. One prerequisite for chunking is that each of the individual concepts be present simultaneously in working memory (Shiffrin & Schneider, 1977). Consequently, the more concepts there are to be organized into a single chunk, the more working memory will be implicated. Although the process of forming rich chunks would temporarily strain working memory capacity, these chunks would have a quantitative payoff. The recoding of many concepts and relations into

a single chunk would have the economizing effect of reducing the load on working memory and hence would increase the functional working capacity for subsequent processing.

The chunks of a good reader might differ qualitatively from a poor reader. The good reader has more functional working memory capacity available for the demands of chunking. He is more likely to have more concepts and relations from preceding parts of the text still active in working memory. Consequently, he should be able to detect more interrelations among these concepts, and to note their relative importance. The good reader's chunks should be richer, and more coherent, and contain different information. The presence of different interrelations could subsequently allow different inferences and generalizations to be drawn.

Good readers might capitalize on their chunking efficiency in the comprehension and working memory span tasks. In the comprehension test passages, the chunks of the good reader would correspond to the interrelations of clauses and sentences that form superordinate discourse units. Such chunks might aid readers in the storage and retrieval of facts and themes. In the span test, which consisted of unrelated sentences, the role of chunking strategies is less obvious. The chunks might be the interrelations among the sentences of a set or even the interrelations of propositions within a single test sentence. As mentioned before, there is some suggestion from the verbal reports of our subjects that both types of chunking occurred.

The present theory that working memory capacity is the crucial source of individual differences in reading was based on a circumscribed population, that is, on a group of bright university students. The mean Verbal SAT score of the group was above the national mean and presumably even the poorest readers were functioning adequately with undergraduate level reading materials. The theory does not address itself to other possible sources of individual

differences in reading, such as perceptual disabilities and motivational factors, that may be characteristic of the extremely poor or immature readers.

This research on individual differences has a parallel to the research on memory development. The argument has been that a major difference between good and poor readers is the efficiency of their processing, rather than static memory capacity. Similarly, developmental research suggests that children and adults differ in processing capacity and that there is little difference in their static capacities. If adults are prevented from using special strategies and if the material is equated for familiarity, children and adult memory spans are very similar (Case, 1978; Chi, 1976; Huttenlocher & Burke, 1976). But children have much slower and less efficient processes. The more of these processes the task requires, the more marked are the developmental differences. Thus both individual and developmental differences are most evident when the subjects differ in the efficiency with which they can execute required processes.

The converse of this argument is that if subjects are equally adept at a particular process, there should be few individual differences in a span task that requires that process. One example of such a process is simple counting. In a counting span task that paralleled the reading span task, Case, Kurland, and Daneman (Note 2) had adults and children count arrays and recall the number of items in each array in a set. The number of arrays increased until the subject could no longer recall accurately. Adults seemed to be equally good at counting. There were only small differences in the speed of counting, a measure of processing efficiency, and in the counting span. Moreover, there was no correlation between the speed and span measures. Individual differences among adults were found only when their processing efficiency was impaired by having them count in an artificial language. By contrast, children are not equally efficient in the process of simple counting. For them, there was a high correlation between their counting span and counting speed.

The asymptotic efficiency of a simple skill like counting highlights the lack of asymptotic performance in reading. If one assumes that a reading span of 5 is asymptotic, the conclusion is that most readers (even a group of bright undergraduate college students) are not performing at ceiling. Moreover, the results of Experiment 2 show that this conclusion is not peculiar to reading. There were significant individual differences in listening comprehension as well. The interesting implication is that both reading and listening comprehension are not easy for adults and that there may be potential room for improvement well into adulthood.

REFERENCES

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

CARROLL, J. B. Defining language comprehension: Some speculations. In R. O. Freedle & J. B. Carrol (eds.), *Language comprehension and the acquisition of knowledge.* Washington, D.C.: Winston, 1972.

