Journal of Educational Psychology
2005, Vol. 97, No. 1, 70–80

0022-0663/05/$12.00 DOI: 10.1037/0022-0663.97.1.70

Rereading Effects Depend on Time of Test

Katherine A. Rawson and Walter Kintsch
University of Colorado at Boulder

Previous research has shown better text learning after rereading versus 1 reading of a text. However, rereading effects have only been explored using immediate tests, whereas most students face delays between study and test. In 2 experiments, 423 college students read a text once, twice in massed fashion, or twice with 1 week between trials. Students were tested either immediately or 2 days after study. On an immediate test, performance was greater after massed versus single reading, whereas performance for distributed rereading was not significantly greater than after single reading. On a delayed test, performance was greater after distributed versus single reading, whereas performance for massed rereading and single reading no longer differed significantly.

Rereading is a study strategy commonly recommended to students when preparing for upcoming examinations and one they often report using (see, e.g., Amlund, Kardash, & Kulhavy, 1986; Feldt & Ray, 1989; Nist, Simpson, & Hogrebe, 1985). For example, Carrier (2003) found that 65% of his sample of upper level college students reported rereading textbook chapters as a study strategy, which was the most frequently reported strategy. The use of rereading as a study strategy is not unfounded. A wealth of research has shown that test performance is greater after a second reading trial than after one reading trial (see, e.g., Amlund et al., 1986; Barnett & Seefeldt, 1989; Bromage & Mayer, 1986; Glover & Corkill, 1987; Krug, Davis, & Glover, 1990; Mayer, 1983; Meyer & McConkie, 1973; Rawson, Dunlosky, & Thiede, 2000; Rothkopf, 1968). The effect of rereading on text learning appears to be particularly robust given that it has been demonstrated using text materials on various topics and of variable length, with both higher skilled and less skilled readers, and using various kinds of test questions.

Although an increasing amount of research has been conducted to establish and explain rereading effects, there are notable gaps in the literature. Perhaps of greatest concern, almost all previous research on rereading effects has used immediate testing. The question of whether rereading effects obtain when tests are administered at a delay has not yet been adequately evaluated. In one of the few studies examining this variable, Dyer, Riley, and Yekovich (1979) had participants read a text either once or twice within an experimental session (the delay between trials was not specified), and they were then tested both immediately and again after 1 week. Performance was greater after rereading than after a single reading on both the immediate and delayed tests, although the specific significance tests for these comparisons were not reported. Additionally, it is unclear whether the preservation of the rereading effect across a delay may have been influenced by the retrieval practice provided by the immediate test. Other studies that have included delayed testing have interleaved study and test trials, raising similar concerns (Haenggi & Perfetti, 1992; Howe, 1970). Nonetheless, these results suggest that improvements in learning with rereading may be durable.

This question of whether rereading effects obtain on delayed tests is particularly important with respect to student scholarship because learning in educational settings is seldom tested immediately after study. Additionally, the ultimate goal of learning strategies or tools is to promote long-term, durable learning and not just transient gains. Thus, given that the previous research supporting the recommendation of rereading as a strategy for improving student learning has only used immediate testing, examining rereading effects with delayed testing is of obvious importance. Accordingly, the primary goal of the present research was to examine rereading effects with delayed testing in addition to immediate testing.

A second issue that needs further research concerns the interval between study trials. In almost all previous research, the advantage of rereading over a single reading trial has been shown using massed rereading (i.e., when the first and subsequent study trials are administered in immediate succession). Few studies have examined (a) the extent to which distributed rereading (i.e., when the second study trial is administered at some delay after the initial trial) also yields a performance advantage over single reading and (b) the extent to which performance is greater for massed or distributed rereading (see, e.g., Glover & Corkill, 1987; Krug et al., 1990). In the study by Glover and Corkill (1987), readers studied a short paragraph twice, either in massed fashion or with a 30-min delay between study trials (the study did not include a single reading baseline). Free recall was greater after distributed rereading than after massed rereading. Similarly, Krug et al. (1990) demonstrated greater free recall after distributed rereading (with 1 week between study trials) than after a single reading or after massed rereading, suggesting that distributed rereading may be a more effective strategy than massed rereading for improving learning outcomes. However, given that distributed rereading has been

Katherine A. Rawson and Walter Kintsch, Department of Psychology, University of Colorado at Boulder.

This research was supported by a National Science Foundation Graduate Research Fellowship, and preparation of this report was supported by a National Institute of Mental Health National Research Service Award Predoctoral Fellowship to Katherine A. Rawson. Many thanks to Jessica Able and Greg Hofer for their assistance in data collection and scoring.

Correspondence concerning this article should be addressed to Katherine A. Rawson, who is now at Kent State University, Department of Psychology, P.O. Box 5190, Kent, OH 44242-0001. E-mail: krawson1@kent.edu

examined in relatively few studies using a limited range of texts and test measures, the generality and reliability of distributed rereading effects have yet to be established. Related research on spacing effects has suggested that performance advantages from distributed practice may not be as reliable as is commonly assumed. Although spacing effects are quite robust with motor and simple verbal learning tasks, a recent meta-analysis by Donovan and Radosevich (1999) suggested that the effects are weaker, are less reliable, and may even reverse as the overall complexity of the cognitive task increases. Thus, in addition to the primary goal of examining rereading effects under both immediate and delayed testing, the secondary goal of the present research was to further investigate the effects of distributed rereading.

Experiment 1

In Experiment 1, readers studied a lengthy expository text either once, twice in massed fashion, or twice with a 1-week delay between study trials. Half of the readers were tested immediately after the final study trial (i.e., after the first study trial for those who only read once and after the second study trial for those who reread the text), and half of the readers were tested after a 2-day delay. The test consisted of a recall measure as well as short answer questions tapping comprehension of the material.

The immediate test group provided a replication of the procedures used in previous research. On the basis of the results of this earlier research, we predicted that performance would be greater after rereading than after a single reading in this group. The primary question of interest in this experiment concerned the extent to which the same pattern would obtain in the delayed test group. The secondary question of interest concerned the reliability and generality of the effects of distributed rereading demonstrated in previous research (e.g., we used lengthier text than in the earlier studies and administered both memory and comprehension tests). On the basis of the initial results reported in previous research, we expected performance would be greater after distributed rereading than after single reading or after massed rereading.

Method

Participants and design. Participants were 235 undergraduates from University of Colorado who participated to partially satisfy a course requirement in an introductory psychology class, which primarily consisted of freshman and sophomore students. Demographic information was not collected for participants in Experiments 1 or 2. However, given that no demographic exclusion criteria for participation were used, our samples presumably approximate the demographic characteristics of the population of University of Colorado undergraduates from which they were drawn (47% female, 13% minority, 1% international). Participants were randomly assigned to one of six experimental groups, defined by the factorial combination of two independent variables: study (single, massed, or distributed) and test (immediate or delayed). Each group contained 39 participants except for the massed/delayed test group ($n = 40$).

