

Metamemory for narrative text

RUTH H. MAKI and SHARON SWETT
North Dakota State University, Fargo, North Dakota

In this experiment, we investigated metamemory for narrative text passages. Subjects read two stories and made memory predictions for the idea units in one and rated the importance of ideas in the other. Half of the subjects were asked to recall the story immediately after reading the passages and half were asked to recall 1 week later; half received passages with single inconsistent idea units and half received passages with corresponding consistent idea units. All subjects made confidence judgments about the accuracy of their recall. Subjects' prediction ratings were related to recall, as shown by significant prediction accuracy quotients. Importance ratings were related to recall on the delayed test but not on the immediate test. Memory prediction ratings predicted recall better than did importance ratings. The absolute level of memory predictions did not differ with delay, but subjects did give higher confidence judgments on an immediate than on a delayed test. Subjects recalled the inconsistent idea better than the consistent idea for one story but not for the other. For both stories, subjects predicted that they would remember the inconsistent ideas better, suggesting that they have a von Restorff-type view, rather than a schema view, of memory. We conclude that subjects can predict their memory for the idea units in narrative text.

The present study was designed to investigate questions about metamemory for text. The results reported in the literature are mixed with respect to subjects' abilities to assess how well they are comprehending and to predict how well they will perform on a memory test involving the text material. Maki and Berry (1984) had subjects predict the likelihood that they would answer test questions correctly for sections from an introductory psychology textbook. They found that subjects who performed above the median on a multiple-choice test made predictions that were more accurate than would be expected by chance. Subjects who scored below the median on the multiple-choice test, however, did not accurately predict to which sections they would give correct and incorrect test answers. Thus, Maki and Berry concluded that better students could predict test performance but that poorer students could not. Maki and Berry also found that prediction accuracy was better on an immediate test than on a delayed test, particularly for the poorer students. In contrast, Glenberg and Epstein (1985) found that subjects at all ability levels, after reading passages of prose, could not predict performance on an inference test any better than expected by chance. They found, however, that subjects' ability to predict inference test performance improved if they had had experience with an inference test item for a specific passage. The difference between the paradigm of Maki and Berry and that of Glenberg and Epstein might be that subjects knew more about what type of multiple-choice questions to expect in Maki and Berry's

study, whereas subjects in Glenberg and Epstein's study were unfamiliar with the specific types of questions that might be asked in the inference test. If subjects know what will appear on a test, they may be able to make predictions about memory for parts of a text as they read it.

In contrast to the results with memory predictions about text, several studies have shown that subjects are fairly good at predicting their memory for unrelated words or sentences. Both Lovelace (1984) and Vesonder and Voss (1985) found a strong relationship between subjects' predictions of memory and their actual memory for both words and sentences. Studies using text differ from those using lists in at least two ways: first, text consists of material that is much more integrated, and second, text studies have required subjects to make predictions for larger units than have list studies. In both Lovelace's and Vesonder and Voss's studies, predictions were made for specific word pairs or sentences. The subject thus had a good idea of exactly what would be tested when the prediction was made. In Maki and Berry's (1984) and Glenberg and Epstein's (1985) studies, predictions were made about fairly long sections of text. Subjects did not know exactly what would be tested. In addition, comprehension difficulty may have been a factor in the text studies, but probably was not a factor in the word list and sentence studies.

In the present experiment, subjects made predictions for single idea units in a passage that was easy to understand. Each prediction was specific to an idea unit. In addition, subjects knew that they would be asked to recall the entire text, so they did not need to guess which portion of the text would actually be tested. The texts used were simple narrative stories, so memory predictions should be based on the memorability of the ideas, not on comprehensibility. These factors should increase the prob-

This report was prepared while the first author was on sabbatical at the Claremont Graduate School. We express our appreciation to Kathy Pezdek for her helpful comments on the paper. Requests for reprints should be sent to Ruth H. Maki, Department of Psychology, North Dakota State University, University Station-Box 5075, Fargo, ND 58105.

ability of finding a significant relationship between subjects' predictions and their recall.

Of additional interest in the present experiment was whether subjects' predictions about recall would differ according to whether the ideas were more or less consistent with the rest of the story. Schema theory (Alba & Hasher, 1983; Taylor & Crocker, 1981) predicts that facts that are consistent with the schema should be processed better, in terms of the schema, and, therefore, should be recalled better. One type of research that supports this prediction is research involving the relevance of statements or events to the schema. Generally, ideas that are rated as important to the structure of a story are remembered better than ideas that are rated to be less important (e.g., Johnson, 1970; Kintsch & van Dijk, 1978). Other research that supports the schema view of text recall involves reconstructions in recall following the presentation of information that is inconsistent with prior expectations. Spiro (1980) presented a balanced story, in which a couple agreed to have or not to have children, and an imbalanced story, in which the couple disagreed. Later, subjects were incidentally informed that the couple married or did not marry. Spiro found that subjects who heard contradictory information about marriage following an imbalanced story distorted the couple's views on having children to make them more consistent with the outcome. However, these accommodative errors occurred only when subjects were told that the experimenter was interested in their reactions to the story, rather than their memory for it, and only after a 3-week or longer delay.

In contrast to the research that supports a schema view of text recall, other research has shown that information that is irrelevant to or inconsistent with a schema is sometimes remembered better than information that is consistent. Interruptions to a script, which is a type of schema for events, are remembered better than actions that are expected in the script (Bower, Black, & Turner, 1979, Experiment 7); material that is irrelevant to a script is remembered better than material that is highly relevant to the script (e.g., Graesser, Woll, Kowalski, & Smith, 1980; Smith & Graesser, 1981); and behaviors that are inconsistent with an expectancy about a person are remembered better than behaviors that are consistent with the expectancy (Hastie, 1980; Srull, 1981; Srull, Lichtenstein, & Rothbart, 1985). Thus, material that stands out is often remembered better than material that is expected. This finding may be similar to the von Restorff effect (cited by Koffka, 1935), in which an item belonging to a different class from the other items in a list is better remembered than items from the same class.

In addition to answering questions about the accuracy of subjects' memory predictions for narrative text, the present study also allowed us to determine whether subjects predict better or poorer memory for material that is moderately inconsistent with the rest of the story. Half of the subjects read a story that was internally consistent, and half read a story that contained an inconsistent idea unit. If subjects have an inherent schema theory of

memory, they might predict that they would be more likely to remember a consistent idea that fits in with the rest of the schema than an inconsistent idea that fits less well. If, however, they have a theory of memory similar to that of von Restorff (cited by Koffka, 1935), they might predict that something that stands out or that is somewhat surprising might be better remembered.

