

The Positive and Negative Consequences of Multiple-Choice Testing

Henry L. Roediger III
Washington University in St. Louis

Elizabeth J. Marsh
Duke University

Multiple-choice tests are commonly used in educational settings but with unknown effects on students' knowledge. The authors examined the consequences of taking a multiple-choice test on a later general knowledge test in which students were warned not to guess. A large positive testing effect was obtained: Prior testing of facts aided final cued-recall performance. However, prior testing also had negative consequences. Prior reading of a greater number of multiple-choice lures decreased the positive testing effect and increased production of multiple-choice lures as incorrect answers on the final test. Multiple-choice testing may inadvertently lead to the creation of false knowledge.

Keywords: testing effect, negative suggestion effect, false memory, multiple-choice testing

Multiple-choice testing is pervasive in university education. Many large introductory courses in natural and social sciences rely on multiple-choice tests (and to a lesser extent, true–false testing) as the primary means of student assessment. Furthermore, because good multiple-choice tests are so difficult to construct, the same questions are often used across semesters. The result is that the test bank needs to be protected, meaning that many professors neither review the test in class nor return the tests to students. Rather, professors let students review the tests in their offices, although in reality few students do so. Thus, in many classes, students take multiple-choice tests and receive no feedback except for their overall score; many students never review their wrong answers and correct their misconceptions. The experiment we report examines the effects that this practice may have on students' knowledge as displayed on later tests (such as a cumulative final exam) and the knowledge that students may carry away from the course.

There are many advantages to multiple-choice testing. Although difficult to create, they are easy to score and therefore are the evaluation method of choice in large classes. The added benefit is that taking a test generally improves students' performance on a later test; this is referred to as the *testing effect* (Bjork, 1975; Carrier & Pashler, 1992; Glover, 1989; Hogan & Kintsch, 1971; Izawa, 1970; Kuo & Hirshman, 1996; McDaniel & Masson, 1985; Runquist, 1986; Spitzer, 1939; Thompson, Wenger, & Bartling, 1978; Tulving, 1967).

However, there can be negative consequences of relying exclusively on multiple-choice tests. Students expecting a multiple-choice test (relative to an essay test) spend less time studying for the test (Kulhavy, Dyer, & Silver, 1975), and they take notes on

different material than do students expecting an essay exam (Rickards & Friedman, 1978). Work in other traditions within cognitive psychology suggests that there may be negative effects of testing. For example, a multiple-choice test is structurally similar to the misinformation paradigm begun by Loftus and Palmer (1974), a naturalistic variation of the classic A–B, A–D interference paradigm (e.g., McGeoch, 1932; see Crowder, 1976, chapter 8 for a review). In these studies, subjects typically see a simulated accident or crime scene and later receive either a test or a narrative about the scene that contains a few items of misinformation. Producing or choosing misinformation on an initial test has negative consequences for later memory (see Ackil & Zaragoza, 1998; Meade & Roediger, in press; Roediger, Jacoby, & McDermott, 1996; Roediger, Wheeler, & Rajaram, 1993; Schooler, Foster, & Loftus, 1988; Zaragoza, Payment, Ackil, Drivdahl, & Beck, 2001). Like the eyewitness paradigm, both multiple-choice and true–false tests routinely expose students to wrong answers (misinformation). For example, in a standard four-alternative multiple-choice test, three alternatives are wrong, and only one is correct. If subjects read all four statements carefully, then they are exposed to three statements containing wrong answers (misinformation) and one containing the correct answer. Even if subjects pick the correct answer at the time, reading the wrong statements may make them later seem true, by extension from other research. That is, simply repeating statements has been shown in many studies to increase the probability that the statements will be judged true (e.g., Bacon, 1979; Begg, Armour, & Kerr, 1985; Hasher, Goldstein, & Toppino, 1977).

The idea that multiple-choice tests teach misinformation has not gone unnoticed in experimental circles. For example, Toppino and his colleagues have shown that when students take true–false and multiple-choice tests, the lures are later judged as more true than novel false facts (Toppino & Brochin, 1989; Toppino & Luipersbeck, 1993; see also Rees, 1986), although the rated truth of repeated falsehoods never reaches the level of objectively true statements. Remmers and Remmers (1926) coined the phrase “the negative suggestion effect” to refer to students' increased belief in false statements from the tests. Similarly, Brown (1988) and Jacoby and Hollingshead (1990) showed that exposure to misspelled words led subjects to misspell the words later on a traditional oral

Henry L. Roediger III, Department of Psychology, Washington University in St. Louis; Elizabeth J. Marsh, Psychological and Brain Sciences, Duke University.

This research was supported by the Collaborative Activity Award from the James S. McDonnell Foundation. We thank Lisa Fazio for her help in collecting the data, Elaina Pelky for help in scoring the data, and Holli Sink for assistance with article preparation.

Correspondence concerning this article should be addressed to Henry L. Roediger III, Department of Psychology, Washington University, Campus Box 1125, St. Louis, MO 63130-4899. E-mail: roediger@artsci.wustl.edu

spelling test, again showing a negative suggestion effect. Brown, Schilling, and Hockensmith (1999) found that exposing subjects to misinformation after an initial test influenced subjects on the final cued-recall and multiple-choice tests. This outcome occurred even when wrong information was identified as such during the interpolated task. However, performance was unaffected by the number of incorrect alternatives, the number of presentations, and the length of the retention interval.