Materials and procedure. The text was excerpted from an article in *Scientific American* (Herzog, Eliasson, & Kaarstad, 2000) about carbon sequestration, a method for reducing the amount of carbon dioxide in the atmosphere. The experimental text was 1,730 words in length, with a Flesch reading ease score of 34.4 and a Flesch-Kincaid grade level of 12. The text was divided into five sections, each introduced with a title or subtitle. A Macintosh computer presented all experimental instructions and materials and recorded all data.

All participants were instructed that they would be reading a scientific text and that they would subsequently be tested. Participants assigned to the massed group were informed that they would receive a second study trial immediately after the first. Participants assigned to the distributed group were informed that they would receive a second study trial 7 days later. All participants were told that the test would involve both recall and short-answer questions requiring memory and comprehension. Participants were also informed whether they would be taking the test immediately after study or 2 days later.

Before reading the text, all participants were asked to complete a brief prior knowledge assessment in which they were given a blank text field on the computer screen and were asked to type in what they knew about global warming. (Inclusion of this factor in statistical analyses did not qualitatively change the outcomes of interest for present purposes, and so this measure is not discussed further.) Next, all participants read the text once. Sections of the text were presented one at a time, each in its entirety, on the computer screen for self-paced study. After reading a section of the text, the participant advanced to the next section with a keypress. Participants were not permitted to return to previously read sections once they had advanced. After reading the text once, participants in the massed group immediately reread the text in the same manner as in the first study trial. Participants in the distributed group were dismissed and asked to return 1 week later, at which time they reread the text in the same manner as in the first study trial. In each study group, participants assigned to the immediate test group completed the test immediately after the final study trial (i.e., the first trial for the single group and the second for the other two groups). Participants in the delayed test group were dismissed after the final study trial and asked to return 2 days later to complete the test.

The first component of the test was a self-paced recall test in which participants were provided with the subtitle from one of the text sections and were asked to type in everything they could remember from that section. All participants were asked to recall the same section, which was the second of the five sections. This section was chosen for the recall test because it had the least amount of content overlap with other sections of the text and thus would support relatively unambiguous scoring of recall. After participants indicated that they had completed the recall test with a keypress, they were asked 12 short-answer comprehension questions (e.g., requiring integration of text information, inferencing, or application). Collectively, the questions tapped information from all sections in the text. Questions were presented one at a time, and participants were not permitted to revisit previously answered questions.

Results and Discussion

In both Experiments 1 and 2, all protocols were scored blind with respect to experimental group. Approximately 10% of the protocols were scored independently by two graders. The two graders then compared their scoring of this subset and discussed any discrepancies. The remaining protocols were then scored by one of the two graders.

Scores on the recall test were based on the number of idea units recalled (verbatim or paraphrase) from the target section. Each idea unit approximately corresponded to one phrase of a sentence. For example, the sentence “Sleipner offshore oil and natural gas field is in the middle of the North Sea, off the coast of Norway” consists of three idea units, “Sleipner is an offshore oil and natural gas field,” “(Sleipner) is in the middle of the North Sea,” and “(Sleipner) is off the coast of Norway.” In total, the target section consisted of 43 idea units. Intrusions of information from other sections of the text were relatively infrequent and were not included in the recall score.

Responses to the 12 short-answer comprehension questions were assigned scores from 0 (no answer or completely wrong answer) to 1 (complete, correct answer), with partial credit given when a response contained some but not all of the correct information. For example, for the question "What is an important difference between using solar power and planting trees with respect to how they each address the problem of global warming?" participants received full credit if they explained that (a) trees reduce the current level of carbon dioxide in the atmosphere, whereas (b) solar power does not reduce the current level but does reduce the further release of carbon dioxide. Partial credit was given if one but not both of these parts to the answer was provided. Scores on the comprehension test are reported as a percentage of total points possible. The split-half reliability for the comprehension questions using the Spearman-Brown correction formula was .80. Although not of primary interest in the present study, reading times for each group are reported in the Appendix for purposes of completeness.

Recall performance. For recall of idea units, the mean for each group is reported in Figure 1. A 3 (study group) × 2 (time of test) analysis of variance (ANOVA) revealed a significant main effect of study group, $F(2, 228) = 4.85$, $MSE = 28.82$, $p = .009$; a significant main effect of time of test, $F(1, 228) = 13.57$, $p < .001$; and a significant interaction, $F(2, 228) = 9.74$, $p < .001$. As revealed by a series of follow-up comparisons, on the immediate test, recall was greater after massed rereading than after a single reading, $F(1, 75) = 8.31$, $MSE = 49.78$, $p = .005$, $d = .66$. However, the same advantage was not evident with distributed rereading: Recall for the distributed rereading and single reading groups did not significantly differ, $F < 1$. Furthermore, as is obvious from inspection of Figure 1, recall was not greater with distributed rereading than with massed rereading. Indeed, the trend was in the opposite direction, with recall significantly greater after massed rereading than after distributed rereading, $F(1, 76) = 8.60$, $MSE = 42.07$, $p = .004$, $d = .66$.

The advantage of massed rereading over single reading on an immediate test is consistent with the robust findings reported in much of the previous research. In contrast, our finding that recall performance on an immediate test was better for massed than for distributed rereading is inconsistent with the results of earlier research. This outcome suggests that distributed rereading effects may be less robust than previously assumed. We postpone further discussion of this issue until the General Discussion.

In contrast to performance on the immediate test, recall on the delayed test did not differ for the massed rereading and single reading groups, $F < 1$, whereas recall was greater after distributed rereading than after a single reading, $F(1, 76) = 17.21$, $MSE = 20.24$, $p = .001$, $d = .82$. Furthermore, recall was greater after distributed rereading than after massed rereading, $F(1, 77) = 10.58$, $MSE = 19.59$, $p = .002$, $d = .73$.

As is apparent from inspection of Figure 1, with single reading or massed rereading, recall performance was lower for the delayed test group than for the immediate test group. The difference between the two massed rereading groups was particularly substantial (a 55% decrease), $F(1, 77) = 22.26$, $MSE = 29.92$, $p < .001$, $d = 1.06$. In contrast, after distributed rereading, performance did not significantly differ for the immediate test and delayed test groups. Taken together, these results suggest that although distributed rereading may lead to less initial learning than massed rereading, the learning acquired from distributed rereading is more durable.