In addition to making predictions about single idea units, subjects in the present study also made two other types of metamemory judgments. Because importance has been shown to be related to memory, we had subjects make importance ratings for a second story. Not only did this allow us to determine whether the consistent and inconsistent ideas were judged to be equally important, it also allowed us to compare the accuracy of memory ratings and importance ratings in predicting recall. At the time of the test, subjects gave confidence judgments for their recall. A number of investigators have asked subjects to rate their confidence in test answers (e.g., Koriati, Lichtenstein, & Fischhoff, 1980; Shaughnessy, 1979), but such judgments have not been compared to memory predictions in the same study. We investigated whether memory predictions and confidence judgments are related and whether the two types of judgments are affected by the same variables.

The independent variables in the experiment were consistency, as described above, and delay. Some subjects knew they would be asked to recall during the initial reading session and some subjects knew that they would be asked to recall a week later. Thus, the effect of expected delay on memory predictions and the effect of an experienced delay on confidence judgments could be determined. Confidence judgments for each recalled sentence were given for both stories.

METHOD

Design

The design was a $2 \times 2 \times 2$ mixed design with delay (10 min vs. 1 week) and consistency (consistent vs. inconsistent) as between-subjects variables and type of rating (memory or importance) as a within-subjects variable. One of two stories was randomly assigned to each type of rating, so half of the subjects read one story for the memory ratings and the other story for the importance ratings and half had the opposite pairing.

Materials

Two narrative stories were written. Each had two versions, one that included a fact that was somewhat inconsistent with the rest of the story and one that included a similar but consistent fact. Each story contained four paragraphs. The Michael/Nancy story was patterned after the story used by Spiro (1980), but the names were changed and it was simplified so that idea units could be presented individually. The Susan story was written for use in this experiment. Each story was divided into idea units by mutual agreement of the two authors. The Michael/Nancy story contained 43 idea units and the Susan story contained 42 idea units. There were two versions of each story, which differed in terms of one idea unit. In the consistent version, that idea unit was compatible with the remainder of the story; in the inconsistent version, that idea unit was somewhat incompatible with the remainder of the story, although

the inconsistency was plausible. In the Michael/Nancy story, the consistent idea was that a couple who met, considered marriage, and got married both wanted to have children. The inconsistent idea was that the man wished to have children, but the woman did not. In the Susan story, a young woman who had just graduated from college with a degree in English was searching for a job related to journalism. In the consistent version, she did the best in writing courses in college. In the inconsistent version, she did the worst in writing courses in college. These stories are printed in the Appendix, with each idea unit on a separate line. The consistent and inconsistent idea units are printed in italics.

Procedure

Each subject made memory prediction ratings for the first story and importance ratings for the second story, with the specific story counterbalanced across presentation positions. Memory predictions were always made before importance ratings so that subjects would not use importance as a way of making the memory predictions. The story was presented on a monitor attached to an Apple II+ computer. Each sentence was presented individually in upper- and lowercase letters at the top of the screen. The sentence was written when the screen was blanked off and then the entire sentence appeared when the screen was turned on, so the subject did not see the sentence created word by word. A timing program began when the sentence appeared. The subject was instructed to read each sentence and to press the space bar on the computer when he/she had understood it. This ended the timing program, and study time was measured. After the subject had read the entire story in this manner, each complete sentence was presented at the top of the screen again. In the middle of the screen the subject saw HOW LIKELY ARE YOU TO REMEMBER with an idea unit from the sentence printed in appropriate upper- and lowercase letters beneath it. Below that was a 7-point rating scale ranging from *Very unlikely* (1) to *Very likely* (7). The subject was instructed to judge how likely he/she was to remember that idea after the appropriate time interval, that is, at the end of the session that day or 1 week later. After the subject had entered a rating, that idea unit was erased and the next one was presented for a rating. After each idea unit in a sentence had been rated, the next sentence was printed at the top of the screen and the subject rated each idea unit in that sentence.

Immediately after the subject had finished reading and rating the first story, instructions for the second story were presented. Again, the subject read the entire story, sentence by sentence, and study time was measured for each sentence. Then each sentence was presented at the top of the screen, and HOW IMPORTANT IS and an idea unit were presented in the middle of the screen. Below this was a 7-point scale ranging from *Easily eliminated* (1) to *Very important* (7). The subject was instructed to decide how important each idea was by deciding how easily it could be eliminated. The subject was told to press 1 if the idea could be easily eliminated without changing the story, and to press 7 if the idea was critical to the story. After all of the idea units in a sentence had been rated, a new sentence was presented and its idea units were rated for importance.

After both stories had been read, subjects in the delayed condition were dismissed and reminded to return 1 week later. Subjects in the immediate condition were given a paper and pencil number progression task to work on for 10 min. After the 10-min or 1-week intervals, all subjects were asked to recall the first story they had read. They were given a minimal retrieval cue for the appropriate story. The cue for the Michael/Nancy story was that it was about a couple who just met. The cue for the Susan story was that it was about a woman looking for a job. The subjects were asked to write down as many ideas as they could from the story, but were told that they did not need to try to write the story exactly as it had been presented. They were asked to start each new sentence on a new line. They were told to take as long as they needed and to tell the

experimenter when they were finished. When the subjects indicated that they were finished, the experimenter asked them to look back over what they had written and decide how sure they were that they had written the ideas in each sentence correctly, that is, that the ideas that they had written had really appeared in the story. The subjects were shown a rating scale on a card with 1 = *Not at all sure* and 7 = *Very sure*. They were asked to write a number beside each sentence they had written to indicate how sure they were that the ideas in that sentence really had occurred in that way in the story. After the first story was complete, the second story was recalled and rated in exactly the same way.

Subjects

A total of 64 volunteer subjects from introductory psychology classes at North Dakota State University were randomly assigned to the immediate and delayed, consistent and inconsistent conditions. There were 15 subjects in each of the two immediate conditions, 18 subjects in the delayed consistent condition, and 16 subjects in the delayed inconsistent condition. Three subjects in the delayed condition did not return on the appropriate day after 1 week, but they were tested 24, 48, or 72 h later.

