

Working memory and comprehension: A replication

LINDA BAKER

University of Maryland Baltimore County, Catonsville, Maryland

(Marilyn E. Demorest, *sponsor*)

Daneman and Carpenter (1980, *Journal of Verbal Learning and Verbal Behavior*, **19**, 450-466) developed a measure of working memory capacity, the reading span test, which they related in several experiments to individual differences in reading comprehension skill. Because previous attempts to establish a relation between working memory and comprehension had failed, Daneman and Carpenter's work has received a great deal of attention. The purpose of the present study was to provide a replication using independently constructed materials for both the reading span test and the comprehension tasks. The results essentially replicated those of Daneman and Carpenter, but there were some important qualifications.

One hypothesized source of individual differences in reading comprehension skill is working memory capacity. Because readers must maintain several propositions in working memory simultaneously in order to establish a cohesive representation of a text, those with more limited capacities are at a disadvantage. Initial tests of this hypothesis met with little success; Daneman and Carpenter (1980) attributed these failures to the fact that traditional measures, such as digit span and word span, tap the storage component of short-term memory but not the processing component. They reasoned that because poor readers differ from good readers in a variety of processing measures related to reading, such as speed of verbal coding and lexical access, working memory capacity may be differentially taxed by the processing demands of reading. Accordingly, Daneman and Carpenter developed a measure of working memory capacity designed to tap both storage and processing. This task, called the reading span test, required subjects to read sets of sentences while simultaneously maintaining the terminal words in memory. College students differed in the number of final words they were subsequently able to recall, and these differences were systematically related to three different measures of reading comprehension. Daneman and Carpenter (1983) recently demonstrated a relation between reading span and still other measures of comprehension.

Daneman and Carpenter's (1980) paper has attracted a great deal of attention and has been widely cited as evidence that working memory capacity plays an important role in comprehension. Citing the limited value of existing measures of comprehension, Daneman (1982) recently extended her argument and proposed that the reading span test itself be used as a measure of comprehension. In view

of the interest generated by the reading span test and Daneman's (1982) somewhat unorthodox proposal, replication is warranted. The present study is a replication of Daneman and Carpenter's (1980) Experiment 1 using independently constructed materials for both the reading span test and the comprehension tasks. This report will focus on points of departure from the original study; the reader should refer to Daneman and Carpenter for details.

METHOD

Reading Span Task

The materials for the reading span task consisted of 75 sentences that ranged in length from 13 to 15 words (mean = 13.78, SD = 0.70) and that were roughly comparable in number of syllables. All of the terminal words used here contained three syllables, whereas Daneman and Carpenter did not control number of syllables. If subjects use some form of articulatory code to rehearse the terminal items, then number of syllables could be an important consideration if one wishes to make normative claims about reading span, as Daneman and Carpenter (1980) seem to do.

The arrangement of sentence sets, the procedures, and the scoring system were similar to those of Daneman and Carpenter (1980). However, in the present study, the subjects controlled sentence presentation themselves; that is, as soon as they had read one sentence aloud, they turned the card over and went on to the next sentence. This change from presentation by experimenter was made for ease of administration. Because Daneman and Carpenter did not specify that there were any serial order constraints on subjects' recall, none were imposed here.

Comprehension Task

Materials consisted of 12 12-sentence narrative passages (mean passage length = 147 words, SD = 11.93). Two questions that tested for fact memory were generated for each passage; a third question tested for anaphora resolution. The anaphoric referent, which was always presented in the last sentence, was separated from its anaphor by 2, 4, or 6 sentences. (Daneman & Carpenter, 1980, varied this distance from 2 to 7 sentences but then analyzed the distances as small, medium, and large.) The task was administered as in Daneman and Carpenter, but with one modification: After the subjects had read and answered questions for all 12 passages, they were re-presented with the passages containing questions they had answered incorrectly and were asked to find the answers to these questions in the passages. The purpose of this ad-

The author's mailing address is: Department of Psychology, University of Maryland Baltimore County, Catonsville, MD 21228.

ditional task was to discover whether the subjects would show comprehension differences when memory for the material was not required.

Subjects

The subjects were 28 students enrolled in an introductory psychology course who participated for extra credit. All subjects consented to their SAT scores being obtained from the Admissions Office. There was a greater range in SAT-Verbal scores and a lower mean than in Daneman and Carpenter (1980) (ranges = 270-640 vs. 400-710; means = 441 vs. 570; SDs = 105 vs. 79.8).