Comprehension performance. Mean comprehension performance for each group is reported in Figure 2. A 3 × 2 ANOVA revealed a significant main effect of study group, $F(2, 229) = 5.94$, $MSE = 255.76$, $p = .003$; a significant main effect of time of test, $F(1, 229) = 4.80$, $p = .029$; and a significant interaction, $F(2, 229) = 3.30$, $p = .039$. In sum, the pattern of results for the comprehension test was highly similar to the pattern of recall. On the immediate test, comprehension performance was greater after massed rereading than after a single reading, $F(1, 76) = 5.90$, $MSE = 276.22$, $p = .018$, $d = .55$, whereas comprehension performance for the distributed rereading and single reading

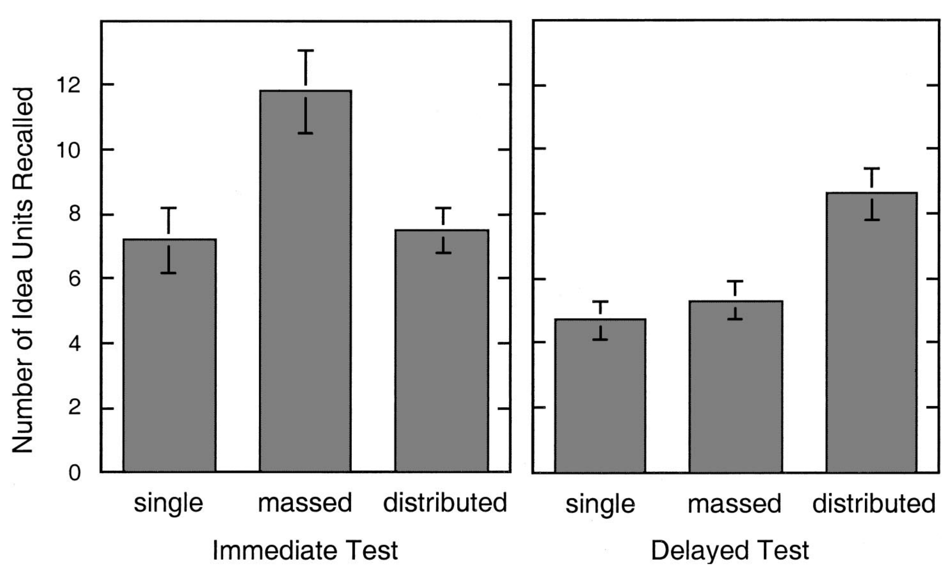

Figure 1. Number of idea units recalled as a function of study group (single, massed rereading, or distributed rereading) and time of test (immediately after study or with a 2-day delay between study and test) for Experiment 1. Error bars represent standard errors of the mean.

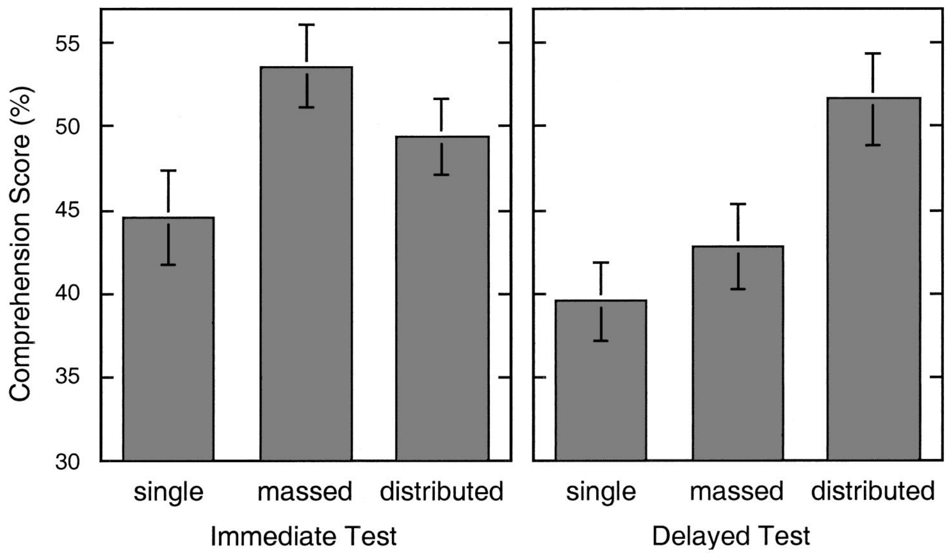

Figure 2. Performance on short-answer comprehension questions (plotted as percentage correct) as a function of study group (single, massed rereading, or distributed rereading) and time of test (immediately after study or with a 2-day delay between study and test) for Experiment 1. Error bars represent standard errors of the mean.

groups did not significantly differ, $F < 1.86$. Furthermore, comprehension performance was not greater after distributed rereading than after massed rereading, with a nonsignificant trend in the opposite direction, $F < 1.53$.

On the delayed test, comprehension performance was greater after distributed rereading than after a single reading, $F(1, 76) = 11.29$, $MSE = 255.20$, $p = .001$, $d = .76$, whereas comprehension performance for the massed rereading and single reading groups did not significantly differ, $F < 1.1$. Comprehension performance was also greater after distributed rereading than after massed rereading, $F(1, 77) = 5.61$, $MSE = 275.72$, $p = .02$, $d = .53$.

As was the case with recall, with single reading or massed rereading, comprehension performance was lower for the delayed test group than for the immediate test group, with the most marked drop after massed rereading (20%), $F(1, 77) = 9.03$, $MSE = 258.93$, $p = .004$, $d = .68$. However, after distributed rereading, performance did not significantly differ for the immediate test and delayed test groups, $F < 1$.

Summary. The primary goal of this research was to examine whether the effects of rereading depend on the delay between study and test. The results clearly demonstrate that they do. On an immediate test, performance was greater after massed rereading than after distributed rereading or a single reading, which did not differ from one another. In contrast, on a delayed test, performance was greater after distributed rereading than after massed rereading or a single reading, which did not differ from one another. The shift in the overall pattern from immediate to delayed testing was primarily due to the fact that learning with distributed rereading was durable and did not differ with test delay, whereas learning with massed rereading was more fragile and dropped sharply with test delay.

Experiment 2

Experiment 1 provided the first demonstration in the literature that the overall pattern of rereading effects shifts from immediate to delayed testing. As is the case with any novel finding, establishing the reliability and generality of the effect was important. Thus, Experiment 2 was designed to provide a conceptual replication of Experiment 1, using different text material and different memory tests. Experiment 2 was also intended to replicate the finding that distributed rereading does not yield greater performance than single reading or massed rereading on the immediate test, unlike results reported in the earlier research on distributed rereading effects.