RESULTS

A number of different types of results are presented. First, the interrater reliabilities for the recall data are presented to show that the scoring procedure was fairly reliable. Next, the recall scores are presented to give the reader some idea of the level of recall that subjects were predicting or making confidence judgments about. Recall of the sentences that were common to the two consistency groups are presented first, followed by recall of the sentence that differed in the consistent and inconsistent conditions. Next, various aspects of the memory prediction ratings are described. First, the average memory prediction ratings given to sentences that were common to the two consistency groups are discussed in order to determine whether the absolute level of prediction ratings differed as a function of delay. Second, the accuracy of the memory predictions is described and we examine the relationship between the predictions and recall. Next, the relationship between the memory predictions and study time is examined. In the final analysis of memory predictions, predictions of recall of the consistent and inconsistent sentences are compared. This same set of analyses is then described for the importance ratings, followed by a discussion of the relationships between the memory predictions and importance ratings. Finally, the confidence judgments given to recalled ideas at the time of the test are presented and their relationship to recall and to the memory predictions and importance ratings are described in order to examine the relationship among these measures of metamemory.

Recall

Interrater reliabilities. Two raters independently scored each recall protocol. The raters evaluated gist rather than verbatim recall. Each idea unit was given a recall score of 0 if it was not recalled, 1 if it was partially recalled, and 2 if it was correctly recalled. This resulted

in two sets of 42 (for the Susan story) and two sets of 43 (for the Michael/Nancy story) recall scores for each subject. The interrater reliabilities for each subject were determined by calculating Cohen's (1960) Kappa value. This measure involves dividing the percentage of agreement minus chance by 1 minus chance, giving the proportion of agreement above chance. There was one Kappa value for each subject for the Susan story and one for the Michael/Nancy story. Each of the individual Kappa values was significant at $p < .001$, showing good agreement between the two raters. The overall average of these Kappa values was .77. The Kappa values were entered into an unweighted-means analysis of variance (ANOVA) with delay and consistency as between-subjects variables and story as a within-subjects variable. The Kappas were equivalent for the immediate and delay conditions [.786 vs. .758; $F(1,59) = 1.63$, $MSe = .015$], and they were equivalent for the consistent and inconsistent conditions (.768 vs. .774; $F < 1$). Mean Kappas were somewhat higher for the Michael/Nancy story than for the Susan story [.790 vs. .752; $F(1,59) = 4.01$, $MSe = .011$]. (The significance level used throughout this report, unless noted otherwise, is $p < .05$.) Because interrater reliabilities were less than perfect, the scores of the two raters were averaged for each idea unit and these average scores were used for the recall analyses reported below.

Overall recall. Mean recall of the sentences that were common in the consistent and inconsistent conditions was analyzed first. The mean rater-assigned recall scores for each idea unit were averaged for each story. Whether the subject had made memory predictions or importance ratings during reading was not considered in this analysis; thus, story was a within-subjects variable, because each subject had read both stories. Mean recall scores are shown in the left half of Table 1. These were analyzed in a $2 \times 2 \times 2$ mixed-design unweighted-means ANOVA with delay and consistency as between-subjects variables and story as a within-subjects variable. Although the sentences used in this analysis did not differ between the consistent and inconsistent conditions, it is possible that the presence of an inconsistent idea might influence recall of the rest of the story, so the consistency variable was included in this analysis. Recall was better for the immediate than for the delayed group [$F(1,60) = 88.29$, $MSe = .128$], and recall was better for the Michael/Nancy story than for the Susan story [$F(1,60) = 19.79$, $MSe = .023$]. No other differences were significant.

Consistent and inconsistent idea recall. Recall for sentences containing the consistent and inconsistent idea units was compared. Each critical sentence contained two idea units, resulting in two recall scores for each sentence. Mean recall scores for the two idea units contained in the consistent and inconsistent sentences for each story are presented in the right half of Table 1. These were analyzed in a $2 \times 2 \times 2 \times 2$ mixed-design unweighted-means ANOVA, with delay and consistency as between-subjects variables and story and idea unit (first vs. second half of the sentence) as within-subjects variables. As can be seen in Table 1, recall was better in the immediate than in the delayed condition [$F(1,60) = 18.54$, $MSe = 1.747$], and recall was better in the inconsistent condition than in the consistent condition [$F(1,60) = 4.73$, $MSe = 1.747$]. However, consistency and story interacted [$F(1,60) = 4.99$, $MSe = .901$]. Inconsistent idea units were better recalled than were consistent units in the Susan story [$F(1,60) = 10.33$, $MSe = .827$], but not in the Michael/Nancy story ($F < 1$). Recall of the inconsistent idea units was better than recall of similar consistent idea units in the Susan story, but the same trend was not significant in the Michael/Nancy story.

Memory Predictions

Each subject first made memory prediction ratings for either the Michael/Nancy or the Susan story. These ratings were made on a 7-point scale (1 = *Very unlikely to remember*, 7 = *Very likely to remember*). First we performed analyses that included all of the sentences that were common to the consistent and inconsistent groups, followed by separate analyses of the consistent versus inconsistent sentences.

Overall memory predictions. The ratings given to the idea units contained in sentences that were common to the consistent and inconsistent conditions were averaged across idea units. The mean ratings did not differ for the consistent and inconsistent groups (4.400 vs. 4.373; $F < 1$), showing that the presence of an inconsistent idea did not affect memory predictions for the rest of the story. Overall, ratings in the immediate and delayed recall groups (4.324 vs. 4.442) did not differ ($F < 1$), but the delay \times story interaction was significant [$F(1,56) = 5.83$, $MSe = .490$]. For the Susan story, the immediate and delayed groups did not differ in their mean ratings [4.398 vs. 4.093; $F(1,27) = 1.832$, $MSe = .400$], but for the Michael/Nancy story, the ratings for the immediate group actually tended to be lower than the ratings for the delayed

Table 1
Immediate and Delayed Mean Recall (0-2 Scale) of Idea Units in Common Sentences
and in Sentences that Varied Between the Consistent and Inconsistent Groups

	Common Sentences				Consistent-Inconsistent Sentences			
	Michael/Nancy Story		Susan Story		Michael/Nancy Story		Susan Story	
	Immediate	Delayed	Immediate	Delayed	Immediate	Delayed	Immediate	Delayed
Consistent	1.275	.592	1.045	.538	.933	.361	.867	.208
Inconsistent	1.093	.536	1.037	.399	1.233	.250	1.483	.849

group [4.259 vs. 4.792; $F(1,29) = 4.18$, $MSe = .573$, $p < .06$]. The results for both stories show that subjects' memory predictions were not higher in the immediate than in the delayed condition, although recall for both stories was considerably higher when tested immediately.