RESULTS AND DISCUSSION

The mean reading span was 4.13, with a range of 2 to 7 (SD = 1.56). Since the maximum set size presented to the subjects was 7 sentences, we cannot rule out the possibility that some subjects would not have exceeded a reading span of 7. Daneman and Carpenter (1980) reported a mean span of 3.15 and a range of 2 to 5 (SD = 0.93). (The 1983 study reported a maximum span of 5.5.) There are two factors that may have contributed to the obtained differences in maximum reading spans. The subjects in the present study controlled sentence presentation themselves. Perhaps the self-pacing permitted the subjects more of an opportunity for rehearsal, even though they seemed to comply with the instruction to turn the sentence cards immediately after having read them aloud. It is also possible that Daneman and Carpenter did indeed stipulate that subjects should not recall the last word in the set first, but they did not state this in their article. (The 1983 study did make such a stipulation explicit.) The subjects in the present study typically did recall the last word first, so this procedural difference may have been a second factor. The fact that all of the to-be-recalled words had three syllables, rather than one or two, did not seem to matter; the effect of this change should have been to decrease reading spans, not to increase them. Daneman (1982) seemed to attribute special significance to the specific range of reading spans observed: "most university students can only recall 2-3 final words, with a very small percentage recalling 4 or 5" (p. 342). However, it is clear from the present study that span size should be interpreted in relative, rather than absolute, terms.

Table 1 presents the intercorrelations among reading span, SAT-Verbal scores, number of pronoun questions answered correctly, and number of fact questions answered correctly. All of these correlations were reliable at the .05 level or better and thus replicated Daneman and Carpenter (1980). However, whereas Daneman and Carpenter found reading span to be more strongly related to fact questions and pronoun questions than to SAT scores, the opposite was true here. Daneman and Carpenter did not report the correlations between SAT scores and the comprehension measures, but the present study indicated that SAT scores were more strongly related to performance on pronoun questions than was reading span. This is a troublesome finding for Daneman's (1982) argument that reading span is a better indicator of comprehension than standardized comprehension tests.

Table 1
Intercorrelations Among the Dependent Measures

| | Reading Span | SAT-Verbal | Pronoun Questions | Fact Questions |
|-------------------|--------------|------------|-------------------|----------------|
| Reading Span | | .67* | .46* | .53* |
| SAT-Verbal | | | .65* | .55* |
| Pronoun Questions | | | | .64* |
| Fact Questions | | | | |

**p* < .05.

Also examined was the relation between reading span and performance on the pronoun questions as a function of anaphor-referent distance. Daneman and Carpenter (1980) found that subjects with larger reading spans were able to answer questions with greater distances than were subjects with smaller spans and that the difference between individuals was smaller for short distances. The present study revealed a similar pattern: The correlation between span and number of questions answered with a distance of 2 was only .13, whereas the correlations between span and number of questions answered with distances of 4 and 6 were .49 and .39, respectively, reliable at the .05 level. The discrepant aspect of the present data is that the subjects were no more accurate on the questions with distances of 4 than on those with distances of 6. Daneman and Carpenter suggested that there is a linear relation between reading span and the distance over which anaphoric relations can be established, but it appears that their finding may be materials specific.

The study also included the additional task of answering previously missed questions in the presence of the passages. If subjects with shorter reading spans continue to answer fewer questions correctly than do subjects with longer reading spans, this would implicate basic comprehension deficiencies. However, if the subjects now perform comparably, this would implicate memory as the original source of the difference. Overall, the subjects correctly answered a high proportion (.87) of the missed questions, with no consistent differences related to reading span. This suggests that comprehension problems apparent in a test requiring memory may be eliminated when students are given an "open-book" test. In other words, the relation between working memory capacity and comprehension seems to depend on the type of comprehension task given.

In conclusion, the present study essentially replicated Daneman and Carpenter (1980): Reading span was related to performance on two types of comprehension questions and to performance on the SAT-Verbal, suggesting that working memory capacity is indeed related to comprehension. Thus, even with completely new sets of materials and some procedural differences, the same basic pattern was obtained. Nevertheless, it is clear that reading span should not be interpreted in absolute terms as Daneman and Carpenter seem wont to do; it is a relative measure that varies with task specifics. Additionally, it is not clear that reading span is a better predictor of reading comprehension than the SAT-Verbal. It is premature to accept

Daneman's (1982) argument that reading span is in fact a better measure than traditional standardized test scores. It remains to be demonstrated that the correlations reflect other than the fact that the measures tap some similar processes.

REFERENCES

DANEMAN, M. (1982). The measurement of reading comprehension:

How not to trade construct validity for predictive power. *Intelligence*, **6**, 331-345.

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**, 561-584.

(Manuscript received for publication October 5, 1984.)