The secondary goal of Experiment 2 was to explore one hypothesis for why massed and distributed rereading showed different patterns of performance on immediate versus delayed tests. To revisit, massed rereading produced a sizable recall advantage on the immediate test (relative to a single reading), but this advantage was almost completely lost after a 2-day delay. Previous research exploring the relative durability of information at various levels of representation suggests that memory for higher level content tends to survive across delay intervals, whereas lower level content is much less likely to be retained across a delay (see, e.g., Kintsch, Welsch, Schmalhofer, & Zimny, 1990). These findings suggest that the advantage of massed rereading on the immediate test may have been due to greater emphasis on encoding and recall of lower level text content, given that the advantage was no longer evident after a 2-day delay. In contrast, learning with distributed rereading was quite durable and did not differ with test delay. Results from the previous research suggest that the durability of learning with distributed rereading may have reflected greater emphasis on encoding and recall of higher level text information. To explore the hypothesis that the interaction of time of restudy and time of test was due to differential emphasis on encoding and recall of content at different levels (hereafter referred to as the *levels hypothesis*), the text used in Experiment 2 included target sentences that were designed to primarily contain either main ideas, important details, or unimportant details (described further below). Thus, in addition to overall levels of recall, we examined the qualitative pattern of recall across these kinds of information in the rereading groups to test the levels hypothesis.

Method

Participants and design. The participants were 168 undergraduates from University of Colorado who participated to partially satisfy a course requirement in introductory psychology. Participants were randomly assigned to six experimental groups defined by the factorial combination of two independent variables: study (single, massed, or distributed) and test (immediate or 2-day delay). Each group contained between 26 and 30 participants (because of some participants who failed to return to complete the experiment).

Materials and procedure. The text was adapted from a popular-reading nonfiction book (Carnes, 1995) about how Hollywood films portray history. The text discussed how feature films usually portray history inaccurately, explained the reasons for inaccuracies, and then illustrated these points by comparing a particular film (*The Charge of the Light Brigade* [Curtiz, 1936]) to the actual historical events on which it was based (the Crimean War). The text was 1,541 words in length, with a Flesch reading ease score of 45.9 and a Flesch-Kincaid grade level of 11.5. The text was presented in five sections as in Experiment 1, except that the sections did not have subtitles.

A subset of the sentences within the text was specifically revised so that the content of a sentence primarily expressed either a main idea (e.g., "Much of the 'history' that is portrayed on screen is fiction, not fact"); an important detail, which was more specific information that further expanded or supported a main idea (e.g., "One important way in which history is fictionalized is by leaving out important historical events altogether"); or an unimportant detail, which was specific information that was not relevant to understanding the main ideas (e.g., "The scene of the cavalry charge was one of the most expensive scenes ever filmed at the time"). A pilot study was then conducted to collect normative ratings of the kind of information contained in each sentence. Forty-three undergraduates were presented with the text once to familiarize them with the content and were then presented with the text again, one sentence at a time in order. Although all text sentences were represented, ratings were collected only for those sentences revised as described above. After each of the revised sentences, participants were asked to indicate whether they thought the sentence expressed a main idea, an important detail, or an unimportant detail (each of these categories was explained to the participants in the instructions) or whether they were unsure to which category the sentence belonged. From the subset of sentences for which ratings were collected, those sentences of each kind showing highest agreement (5 main-idea sentences, $M = 67\%$ agreement; 10 important-detail sentences, $M = 75\%$ agreement; 10 unimportant-detail sentences, $M = 75\%$ agreement) were selected as target sentences for the secondary analyses described in the *Results and Discussion* section below.

Instructions and procedure were the same as in Experiment 1, with the following exceptions. First, for the prior knowledge assessment, the prompt was as follows: "(1) What do you know about the Crimean War? (2) Have you ever seen the movie *Charge of the Light Brigade*? If so, what do you remember from the movie?" Almost no participant knew anything about the Crimean War, and none had seen the movie, so the results of the prior knowledge assessment are not discussed further. The first component of the test was a self-paced free-recall test in which the participants were asked to type in as much of the entire text as they could remember (rather than just one section cued with a subtitle, as in Experiment 1). Participants were then presented with 19 short-answer and fill-in-the-blank questions (hereafter referred to collectively as short-answer questions) that tapped memory for information explicitly stated in the text.

Results and Discussion

The scoring of overall recall was highly similar to Experiment 1, except that a slightly larger grain size of analysis was used in designating the idea units contained in the text (roughly equivalent to scoring based on the central content of each clause rather than each phrase). For example, for the text excerpt "Movies have a tremendous power to shape what we know about history. Many of the greatest figures of history have become inextricably linked to their screen images. Joan of Arc has become Ingrid Bergman in our minds, and Malcolm X is the person Denzel Washington has shown us on screen," the following four idea units were included on the scoring key: "movies can shape our knowledge of history," "historical figures are linked to their screen images," "Joan of Arc is Ingrid Bergman," and "Malcolm X is Denzel Washington." This change in the grain size of analysis for scoring from phrase in Experiment 1 to clause in Experiment 2 was largely pragmatic, given that the entire text was scored in Experiment 2 versus just one section in Experiment 1. Even segmenting the text based roughly on clauses rather than phrases resulted in a sizable number of units for scoring—125 possible idea units to be scored—and using even smaller units of analysis would have resulted in an impractically unwieldy scoring key. We should note that because the effects of interest are within experiment rather than between experiments and, to foreshadow, the same pattern of results obtains in both experiments, this slight change in the unit for recall scoring should not present interpretive difficulties.

For the target sentences, three recall scores were computed, one for each kind of sentence. For each kind, the number of idea units recalled from those sentences was computed and then converted to a percentage of the total number of idea units of that kind. For the short-answer questions, responses were assigned scores from 0 (no answer or completely wrong answer) to 1 (complete, correct answer), with partial credit given when a response contained some but not all of the correct information. Scores for the short-answer test are reported as a percentage of total points possible. The split-half reliability of the short-answer instrument using the Spearman-Brown correction formula was .73. For purposes of completeness, reading times for each group are reported in the Appendix.

Overall recall performance. Mean number of idea units recalled for each group is reported in Figure 3. A 3 × 2 ANOVA revealed a significant main effect of study group, $F(2, 162) = 4.26$, $MSE = 92.36$, $p = .016$; a significant main effect of time of test, $F(1, 162) = 25.73$, $p < .001$; and a significant interaction, $F(2, 162) = 5.50$, $p = .005$.

As revealed by follow-up comparisons, the pattern of recall was highly similar to that observed in Experiment 1. On the immediate test, recall was greater after massed rereading than after a single reading, $F(1, 51) = 9.07$, $MSE = 128.51$, $p = .004$, $d = .83$, whereas recall for the distributed rereading and single reading groups did not significantly differ, $F < 1$. Furthermore, as is obvious from inspection of Figure 3, recall was not greater with distributed rereading than with massed rereading, with a significant trend in the opposite direction, $F(1, 51) = 6.87$, $MSE = 146.42$, $p = .012$, $d = .72$.