Prediction accuracy quotients. The purpose of the next set of analyses was to investigate how accurately subjects predicted their recall. There is some disagreement in the literature about how the accuracy of memory predictions should be measured. Nelson (1984) argued for the use of nonparametric correlational measures, whereas other investigators have treated the rating data as interval in nature and have used parametric measures. The ratings in the present study were treated as interval in nature in the ANOVAs reported above; therefore, we will continue to treat them as interval data. Parametric prediction accuracy quotients (PAQs), which were used by King, Zechmeister, and Shaughnessy (1980), and more recently by Lovelace (1984), will be reported. However, we also computed Spearman rank-order correlations, and their pattern was identical to that of the PAQs. We conducted these analyses only for the idea units that were common to the consistent and inconsistent conditions, so that we could examine the overall relationship between ratings and recall without including two consistent and inconsistent idea units.

The PAQs were calculated by subtracting the mean rating for nonrecalled items from the mean rating for recalled items. Recall was scored on a 3-point scale (0, 1, or 2) in the present study, but for the PAQ, nonrecalled items were those given a score of 0 by both raters and recalled items were considered to be items having average scores above 0. Overall, the proportion correct using this criterion was .53. The correct-incorrect rating difference was then divided by the square root of the pooled variance for the correct and incorrect ratings in order to take the range of ratings into account for each subject. A PAQ of 0 indicates that subjects were unable to accurately predict recall, and a positive PAQ indicates that correctly recalled items were rated higher than incorrectly recalled items, that is, that predictions showed some accuracy. Because these PAQ values should be normally distributed around 0 if the null hypothesis is true, we used parametric statistical tests on them.

The mean PAQ values for the memory predictions in the two delay conditions and for the two stories can be seen in Table 2. These were analyzed in a $2 \times 2 \times 2$ between-subjects unweighted-means ANOVA with delay, consistency, and story as between-subjects variables.

There were no significant effects [$F_s(1,56) \leq 1.54$, $MSe = .172$], showing that subjects predicted their recall of the two stories equally well and also that they predicted their immediate and delayed recall equally well. However, the mean PAQ of .342 was significantly higher than 0, as shown by a single-sample t test [$t(63) = 6.83$].

Maki and Berry (1984) showed that subjects who performed poorly also made less accurate memory predictions, although Glenberg and Epstein (1985) did not find such a relationship with inference tests on text and Lovelace (1984) failed to find a similar relationship with word pairs or sentences. In order to look for a such a relationship in the present data, we calculated Pearson correlation coefficients (r_s) between recall scores and the PAQs for subjects in the immediate and delayed conditions. The r between the memory rating PAQs and recall was $-.30$ in the immediate condition and $-.10$ in the delayed condition. Neither value was significantly different from 0 ($p_s > .10$), but the correlations were in the opposite direction from those in Maki and Berry's (1984) study; here, subjects with higher recall tended to have lower PAQs.

Relationship between study time and memory predictions. Subjects may have based their memory predictions on how rapidly they read sentences. If so, then reading speed should be correlated with memory predictions. The time that each subject spent reading each sentence when it was originally presented was converted to reading speed per syllable by dividing reading time for each sentence by the number of syllables it contained. Because only one reading speed per sentence was available, the ratings given to the idea units in a sentence were averaged so that there was one mean rating per sentence. We computed r for each subject, using milliseconds per syllable and the mean rating for each sentence in the story read in the memory prediction phase of the experiment. For the Michael/Nancy story, r was computed across 23 sentences; for the Susan story, r was computed across 17 sentences. The mean r across subjects was .024, which was not significantly different from 0 ($t < 1$). In addition, an ANOVA showed that the correlations did not differ as a function of delay, consistency, or story ($F_s < 1$). Thus, there was no evidence that memory prediction ratings were based on reading speed.

Predictions for consistent and inconsistent idea units. Ratings of consistent and inconsistent ideas were analyzed in a $2 \times 2 \times 2 \times 2$ mixed-design unweighted-means ANOVA with delay, consistency, and story as between-subjects variables and idea unit within the critical sentence as a within-subjects variable. Mean prediction ratings for

Table 2
Mean Prediction Accuracy Quotients for Memory and Importance Ratings

	Immediate			Delayed		
	Michael/Nancy	Susan	Mean	Michael/Nancy	Susan	Mean
Memory	.267	.419	.338*	.399	.293	.346*
Importance	.023	.165	.099	.266	.348	.307*

*Significantly different from 0 with single-sample t test ($p < .01$).

Table 3
Memory Prediction Ratings for First and Second Idea Units in Consistent and Inconsistent Sentences by the Immediate and Delayed Groups

	Michael/Nancy Story				Susan Story			
	Immediate		Delayed		Immediate		Delayed	
	Idea 1	Idea 2*	Idea 1	Idea 2*	Idea 1*	Idea 2	Idea 1*	Idea 2
Consistent	4.625	4.375	4.778	4.667	3.857	5.143	3.889	5.000
Inconsistent	5.125	5.375	5.375	5.500	4.429	5.286	5.500	5.625

*Differed between consistent and inconsistent conditions.

the two idea units in the consistent and inconsistent sentences are presented in Table 3 as a function of story and delay. Mean ratings did not differ between the delayed and immediate groups or between the Michael/Nancy story and the Susan story ($F_s < 1$). Higher ratings of the critical sentence by the inconsistent group than by the consistent group (5.290 vs. 4.545) were marginally significant [$F(1,56) = 3.40$, $MSe = 5.04$, $p = .07$], as was the interaction among consistency, story, and idea unit [$F(1,56) = 3.85$, $MSe = .597$, $p = .054$]. As can be seen in Table 3, the difference between the consistent and inconsistent ratings was greatest for the second idea unit (the one that differed) in the Michael/Nancy story but for the first idea unit (the one that differed) in the Susan story. Only these inconsistent idea units were analyzed in a subsequent unweighted-means ANOVA, with delay, consistency, and story as between-subjects variables. The only significant effect was higher ratings for the inconsistent idea than for the consistent idea [5.226 vs. 4.212; $F(1,56) = 5.56$, $MSe = 2.88$]. Thus, subjects predicted that they would remember an idea that was inconsistent with the rest of the story better than an idea that was consistent. In contrast to the recall data, predictions did not differ for the two stories.

Prediction accuracy for consistent and inconsistent ideas. Because there were only two idea units in the consistent and inconsistent sentences, PAQs could not be calculated for the sentence that differed in the two consistency conditions. Therefore, we assessed accuracy of the memory prediction ratings by correlating the mean recall score of the idea units in the consistent or inconsistent sentence with the mean prediction rating for that sentence. Therefore, we computed Pearson r s across subjects, essentially to determine whether subjects who gave higher memory predictions for the critical sentence also had higher recall scores. The resulting r s in the various conditions are shown in the left half of Table 4. The rela-

tionship between memory prediction ratings and recall was significant only in the immediate, consistent condition.¹ It is possible that subjects did not make accurate predictions in the immediate, inconsistent condition because they tended to give high prediction ratings for the inconsistent sentence whether or not they later recalled it. In the delayed condition, the correlations were not significant in either the consistent or inconsistent conditions. The results of these between-subjects correlations based on one sentence per subject are somewhat at variance with the PAQ results, which showed significant relationships between recall and memory predictions on both the immediate and delayed tests. A possible reason for these different patterns is that the PAQ is a within-subjects measure and the Pearson r , as computed across subjects here, is a between-subjects measure, a point that will be discussed later.