CASE, R. Intellectual development from birth to adulthood: A Neo-Piagetian interpretation. In R. Siegler (ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

CHAFE, W. L. Discourse structure and human knowledge. In R. O. Freedle & J. B. Carroll (eds.), *Language comprehension and the acquisition of knowledge.* Washington, D.C.: Winston, 1972.

CHI, M. T. H. Short-term memory limitations in children: Capacity or processing deficits? *Memory and Cognition,* 1976, 4, 559–580.

COLLINS, A. M., & LOFTUS, E. F. A spreading-activation theory of semantic processing. *Psychological Review,* 1975, 82, 407–428.

DAVIS, F. B. Fundamental factors in reading. *Psychometrika,* 1944, 9, 185–197.

DAVIS, F. B. Research in comprehension in reading. *Reading Research Quarterly,* 1968, 3, 499–545.

FARNHAM-DIGGORY, S., & GREGG, L. W. Short-term memory function in young readers. *Journal of Experimental Child Psychology,* 1975, 19, 279–298.

GUYER, B. L., & FRIEDMAN, M. P. Hemispheric pro-

cessing and cognitive styles in learning-disabled and normal children. *Child Development*, 1975, 46, 658–668.

HITCH, G. J. The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, 1978, 10, 302–323.

HUNT, E., FROST, N., & LUNNEBORG, C. Individual differences in cognition: A new approach to intelligence. In G. H. Bower (ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press, 1973. Vol. 7.

HUNT, E., LUNNEBORG, C., & LEWIS, J. What does it mean to be high verbal? *Cognitive Psychology*, 1975, 7, 194–227.

HUTTENLOCHER, J., & BURKE, D. Why does memory span increase with age? *Journal of Verbal Learning and Verbal Behavior*, 1976, 8, 1–31.

JACKSON, M. D., & MCCLELLAND, J. L. Processing determinants of reading speed. *Journal of Experimental Psychology: General*, 1979, 108, 151–181.

JUST, M. A., & CARPENTER, P. A. A theory of reading: From eye fixations to comprehension. *Psychological Review*, in press.

KINTSCH, W., & VAN DIJK, T. A. Toward a model of text comprehension and production. *Psychological Review*, 1978, 85, 363–394.

LABERGE, D., & SAMUELS, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, 6, 293–323.

LESGOLD, A. M., & PERFETTI, C. A. Interactive processes in reading comprehension. *Discourse Processes*, 1978, 1, 323–336.

LYON, D. R. Individual differences in immediate serial recall: A matter of mnemonics? *Cognitive Psychology*, 1977, 9, 403–411.

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

NEWELL, A. Production systems: Models of control structures. In W. G. Chase (ed.), *Visual information processing*. New York: Academic Press, 1973.

PERFETTI, C. A., & GOLDMAN, S. R. Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior*, 1976, 14, 33–42.

PERFETTI, C. A., & LESGOLD, A. M. Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

REITMAN, J. S. Without surreptitious rehearsal information in short-term memory decays. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 365–377.

RIZZO, N. D. Studies in visual and auditory memory span with specific reference to reading disability. *Journal of Experimental Education*, 1939, 8, 208–244.

SHIFFRIN, R. M., & SCHNEIDER, W. Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 1977, 84, 127–190.

SIMON, H. A. How big is a chunk? *Science*, 1974, 183, 482–488.

STICHT, T. G. Learning by listening. In R. O. Freedle & J. B. Carroll (eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: Winston, 1972.

VALTIN, R. *Reports of research on dyslexia in children*. Paper presented at International Reading Association, Denver, 1973. (ERIC Document Reproduction Service No. ED 079713)

REFERENCE NOTES

1. CARPENTER, P. A., & DANEMAN, M. *Lexical retrievel and error recovery in reading: A model based on eye fixations*. Mimeograph manuscript. Carnegie–Mellon University, Pittsburgh, 1979.

2. CASE, R., KURLAND, M., & DANEMAN, M. *Operational efficiency and the growth of M-space*. Paper presented at the Society for Research in Child Development, San Francisco, March 1979.