On the delayed test, recall did not differ for the massed rereading and single reading groups, $F < 1$, whereas recall was significantly greater after distributed rereading than after a single read-

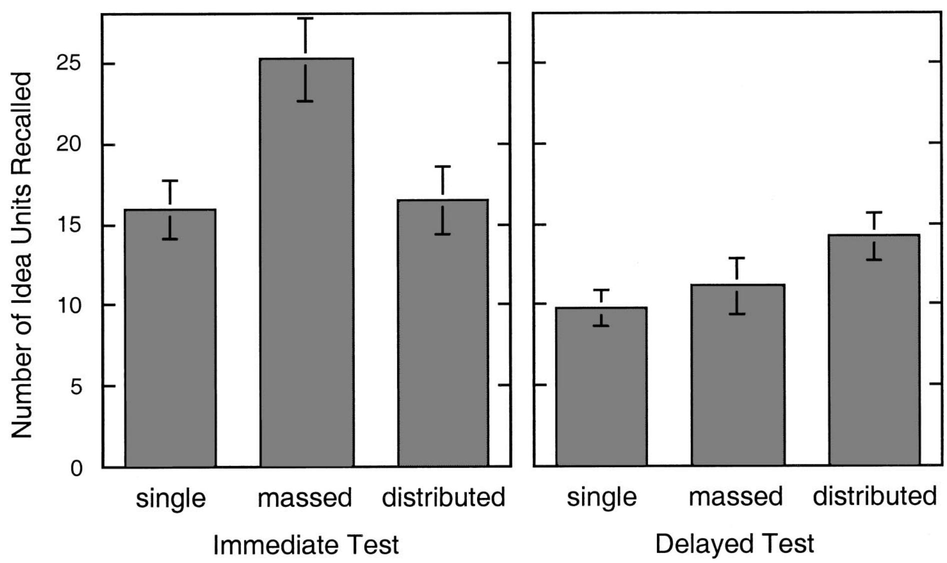

Figure 3. Number of idea units recalled as a function of study group (single, massed rereading, or distributed rereading) and time of test (immediately after study or with a 2-day delay between study and test) for Experiment 2. Error bars represent standard errors of the mean.

ing, $F(1, 57) = 5.64$, $MSE = 49.95$, $p = .021$, $d = .62$. Recall was greater after distributed rereading than after massed rereading, although the trend failed to reach significance, $F(1, 56) = 1.92$, $p = .171$.

As is apparent from inspection of Figure 3, with single reading or massed rereading, recall performance was lower on a delayed test than on an immediate test. The difference between the two massed rereading groups was again substantial (56% decrease), $F(1, 53) = 22.05$, $MSE = 125.28$, $p < .001$, $d = 1.27$. In contrast, after distributed rereading, performance did not significantly differ for the immediate test and delayed test groups, $F < 1$.

Recall of target sentences. The secondary goal of Experiment 2 was to evaluate the levels hypothesis, which states that the shift in rereading effects from immediate to delayed testing is due to differential emphasis on encoding and recall of lower level versus higher level information in the massed and distributed rereading groups, respectively. In Figure 4, recall of the three kinds of target sentence is plotted for the massed and distributed rereading groups. For each group, target sentence recall is plotted at three different times. The leftmost set of points in each panel reflects recall immediately prior to the second study trial in each group, to estimate the relative accessibility of information that participants in the rereading groups would have had right before rereading. For the massed rereading group, these values were obtained by scoring recall of target sentences for participants in the single reading group who were tested immediately. For the distributed rereading group, these values were obtained by scoring recall of target sentences from an additional 20 participants (from the same pool of participants from which the six experimental groups were drawn) who were randomly assigned to a group in which the text was presented once. These participants were then tested after a 7-day delay, to reflect the interval between first and second study trials for the distributed rereading group. The middle set of points in each panel reports recall on the immediate test in the two rereading groups, and the rightmost set of points reports recall on the delayed test in the two rereading groups.

First, consider the comparisons between the "before Trial 2" points in the two panels. Taken together, these points conceptually replicate the earlier research on the durability of information at different levels (see, e.g., Kintsch et al., 1990), which provides a further check of the manipulation of target sentence level. After reading a text only once, recall of main ideas was preserved across a week interval ($F < 1$), whereas recall for detail information was significantly lower after a week, $F(1, 45) = 14.90$, $MSE = 81.37$, $p < .001$, $d = 1.14$, and $F(1, 45) = 10.78$, $MSE = 59.39$, $p = .002$, $d = .96$, for important and unimportant details, respectively.

Next, consider the pattern of results in the left panel, which plots results for massed rereading. Inconsistent with the levels hypothesis, the advantage of massed rereading on the immediate test was not due to greater emphasis on lower level information than on higher level information, as can be seen by comparing the gains in recall from before Trial 2 to after Trial 2 for each kind of idea unit. A 2 (time of recall) × 3 (kind of idea unit) mixed-factor ANOVA yielded significant main effects for both factors, $F(1, 51) = 23.85$, $MSE = 163.97$, $p < .001$, and $F(2, 102) = 20.75$, $MSE = 105.06$, $p < .001$, respectively, as well as a significant interaction, $F(2, 102) = 3.74$, $MSE = 105.06$, $p < .031$. In contrast to the predictions of the levels hypothesis, gains with massed rereading were most pronounced for main ideas and least pronounced for unimportant details; for main ideas, $F(1, 51) = 33.52$, $MSE = 100.77$, $p < .001$, $d = 1.59$; for important details, $F(1, 51) = 5.38$, $MSE = 173.58$, $p = .024$, $d = .64$; and for unimportant details, $F(1, 51) = 3.87$, $MSE = 99.73$, $p = .055$, $d = .54$. Somewhat surprisingly, these gains in main idea recall did not hold up across a test delay, as can be seen by comparing recall from after Trial 2 to recall after 2 days. Recall after massed rereading for all three kinds of information was lower on a delayed test; for main ideas, $F(1, 53) = 29.56$, $MSE = 104.68$, $p < .001$, $d = 1.47$; for important details,