Importance Ratings

After the first story had been read and memory predictions had been made, each subject read the second story and made importance ratings for each idea unit. These ratings were also made on a 7-point scale with 1 = *Easily eliminated* and 7 = *Very important*. The ratings for idea units that were common to the consistent and inconsistent conditions were analyzed first, followed by the ratings given to the consistent and inconsistent sentences.

Overall importance ratings. Mean importance ratings for idea units that were the same in the consistent and inconsistent conditions were analyzed in a 2 (delay) \times 2 (consistency) \times 2 (story) between-subjects unweighted-means ANOVA. The overall rating (4.26) did not differ as a function of delay, consistency, or story ($F_s \leq 1.23$, $MSe = .356$). Thus, the presence of an inconsistent sentence in the story did not affect subjects' perception of the importance of the rest of the story, and the expected retention interval had no effect on perceived importance.

Prediction accuracy quotients. PAQs have not been used for importance ratings before, but the logic is the same as for the memory prediction ratings, and use of a similar measure allows us to compare these two types of ratings. As with the memory predictions, the mean importance rating given to nonrecalled idea units was subtracted from the mean rating given to recalled idea units, and this difference was divided by the pooled variance in the importance ratings. The PAQs associated with the importance ratings are shown in the bottom of Table 2. These were analyzed in a 2 \times 2 \times 2 between-subjects

Table 4
Pearson r Coefficients Relating Critical Sentence Recall with Memory Predictions and Importance Ratings for the Immediate and Delayed, Consistent and Inconsistent Groups

	Memory Predictions and Recall		Importance Ratings and Recall	
	Immediate	Delayed	Immediate	Delayed
Consistent	.59*	-.12	.20	.11
Inconsistent	-.21	.23	.29	-.12

*Significantly different from 0 ($p < .05$).

design with delay, consistency, and story as factors. There was a marginally significant trend for larger PAQs in the delayed than in the immediate condition [$F(1,56) = 3.70$, $MSe = .191$, $p < .06$]. Single-sample t tests showed that the mean PAQ in the immediate condition (.099) did not differ from 0 [$t(29) = 1.22$], but the mean in the delayed condition (.307) was significantly higher than 0 [$t(33) = 4.11$]. Thus, it appears that subjects' importance ratings predicted their recall better after 1 week than after 10 min.

Relationship between importance ratings and reading speed. Pearson correlation coefficients were calculated for each subject to determine whether reading speed per syllable and the mean importance rating given to each sentence were related. The mean r across subjects was .077, which, although small, was significantly different from 0 [$t(63) = 3.35$]. Generally, subjects read sentences that they rated as more important more slowly than they read sentences that they rated as less important. However, on the average, this relationship was not very strong.

Relationship between importance and memory prediction ratings. Because each subject made both memory prediction ratings and importance ratings, these can be compared. These ratings were averaged across story, because about half of the subjects made memory predictions for the Michael/Nancy story and importance ratings for the Susan story and half made importance ratings for the Michael/Nancy story and memory predictions for the Susan story. The PAQs were analyzed in a $2 \times 2 \times 2$ mixed-design unweighted-means ANOVA with delay and consistency as between-subjects variables and type of rating (memory vs. importance) as a within-subjects variable. The results showed that the PAQs associated with memory ratings were higher than the PAQs associated with importance ratings [.342 vs. .210; $F(1,60) = 4.45$, $MSe = .139$]. Although the previous analysis suggested that delay and type of ratings interacted, the delay \times type of rating interaction was not significant [$F(1,60) = 2.25$, $MSe = .139$].

To determine whether subjects who predicted their recall accurately with memory predictions also predicted their recall with their importance ratings, we computed Pearson correlation coefficients for subjects' memory and importance PAQ scores. The r in the immediate condition was .14 ($p > .10$) and the r in the delayed condition was .35 ($p < .05$). Therefore, in the delayed condition, subjects whose memory prediction ratings predicted their recall scores also made importance ratings that predicted their recall scores. This relationship did not hold in the immediate condition, probably because importance ratings were generally not related to recall in that condition.

Although memory predictions and importance ratings differed in how well they predicted recall, these two types of ratings must be somewhat related; it stands to reason that memory would be predicted to be higher for important ideas. Each subject did not make both types of ratings on the same story, so correlations cannot be computed for individual subjects. However, the group's mean

memory prediction and mean importance rating for each idea unit can be determined, and these can be correlated. These mean ratings for each idea unit are shown in the Appendix. Pearson r correlations were computed between the group ratings for memory predictions and for importance ratings. The correlation between the two types of ratings was .74 ($p < .01$) for the Michael/Nancy story and .61 ($p < .01$) for the Susan story. These significant correlations suggest that ideas that were rated as important were ideas that were also rated as likely to be recalled, although the relationship was far from perfect.

Importance ratings for consistent and inconsistent idea units. Subjects' importance ratings for the two idea units in the consistent and inconsistent sentences were analyzed in a $2 \times 2 \times 2 \times 2$ mixed-design unweighted-means ANOVA, with delay, consistency, and story as between-subjects variables and idea unit (first vs. second half of the critical sentence) as a within-subjects variable. The only significant effect for the critical sentence was that the idea units in the Michael/Nancy story were rated as more important than the idea units in the Susan story [5.565 vs. 4.727; $F(1,56) = 5.39$, $MSe = 4.35$]. Consistent sentences were rated about as important as inconsistent sentences (5.015 vs. 5.258; $F < 1$), and subjects expecting an immediate test gave importance ratings similar to those given by subjects expecting a delayed test (5.083 vs. 5.176; $F < 1$).

Correlation of importance and consistent or inconsistent idea recall. The next analysis was conducted to determine whether subjects who had higher recall scores for the consistent or inconsistent ideas also gave these ideas higher importance ratings than subjects who had lower recall scores. We computed Pearson r s for the importance ratings and recall scores for the four consistency and delay groups. As can be seen in the right half of Table 4, none of these correlations was significantly different from 0. Thus, this analysis, based on only one sentence per subject, suggested that there was no relationship between importance ratings and recall. This is in contrast to the results of the overall analysis with PAQ scores, in which a significant relationship between importance and recall was found in the delayed condition.