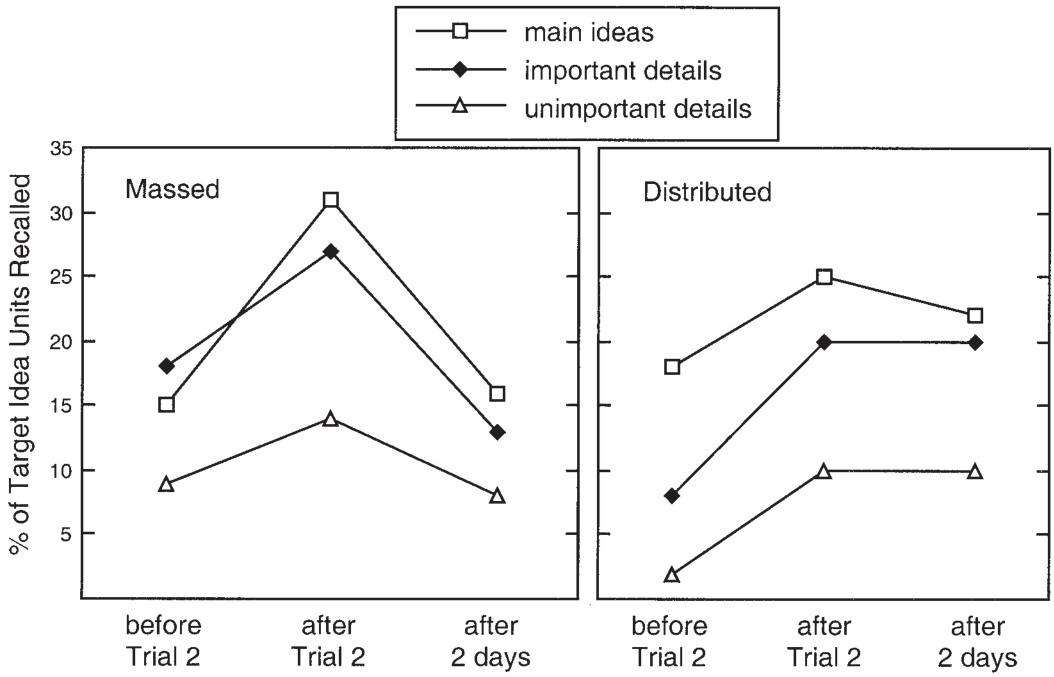

Figure 4. Percentage of idea units recalled from target sentences as a function of rereading group (massed or distributed), kind of target sentence (main idea, important detail, or unimportant detail; see text for further description of these distinctions), and time of recall (immediately prior to the second study trial, immediately after the second study trial, or 2 days after the second study trial in each group) in Experiment 2.

$F(1, 53) = 13.86$, $MSE = 181.94$, $p < .001$, $d = 1.01$; and for unimportant details, $F(1, 53) = 4.17$, $MSE = 110.82$, $p = .046$, $d = .55$.

Finally, consider the pattern of results in the right panel, which plots results for distributed rereading. As is evident in the leftmost set of points, 1 week after a single reading, recall of text content primarily consisted of main ideas and was much lower for detail information. Comparison of recall from before Trial 2 to after Trial 2 suggests that during a subsequent rereading trial, readers focused on encoding information at all levels, rather than primarily on further encoding of main idea information, as suggested by the levels hypothesis. A 2 (time of recall) $\times$ 3 (kind of idea unit) mixed-factor ANOVA yielded significant main effects for both factors, $F(1, 45) = 13.95$, $MSE = 216.12$, $p = .001$, and $F(2, 90) = 37.84$, $MSE = 72.68$, $p < .001$, but the interaction was not significant, $F < 1$. Recall was greater for all three kinds of information; for main ideas, $F(1, 45) = 5.22$, $MSE = 129.98$, $p = .027$, $d = .68$; for important details, $F(1, 45) = 11.98$, $MSE = 132.51$, $p = .001$, $d = 1.03$; and for unimportant details, $F(1, 45) = 8.63$, $MSE = 99.00$, $p = .005$, $d = .86$. Again somewhat surprisingly, gains in recall at all levels of information held up across the test delay interval, Fs < 1. One possibility is that during rereading, readers not only encoded detail information but also related it to the main ideas in the text. Subsequent recall of those main ideas on the delayed test would then provide links to related detail information, facilitating recall of those details (see Murray & McGlone, 1997).

Short-answer performance. Mean short-answer performance for each group is reported in Figure 5. In sum, the pattern of results is highly similar to those for overall recall. A 3 $\times$ 2 ANOVA revealed nonsignificant main effect of study group, $F(2, 161) = 2.31$, $MSE = 0.04$, $p = .103$; a significant main effect of time of test, $F(1, 161) = 27.36$, $p < .001$; and a significant interaction, $F(2, 161) = 3.70$, $p = .027$.

On the immediate test, short-answer performance was greater after massed rereading than after a single reading, $F(1, 51) = 4.68$, $MSE = 0.04$, $p = .035$, $d = .57$, whereas performance for the distributed rereading and single reading groups did not significantly differ, $F < 1$. There was also a marginal trend for greater short-answer performance on the immediate test after massed rereading than after distributed rereading, $F(1, 50) = 3.69$, $MSE = 0.04$, $p = .061$, $d = .56$.

On the delayed test, short-answer performance was greater after distributed rereading than after a single reading, $F(1, 57) = 6.83$, $MSE = 0.04$, $p = .011$, $d = .66$, whereas performance for the massed rereading and single reading groups did not significantly differ, $F < 1$. Finally, there was a marginal trend for greater short-answer performance on the delayed test after distributed rereading than after massed rereading, $F(1, 56) = 3.58$, $MSE = 0.04$, $p = .064$, $d = .48$.

As was the case for recall, with single reading or massed rereading, short-answer performance was lower for the delayed test group than for the immediate test group, with the most marked drop after massed rereading (39%), $F(1, 53) = 24.00$, $MSE = 0.04$, $p < .001$, $d = 1.30$. However, after distributed rereading, performance did not significantly differ for the immediate test and delayed test groups, $F < 1$.

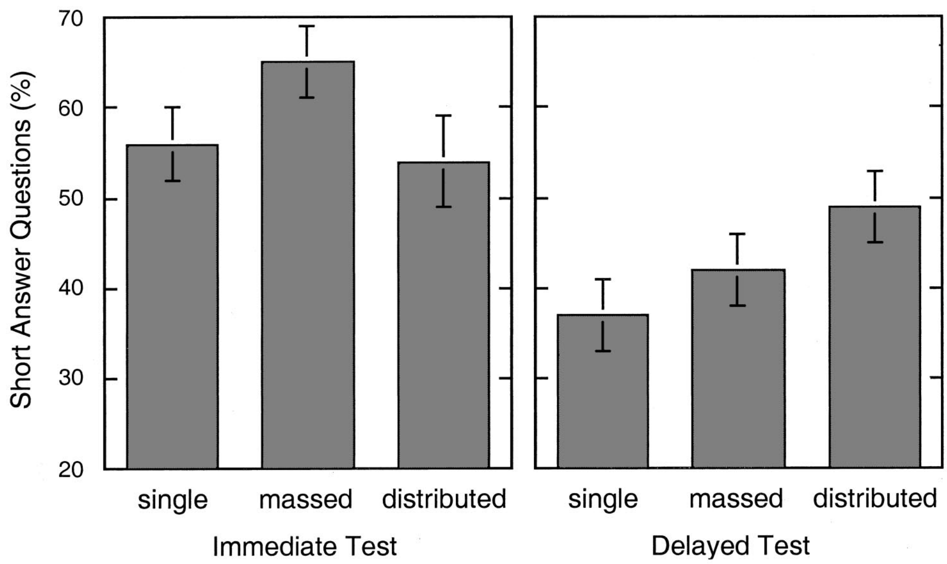

Figure 5. Performance on short-answer memory questions (plotted as percentage correct) as a function of study group (single, massed rereading, or distributed rereading) and time of test (immediately after study or with a 2-day delay between study and test) for Experiment 2. Error bars represent standard errors of the mean.

Summary. The most important outcomes of Experiment 2 concern the replication of the main findings from Experiment 1. Using different text material, we again found that rereading effects depended on time of test. Performance on the immediate test was greater after massed rereading than after a single reading. However, performance was no longer significantly greater after massed rereading on the delayed test. Distributed rereading showed the opposite pattern. Performance on the immediate test did not significantly differ after distributed rereading versus a single reading, whereas performance on the delayed test was greater after distributed rereading than after a single reading. As in Experiment 1, the shift in the overall pattern from immediate to delayed testing was due primarily to a drop in performance with massed rereading and relatively minimal differences in performance with distributed rereading. One account of this pattern, the levels hypothesis, was not supported by the results of secondary analyses, a point that we discuss further below.

General Discussion

The primary goal of the present research was to explore the extent to which rereading effects depend on test delay. Two experiments using different text materials and different tests yielded a consistent pattern of results: Massed but not distributed rereading improves performance on an immediate test, whereas distributed but not massed rereading improves performance on a delayed test. Thus, the present research establishes a finding with important implications for student scholarship, given that rereading is a study strategy prevalently used by students who are typically preparing for delayed tests. Below, we discuss two further issues that may suggest directions for future research. We then briefly consider these findings in the broader context of research on study strategies.

Why Do Rereading Effects Depend on Test Delay?

In Experiment 2, we evaluated one hypothesis for why massed and distributed rereading showed different patterns of performance on immediate and delayed tests. According to the levels hypothesis, the immediate test advantage for the massed rereading group is due to emphasis on encoding and recall of detail information, whereas the durability of performance in the distributed rereading group is due to emphasis on encoding and recall of main ideas. The results of Experiment 2 do not support the levels hypothesis: Comparison of recall prior to versus immediately after rereading revealed improvements for information at all levels with both massed and distributed rereading, with no clear advantage for detailed information with massed rereading nor for main ideas with distributed rereading.

Although the results from Experiment 2 weigh against the levels hypothesis, this in itself is informative—part of the process of understanding a new phenomenon is constraining the set of potential accounts of the effect. Of course, further research is needed before the levels hypothesis can be definitively ruled out. It may be the case that other methods may provide measures that are more sensitive than the recall task used in the present research to evaluate the hypothesis. Online measures would appear to be particularly promising. For example, text material could be presented in a sentence-by-sentence fashion to allow examination of reading times for sentences containing main ideas, important details, or unimportant details in the different experimental groups. The levels hypothesis predicts that readers in the massed rereading group would spend more time on sentences containing detailed information, whereas readers in the distributed rereading group would spend more time on sentences containing main ideas. Similarly, think-aloud protocols could be collected during study, to examine the extent to which readers in the two groups place differential emphasis on the various kinds of information.

Future research could also explore other accounts of why massed and distributed rereading showed different patterns of performance on immediate and delayed tests. One possibility is that the critical difference was not due to what information was processed but to how the information was processed. Readers in the two groups may have differed in the kinds of processing operations applied to the text content. Specifically, during massed rereading, readers may have primarily allocated their processing resources to encoding more pieces of text information, investing fewer resources in the integration and organization of that information. In contrast, during distributed rereading, readers may have allocated fewer resources to the encoding of individual pieces of text information and more to the integration and organization of the text information. If so, distributed rereading would result in a slighter but tighter representation than massed rereading, one with a structure that would be more supportive of recall after a delay. Although speculative, this account provides a plausible explanation of the current pattern of findings that we will be evaluating in future research.

Distributed Rereading Effects: Reconciling With Previous Research

As noted earlier, most previous research on rereading has been limited not only to examination of performance on immediate tests but also to examination of rereading effects using only massed rereading. Few studies have examined rereading effects using distributed study trials. In fact, only two previous studies have examined both massed and distributed rereading within the same experiment. To revisit, in the study by Glover and Corkill (1987), participants read a short text (99 words) twice, either in massed fashion or with a 30-min delay between study trials. Similarly, Krug et al. (1990) had participants read a 600-word text twice in massed fashion or with a week delay between study trials. In both studies, performance on an immediate free-recall test was greater after distributed rereading than after massed rereading. In contrast to these earlier results, the present research shows that performance was greater after massed rereading than after distributed rereading on the immediate test. What might account for these differences?

Methodological comparison of the current study with the earlier research may provide one clue. Although the procedure used in the present research was highly similar to that in the study by Krug et al. (1990), the text materials used in the two studies differ in several respects. Of greatest interest, the texts used in the current research were 2.5–3 times as long as the text used by Krug et al. One reason why text length may be important involves the amount of processing devoted to the text during the second study trial in each condition. Relevant to this point, Krug et al. forwarded the "deactivation hypothesis," according to which "full processing of text will occur only on those learning trials in which readers' representations of the text are absent from working memory or 'deactivated' at the onset of the reading episode" (p. 366). With shorter texts, much less time has elapsed and much less material has intervened between the first study trial and a massed restudy trial than with longer texts, so any given piece of text content is more likely to be active (or more easily reactivated) on restudy than with longer texts. Thus, shorter texts may be less likely than longer texts to receive full processing during massed restudy, which would put massed rereading at a disadvantage to distributed rereading for shorter texts more than for longer texts. Indeed, Krug et al. reported a 59% decrease in reading time from first to second study trial with massed rereading (2.7 vs. 1.1 min, respectively), whereas virtually no decrease was observed with distributed rereading (2.6 vs. 2.5 min). By contrast, with the longer texts used in the present research, reading time from the first to second study trial only declined 22% with massed rereading (10.4 vs. 8.1 min, respectively, averaging across experiments). Furthermore, although no decline from the first to second study trial was observed with distributed rereading (9.6 vs. 9.7 min, respectively), the overall amount of time spent studying did not greatly differ for the massed and distributed groups (18.5 vs. 19.3 min, respectively).

Thus, text length may influence the relative amount of time spent studying the text with massed versus distributed rereading. However, the amount of time spent processing text may not be the only difference between short and long texts. In addition to quantitative differences in processing time, qualitative differences in processing of short and long texts may exist, with meaning construction and integration becoming less successful as text length increases. One early study by Rothkopf and Billington (1983) is particularly suggestive. Participants read one of two versions of a text, each containing the same target content but one also containing additional filler sentences to double the length of the text (around 1,400 vs. 3,000 words, respectively). On a subsequent test of memory for the target content, performance was greater after reading the shorter version than after reading the longer version. Additionally, Rothkopf and Billington reported tentative evidence that the difference in performance was relatively consistent across serial position within the texts (e.g., for content that appeared at the beginning of both texts and for content that appeared near the end of both texts). On the basis of this pattern, they suggested that the processing difference for shorter versus longer texts was "general rather than local . . . one possibility is that the integration of related facts may be more difficult for longer than for shorter texts" (Rothkopf & Billington, 1983, p. 680). Although speculative, the possibility that the extent of integration differs during the reading of shorter versus longer texts converges with discussion in the previous section that the degree of integration may be a key factor in explaining the shift in rereading effects with test delay.