Confidence Judgments for Recalled Ideas

Following recall, subjects were asked to look over what they had written and to decide how sure they were that the ideas they recalled had actually appeared in the original story. The judgments varied from 1 (*Not at all sure that the idea was part of the original story*) to 7 (*Very sure that the idea was part of the story*). Each idea unit within a sentence was assigned the rating that the subject had given to the entire sentence. Confidence judgments could be given only for recalled ideas, so judgments were not available for the consistent and inconsistent ideas for many subjects. Therefore, only overall analyses that involved all recalled idea units are presented.

Overall confidence judgments for recalled ideas. The mean confidence rating for recalled idea units was deter-

mined for each subject. Although this mean was based on different numbers of idea units for different subjects, all subjects recalled some ideas, and, therefore, a mean confidence judgment could be determined. Overall, confidence was higher for the immediate than for the delayed group [6.241 vs. 5.773; $F(1,60) = 6.21$, $MSe = 1.177$] and higher for the Michael/Nancy story than for the Susan story [6.205 vs. 5.780; $F(1,60) = 22.10$, $MSe = .250$]. There was also an interaction between delay and story, because the difference between the immediate and the delayed conditions was greater for the Susan story (6.126 vs. 5.474) than for the Michael/Nancy story (6.356 vs. 6.071) [$F(1,60) = 4.42$, $MSe = .250$]. Although subjects were not sensitive to the expected delay in their memory prediction ratings, they were sensitive to the delay after having experienced it at the time of the test. In addition, the story that was better recalled was given higher confidence judgments.

Relationship between recall and confidence judgments. Because recalled ideas were scored by each rater as 2 if they were correct and 1 if they were partially correct, and because scores were averaged across raters, a number of different recall scores for each idea unit were possible. These scores for recalled ideas were correlated with the corresponding confidence judgments for each subject, using the Pearson r . One subject was dropped from this analysis because only four ideas were recalled, producing an unstable correlation. The correlations were then entered in a 2 (delay) \times 2 (consistency) \times 2 (story) mixed-design ANOVA with story as a within-subjects variable, since each subject recalled each story. The type of ratings that had been given during study were not considered in this analysis. Overall, the average Pearson r between confidence ratings and recall scores was .34, but the ANOVA showed that this differed for the two stories ($r = .45$ for the Michael/Nancy story and .22 for the Susan story). Both of these values differed from 0 [$t(62) = 17.19$ and 6.22, respectively]. Thus, subjects gave accurate confidence judgments, in that higher confidence judgments were related to higher recall scores for both stories.

Relationship between confidence judgments and ratings. For each subject, a correlation coefficient was computed between the confidence judgment and the memory prediction for each recalled idea. The average Pearson r for the confidence judgments and memory predictions was .12, which is significantly different from 0 [$t(62) = 3.03$]. The size of the correlation did not vary as a function of delay, consistency, or story [$F(1,56) \leq 2.79$, $MSe = .081$]. Generally, recalled ideas that received high confidence ratings had also received high memory predictions during study; however, this relationship was fairly weak. Similar correlations were also computed between confidence judgments and importance ratings for each subject. The mean r was $-.016$, which was not significantly different from 0 ($t < 1$). Recalled important ideas were not given higher confidence judgments than less important recalled ideas.

We conducted another correlational analysis to determine whether subjects who had higher memory prediction PAQs also had higher Pearson r s for their confidence judgments and recall. Thus, the question was whether subjects who made accurate memory predictions also made accurate confidence judgments at the time of the test. Overall, the Pearson r between these two measures of accuracy was .036 ($p > .10$), indicating that there was no relationship. Similarly, there was no relationship between subjects' importance-rating PAQs and the confidence-judgment/recall r s ($r = -.11$, $p > .10$). Although subjects' confidence judgments were related to their memory prediction ratings, the level of accuracy of the two types of metamemory judgments was not related. This may have been because only recalled ideas could be used in the confidence judgment analysis, or because subjects who are good at making memory predictions are not necessarily good at making confidence judgments. Further research is necessary to distinguish between these possibilities.

GENERAL DISCUSSION

Generally, subjects were able to predict their recall with greater than chance accuracy. The memory predictions were significantly related to recall on both immediate and delayed tests. Although the average value of the PAQs for memory predictions is not high (.342), it is significantly above zero and is not much less than the PAQs observed by Lovelace (1984) following a single study trial with word pairs (PAQ = .40) or sentences (PAQ = .49). Thus, subjects can predict their memory for idea units in narrative text when they know that they will receive a recall test.

Importance ratings were not as predictive of recall as were memory prediction ratings. Importance ratings were particularly poor predictors of recall on an immediate test. This result contrasts with that of Johnson (1970), who found that subject-rated importance predicted recall quite effectively and that this predictability was not related to retention interval. Although the present results appear to be at odds with those of Johnson, they are consistent with results reported by Kintsch and van Dijk (1978), who found that the ratio of recalled macropropositions (important ideas) to recalled micropropositions (less important, detail information) increased with delay. Thus, importance became more predictive of recall with longer retention intervals. The different effects of delay in the various studies may have been due to the use of different methods to define importance, different measures to relate recall to importance, or both. Alternatively, the role of importance in recall might depend upon the difficulty of the text material. Schmidt (1983) found a negative relationship between subjectively rated importance and recall for difficult technical prose, probably because the most important ideas were abstract generalizations and the less important ideas were concrete examples. Meyer (1985) also reported that objectively determined importance does not produce a systematic effect on recall with some types

of text, especially text containing specific, concrete details. Importance may also play a lesser role in the immediate recall of very easy narrative text, such as that used in the present experiment.

Confidence judgments were significantly related to recall, even though they were given only for ideas that were recalled. The correlation analysis showed that subjects gave higher confidence judgments to ideas that were given higher recall scores by the raters. The relationship between confidence judgments and recall was not affected by delay; that is, subjects' confidence judgments were as strongly related to their recall on the delayed as on the immediate test.

The delay variable influenced the relationship between recall and importance ratings, but did not influence the relationship between recall and memory predictions or the relationship between confidence judgments and recall. As expected, delay influenced the level of recall: subjects recalled less after 1 week than after 10 min. However, delay did not affect subjects' absolute level of memory predictions or importance ratings. This makes sense for importance ratings; material should be considered to be of the same importance whether it is to be recalled immediately or after a week. However, it would be expected that subjects would give lower memory prediction ratings if they know they will be asked to recall 1 week later than if they know they will be asked to recall in the same session. Because delay was a between-subjects manipulation, it may have been that subjects were using the rating scale differently in the two delay conditions. Such insensitivity to delay interval in a within-subjects design would be more interesting, but still the possibility that subjects do not take delay into account in memory predictions is intriguing. After subjects actually experienced the 1-week delay, however, the overall level of their confidence judgments did decline, relative to those of subjects who recalled in the same session. Thus, the level of confidence judgments was affected by delay in this between-subjects design. Schacter (1983) also found that subjects' confidence ratings were lower after a delay. His subjects were less likely to predict that they would recognize a non-recalled word after a 1-week delay than immediately. However, as in the present study, the accuracy of his subjects' confidence judgments was not related to delay.

This experiment also allowed an examination of the relationship between memory prediction ratings and confidence judgments, at least for recalled ideas. The average correlation was significant, showing that ideas that received higher memory prediction ratings also received higher confidence judgments at the time of the test. The correlation was probably depressed because only recalled ideas could be used in this analysis. If it had been possible to include forgotten ideas (which would have been possible with a recognition test), the relationship should have been stronger, because those items were at the low end of the memory prediction rating scale. In addition, a correlational analysis showed that the accuracy of memory predictions and the accuracy of confidence judgments

were not related. Thus, the ability to predict one's memory may not be related to the ability to decide if one has answered a question correctly. Importance ratings and confidence judgments were also unrelated, indicating that these two measures of metamemory are probably based on different psychological processes.

Reading speed, as measured by milliseconds per syllable in a sentence, was not related to memory predictions. In contrast, there was some relationship between reading speed and importance, although the absolute level of the average correlation across subjects was quite low. This finding is similar to that of Cirilo and Foss (1980), who found that subjects took longer to read more important information than they did to read less important information.

In addition to answering questions about metamemory for narrative text, this experiment showed that subjects predict that they are more likely to recall an idea that is somewhat inconsistent with the text than to recall ideas that are more consistent. Although recall of the inconsistent idea was significantly better only in the Susan story, subjects' memory predictions were higher for the inconsistent than for the consistent idea in both stories. This effect was particularly evident for the specific idea unit that differed between the consistent and inconsistent conditions. The differences between the consistent and inconsistent ideas in recall and memory predictions were not due to differences in importance, because subjects rated the importance of the consistent and inconsistent ideas the same in the two stories. Subjects seem to hold a view of memory similar to that of von Restorff (cited in Koffka, 1935), rather than a schema view. That is, they predicted that they would remember the fact that stood out better than the fact that fit in with the rest of the story.

There was also some evidence that subjects' memory predictions were more accurate for a consistent than for an inconsistent idea, because subjects generally gave high ratings to the inconsistent idea whether or not they remembered it. However, a significant relationship between memory predictions and recall was observed only on the immediate and not on the delayed test. This is in contrast to the analysis of the PAQ measure, which was significantly different from zero whether recall was immediate or delayed. These different patterns may have resulted from the different methods used to arrive at the recall-prediction relationship. Because there was only one sentence per subject when the consistent and inconsistent ideas were considered, subjects' prediction accuracy was assessed in a between-subjects design; that is, the analysis was designed to determine whether subjects who recalled any part of the sentence gave higher predictions than subjects who did not recall it. The PAQ measure was a within-subjects measure; that is, it showed whether an individual gave higher predictions to ideas that were later recalled than to ideas that were not recalled. Subjects may be able to discriminate what they know from what they do not know within their own memories, but measures that compare one subject's rating with that of another sub-

ject may show that subjects who recall do not give predictions that differ from subjects who do not recall. Such a difference in design may also explain why studies of the relationship between eyewitness identification accuracy and confidence show that correlations are very low (e.g., Deffenbacher, 1980), but studies of the relationship between subjects' test answers and their confidence (e.g., Koriat et al., 1980; Shaughnessy, 1979) show fairly strong relationships. The eyewitness studies are typically based on one observation per subject and correlations must be between subjects, whereas the confidence judgment studies are based on many observations per subject and relationships are determined within subjects. Between-subjects comparisons, such as that performed in the present study with the consistent and inconsistent ideas, may be less likely to show accurate prediction or confidence judgment accuracy than within-subjects comparisons, such as that performed in the overall analysis with PAQ scores.

Our general conclusion from the present study is that subjects can assess their levels of comprehension well enough to make accurate predictions of their free recall performance and accurate confidence judgments about their recall. This conclusion is in contrast to that of Glenberg and Epstein (1985), who concluded that subjects are very poor at assessing their levels of comprehension of text. Glenberg and Epstein cited two possible reasons for their subjects' very poor performance. First, subjects may not have been able to develop the text structure that was intended by the author; that is, they may not have been able to comprehend the text. Glenberg and Epstein argued that this was unlikely with their materials, which were fairly straightforward narratives. The second, and more likely, source of difficulty that they noted was in the subjects' assessment of the text representation. To make memory predictions, subjects must assess the integrity of the text representation that they have created from reading and they must judge how well the material in that representation matches what will be needed on the test. The ease of making that assessment may be critical in determining whether or not subjects' memory predictions will be accurate. In Glenberg and Epstein's experiment, the test was an inference test, and subjects of all ability levels may have been unable to make an accurate assessment of the usefulness of their text representations in answering inference questions. Maki and Berry (1984) used a more familiar type of test, a multiple-choice test, and subjects of higher ability were able to make accurate memory predictions, possibly because they could predict test questions and assess how well their text representations would serve in answering those questions. Subjects of lower ability did not make accurate predictions, possibly because they could not predict test items very accurately. In the present experiment, subjects of all ability levels made accurate memory predictions, perhaps because they knew the exact nature of the test. If this anal-

ysis is correct, then conclusions about whether or not subjects can predict their memory for text must depend upon the method of assessing memory. Straightforward, predictable tests may lead to the conclusion that subjects can predict their memories; less predictable tests may lead to the conclusion that they cannot.