Thus, the inconsistency of distributed rereading effects in the present research with those reported in previous research may only be apparent. Furthermore, the present pattern of results is consistent with some findings in the larger literature on practice effects. As Schmidt and Bjork (1992) pointed out, several studies on the spacing of practice have shown that massed practice may yield greater performance than distributed practice when learning is tested immediately, whereas distributed practice yields greater performance than massed practice on retention tests. Taken together, these findings suggest that an important direction for future research on distributed rereading effects will be to identify the conditions under which distributed rereading optimizes learning.

Conclusions

The present research shows that the benefit of rereading depends on the delay between study and test. With texts of moderate length, performance on an immediate test is greater after massed rereading than after a single reading or after distributed rereading, whereas performance on a delayed test is greater after distributed rereading. Although this pattern deviates somewhat from the pattern demonstrated in earlier research, it is important to note that it still supports the same conclusions concerning the schedule of rereading that is most likely to benefit students. The results of previous research would recommend distributed rereading, on the basis of the learning advantage shown for this group over single reading or massed rereading. However, this advantage was demonstrated on an immediate test. Although we did not find an advantage for distributed rereading on an immediate test, we did find an advantage for distributed rereading on a delayed test. Given the argument that students are most likely to face delayed testing conditions in real educational settings, we would still recommend distributed study as the most effective rereading schedule for students.

More broadly, how does rereading compare with other study strategies? Historically, in early strategy research, rereading was often included in experimental designs as a control to equate time on task with other study strategies of interest (e.g., summarization, notetaking, outlining). Most studies found that rereading was as effective as these other study strategies for improving learning, a finding that subsequently fostered an interest in rereading itself. However, the study strategies explored in this earlier research tended to focus on rather passive strategies, rereading included. More recent research suggests that study strategies that actively engage a reader may result in greater gains in learning (for discussion and review, see Nist & Simpson, 2000). Active study strategies may include self-generation of effective questions and explanatory-based answers, reciprocal teaching, self-generated elaborations, and visual organization of important text ideas into concept maps or networks (see Nist & Simpson, 2000). Importantly enough, Nist and Simpson (2000) noted that for students to use these more active study strategies effectively, intensive training may be required across a substantial amount of time, with significant amounts of practice within specific contexts and content domains. The trade-offs between passive and active study strategies are thus obvious: Passive study strategies such as rereading require minimal to no training for students to implement effectively, although they generally lead to smaller learning gains than do active strategies. Active study strategies may lead to more pronounced learning gains, but they usually require extended time and effort with trained instruction and supervision for students to use effectively. Thus, depending on the availability of appropriately structured training programs in active study strategies (e.g., "Learning to Learn" classes), many students may continue to rely on study strategies such as rereading. Additionally, Nist and Simpson argued that there is no one right study strategy—rather, students should optimally be supplied with a toolbox of strategies that have varying levels of effectiveness in different task contexts, content domains, and in various combinations with one another. For example, rereading may provide a relatively time-efficient way to establish basic learning of important facts and ideas within a domain upon which more active, time-consuming strategies can subsequently build. Thus, it is important to further establish both how rereading improves learning and the conditions under which it does so maximally.

References

Amlund, J. T., Kardash, C. A. M., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly, 21,* 49–58.

Barnett, J. E., & Seefeldt, R. W. (1989). Read something once, why read it again? Repetitive reading and recall. *Journal of Reading Behavior, 21,* 351–360.

Bromage, B. K., & Mayer, R. E. (1986). Quantitative and qualitative effects of repetition on learning from technical text. *Journal of Educational Psychology, 78,* 271–278.

Carnes, M. C. (1995). *Past imperfect: History according to the movies.* New York: Holt.

Carrier, L. M. (2003). College students' choices of study strategies. *Perceptual and Motor Skills, 96,* 54–56.

Curtiz, M. (Director). (1936). *The charge of the light brigade* [Motion picture]. United States: Warner Bros.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84,* 795–805.

Dyer, J. W., Riley, J., & Yekovich, F. R. (1979). An analysis of three study skills: Notetaking, summarizing, and rereading. *Journal of Educational Research, 73,* 3–7.

Feldt, R. C., & Ray, M. (1989). Effect of test expectancy on preferred study strategy use and test performance. *Perceptual and Motor Skills, 68,* 1157–1158.

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology, 79,* 198–199.

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology, 84,* 182–192.

Herzog, H., Eliasson, B., & Kaarstad, O. (2000, February). Capturing greenhouse gases. *Scientific American, 282*(2), 72–80.

Howe, M. J. A. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology, 61,* 214–219.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

Krug, D., Davis, B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology, 82,* 366–371.

Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology, 75,* 40–49.

Meyer, B. J., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology, 65,* 109–117.

Murray, J. D., & McGlone, C. (1997). Topic overviews and the processing of topic structure. *Journal of Educational Psychology, 89,* 251–261.

Nist, S. L., & Simpson, M. L. (2000). College studying. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 645–666). Mahwah, NJ: Erlbaum.

Nist, S. L., Simpson, M. L., & Hogrebe, M. C. (1985). The relationship between the use of study strategies and test performance. *Journal of Reading Behavior, 17,* 15–28.

Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28,* 1004–1010.

Rothkopf, E. Z. (1968). Textual constraint as a function of repeated inspection. *Journal of Educational Psychology, 59,* 20–25.

Rothkopf, E. Z., & Billington, M. J. (1983). Passage length and recall with test size held constant: Effects of modality, pacing, and learning set. *Journal of Verbal Learning and Verbal Behavior, 22,* 667–681.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

(*Appendix follows*)

Appendix

Reading Times (in Minutes) for Experiments 1 and 2

Group	Immediate test	Delayed test
Experiment 1		
Single reading	10.7 (0.8)	10.9 (0.5)
Massed rereading, Study Trial 1	11.1 (0.4)	10.6 (0.5)
Massed rereading, Study Trial 2	7.9 (0.5)	8.1 (0.7)
Distributed rereading, Study Trial 1	9.3 (0.4)	10.2 (0.5)
Distributed rereading, Study Trial 2	9.0 (0.4)	10.6 (0.7)
Experiment 2		
Single reading	9.6 (0.5)	9.9 (0.5)
Massed rereading, Study Trial 1	9.9 (0.5)	9.9 (0.4)
Massed rereading, Study Trial 2	7.5 (0.7)	8.8 (0.7)
Distributed rereading, Study Trial 1	9.0 (0.4)	9.9 (0.6)
Distributed rereading, Study Trial 2	9.3 (0.8)	10.0 (0.7)

Note. Standard errors are reported in parentheses.

Received October 16, 2003
Revision received April 14, 2004
Accepted June 25, 2004 ■