REFERENCES

- ALBA, J. W., & HASHER, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203-231.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.
- CIRILO, R. K., & FOSS, D. J. (1980). Text structure and reading time for sentences. *Journal of Verbal Learning & Verbal Behavior*, *19*, 96-109.
- COHEN, J. A. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement*, *20*, 37-46.
- DEFFENBACHER, K. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law & Human Behavior*, *4*, 243-260.
- GLENBERG, A. M., & EPSTEIN, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 702-718.
- GRAESSER, A. C., WOLL, S. B., KOWALSKI, D. J., & SMITH, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 503-515.
- HASTIE, R. (1980). Memory for information which confirms or contradicts a general impression. In R. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, D. L. Hamilton, & D. E. Carlston (Eds.), *Person memory: The cognitive basis of social perception*. Hillsdale, NJ: Erlbaum.
- JOHNSON, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning & Verbal Behavior*, *9*, 12-20.
- KING, J. F., ZECHMEISTER, E. B., & SHAUGHNESSY, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, *93*, 329-343.
- KINTSCH, W., & VAN DIJK, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.
- KOFFKA, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace & World.
- KORIAT, A., LICHTENSTEIN, S., & FISCHHOFF, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory*, *6*, 107-118.
- LOVELACE, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 756-766.
- MAKI, R. H., & BERRY, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 663-679.
- MEYER, B. J. F. (1985). Prose analysis: Purposes, procedures, and problems. In B. K. Britton & J. B. Black (Eds.), *Understanding expository text*. Hillsdale, NJ: Erlbaum.
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109-133.
- SCHACTER, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 39-54.
- SCHMIDT, S. R. (1983). The effects of recall and recognition test expectancies on the retention of prose. *Memory & Cognition*, *11*, 172-180.
- SHAUGHNESSY, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, *13*, 505-514.
- SMITH, D. A., & GRAESSER, A. C. (1981). Memory for actions in scripted

- activities as a function of typicality, retention interval, and retrieval task. *Memory & Cognition*, 9, 550-559.
- SPIRO, R. J. (1980). Accommodative reconstruction in prose recall. *Journal of Verbal Learning & Verbal Behavior*, 19, 84-95.
- SRULL, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning & Memory*, 7, 440-463.
- SRULL, T. K., LICHTENSTEIN, M., & ROTHBART, M. (1985). Associative storage and retrieval processes in person memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 11, 316-345.
- TAYLOR, S. E., & CROCKER, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1). Hillsdale, NJ: Erlbaum.
- VESONDER, G. T., & VOSS, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory & Language*, 24, 363-376.

NOTE

1. We also analyzed these data in another way: we divided subjects according to whether or not they recalled any portion of the critical sentence and conducted a 2 (delay) \times 2 (consistency) \times 2 (recalled vs. not recalled) between-subjects ANOVA. This analysis also indicated that the difference in memory predictions between subjects who did and who did not recall the sentence was significant only in the immediate, consistent condition. Thus, the conclusion is the same as that drawn from the correlational analysis.

APPENDIX

Stories, Broken into Idea Units, with Mean Group Memory Predictions and Importance Ratings

Idea Unit	Mean Memory Rating	Mean Importance Rating
Michael/Nancy Story		
On a warm August day	3.95	3.00
shortly before the beginning of Fall Quarter,	4.61	4.07
Michael and Nancy met.	6.41	6.79
She was carrying boxes from her car	4.56	3.48
to the building where she had rented an apartment.	4.63	4.75
Michael was in the hallway	4.33	4.51
checking his mailbox.	4.02	3.08
At first, they did not see each other	3.10	3.34
and they nearly collided.	5.04	4.74
After a brief moment of laughter and apologies,	3.84	3.19
they introduced themselves,	4.82	5.74
commenting that they must be neighbors.	3.94	4.62
Michael offered to help Nancy	5.11	4.64
move the remainder of her things.	3.76	3.04
Afterwards they sat outside	3.84	4.34
drinking iced tea and talking.	3.64	2.94
In the months that followed,	3.52	3.59
Michael and Nancy spent much of their free time together.	5.96	6.39
Since they lived in the same building,	5.48	3.93
hardly a day went by when they did not see each other.	5.40	5.90
By Winter Quarter, they were very close.	5.46	5.75
They went on several ski trips	4.83	3.77
over long weekends that winter,	2.90	2.67
and they decided to spend Spring break together.	5.32	5.11
Over Spring break,	4.03	3.60
Nancy and Michael began to consider marriage.	6.42	6.75
After graduation,	3.83	4.32
they planned to visit relatives	3.83	4.32
before setting a definite date for a wedding.	3.98	4.88
Michael and Nancy had many common interests.	5.06	4.81
Both were musicians.	5.15	4.02
Michael played the flute	4.15	2.69
and Nancy played the guitar.	4.17	2.63
They also enjoyed the same outdoor activities	4.52	4.85
of canoeing, hiking, and camping.	3.85	2.61
Michael was interested in raising a family,	4.98	5.53
and Nancy wanted to have children also. (C)	4.02	5.50
but Nancy did not care to have children. (I)	5.44	5.73
Both desired to travel abroad.	4.07	4.54
They especially wanted to visit Australia.	4.61	2.99
School was out at the end of May.	4.91	3.45
They finally set their wedding date,	4.86	5.72
August 23rd.	4.12	4.88
They were married a year after they met.	5.64	5.37

APPENDIX (Continued)

Idea Unit	Mean Memory Rating	Mean Importance Rating
Susan Story		
Susan	5.72	6.24
recently	4.34	3.77
moved to Cleveland.	5.04	5.06
On her first day there,	4.76	3.38
she bought a newspaper	3.03	2.69
and went apartment hunting.	5.32	4.74
She contacted several landlords	3.71	3.09
and found an apartment in a complex	3.96	4.29
overlooking a park.	4.09	2.88
After she finished unpacking,	2.54	2.34
she explored the city.	3.94	3.50
In her search for employment,	4.15	4.80
Susan applied for a writing assignment	4.63	5.60
with a weekly magazine.	3.50	4.08
She interviewed for a columnist position	4.11	5.48
with a local newspaper.	3.88	4.18
Susan applied as a copy editor	3.88	5.20
for a publishing firm.	3.10	4.13
She also submitted her resume	3.79	3.96
to three high schools	4.56	4.39
for teaching jobs.	5.50	5.27
Susan had just received	3.25	3.48
her B. A.	5.30	5.46
in English.	5.46	5.73
She also has a teaching certificate.	5.32	5.61
She completed her studies at Drake University	4.77	4.48
in Iowa.	4.18	2.89
Susan took an average course load	4.55	2.37
and graduated in four years.	5.47	3.19
<i>The courses that she did the best in (C)</i>	3.88	4.35
<i>The courses that she did the worst in (I)</i>	5.00	4.56
were writing courses.	5.26	5.00
As an undergraduate,	2.82	3.29
she worked for the school paper	4.36	4.56
and was editor	4.07	5.05
of the yearbook.	4.09	4.32
She also participated in theatrical productions.	4.39	3.43
After working to repay	2.85	3.16
her student loans	3.73	3.43
Susan would like to pursue serious writing.	5.00	5.57
She is interested in publishing	3.87	4.84
children's stories	4.40	4.63
and plays.	4.20	4.62

(Manuscript received March 26, 1986;
revision accepted for publication August 18, 1986.)