We examined the possible effects of negative suggestion from multiple-choice tests in situations resembling those of introductory classes outlined in the first paragraph. Students read one half of a collection of educationally relevant materials (nonfiction passages about science and history). They then took a first multiple-choice test (with no feedback) and later a second test, a cued-recall test with a warning against guessing. When subjects were tested on passages they had not read, this corresponded to testing in classes in which students have not done the relevant reading (and also lead to a difficulty-of-question manipulation, because the items were necessarily answered less well in the condition in which the passages had not been read).

On the basis of the literatures of the testing effect and interference effects, we predicted that overall a multiple-choice test (even without feedback) would have a positive effect on a later cued-recall test (the standard testing effect). However, we also expected that the number of lures on the multiple-choice test would negatively affect performance: The more alternatives on the multiple-choice test, the worse performance on the later cued-recall test and the smaller the positive testing effect. More important, we predicted that an increased number of alternatives on the multiple-choice test would also increase errors on the later cued-recall test (similar to a fan effect; Lewis & Anderson, 1976). Furthermore, we examined whether the negative effect of more alternatives would interact with the difficulty of the items, such that increasing the number of alternatives would cause a greater negative effect on later cued recall for difficult items. The reason is that in the literature on retroactive interference, greater interference from A–D learning occurs when A–B learning is less complete (Crowder, 1976; Postman, 1962).

Method

Subjects

A total of 24 undergraduates from Washington University participated in the experiment for partial fulfillment of a course requirement. Subjects were tested either individually or in groups of up to 5 people.

Design

The experiment had a 2 (passage status: read or not) \times 4 (number of alternatives on the multiple-choice test: 0, 2, 4, 6) design. The dependent measures were proportion correct answers and proportion errors on the final cued-recall test.

Materials

We used 36 nonfiction passages from the reading comprehension sections of the Test of English as a Foreign Language and Graduate Record Examination practice test books. These passages spanned a variety of topics, including famous people (e.g., Louis Armstrong), science (e.g., the sun), history (e.g., the founding of New York City), places (e.g., Mount

Rainier) and animals (e.g., sea otters). Passages were only chosen if they allowed the creation of four items, each of which could be presented in all four test formats necessary for the design (2, 4, and 6 alternative multiple-choice, plus cued recall). To create the multiple-choice questions, we generated five plausible lures. Two lures were randomly removed to create the four-alternative version, and two additional lures were randomly removed to create the two-alternative version. Across subjects, the four questions corresponding to each of the passages were rotated through the four multiple-choice conditions (0 [not tested], 2, 4, or 6 alternatives).

The passages were divided into two groups of 18 passages; passages on similar subjects were placed in different groups. One half of the subjects read the first set of passages, and one half of the subjects read the second set of passages; all subjects were tested on the complete set. This feature realized the read–not read aspect of the design.

The multiple-choice test contained 144 questions, 108 of which corresponded to the critical questions. The final cued-recall test contained 216 questions, 144 of which corresponded to the critical questions, and the remainder were fillers. Both tests were given in paper-and-pencil format.

Procedure

The experiment had four phases: passage reading, the multiple-choice test, the visuospatial-reasoning filler task, and the final cued-recall test.

In Phase I, passage reading, subjects read 18 of the 36 passages. The amount of time devoted to each passage was determined in pretesting; on average, subjects were given up to 90 s to read each passage. The goal was for all subjects to finish reading each passage once. Subjects were given a reading recording sheet on which they indicated when they had completed reading the passage; the experimenter monitored the subjects for completion and moved them to the next passage when all had finished reading.

In Phase II, subjects took the multiple-choice test. Prior to the test, they were instructed to provide a response to each and every question, even if they had to guess. They were informed that some of the questions would have two possible answers, some four, and others six. The experimenter instructed them to read each question carefully, being sure to read all possible answers, and then to circle the best answer. They were told to answer the questions in the order in which they appeared, and not to go back and change answers. Subjects were given up to 22 min to answer 144 questions, and the experimenter advised them how time was passing. Subjects who finished early were instructed to turn over their tests and wait quietly for the next set of instructions. No feedback was given as to correctness of answers.

Phase III involved the filler task, the visuospatial brainteasers. Subjects were given 5 min to work on a series of puzzles.

Phase IV involved the cued-recall test. Subjects were strongly warned not to guess just before they took the final cued-recall test. They were told to answer each question only if they were reasonably sure of the answer, and to draw a line through the answer space if they did not know the answer (as opposed to guessing). Subjects were allotted up to 35 min to answer 216 questions, and the experimenter kept them aware of how time was passing. As with the multiple-choice test, subjects who finished early were instructed to turn over their tests and wait quietly for the next set of instructions. At the end of the experiment, subjects were debriefed and thanked for their participation.

Results

All results were significant at the .05 level unless otherwise noted.

Multiple-Choice Test

The multiple-choice test data are shown in Table 1. As expected, more questions were answered correctly when subjects had read

Table 1
Proportion Correct on the Multiple-Choice Test as a Function of Whether Subjects Read the Passages and Number of Alternatives (Including Correct) for Each Question

| Passages | No. of Alternatives | | |
|----------|---------------------|------|-----|
| | Two | Four | Six |
| Read | .90 | .77 | .74 |
| Not read | .72 | .50 | .39 |

Note. The standard error was .015 in the read condition and .013 when the passages were not read.

the relevant passages, $F(1, 23) = 362.68$, $MSE = .01$. Also as expected, subjects answered more questions correctly when tested with fewer alternatives, $F(2, 46) = 102.51$, $MSE = .01$. Reading interacted significantly with number of alternatives, $F(2, 46) = 6.18$, $MSE = .01$. For read passages, subjects correctly answered more two-alternative than four-alternative questions, $t(23) = 3.59$, $SE = .03$, but they performed no worse when tested with six alternatives than with four ($t < 1$). When subjects had not read the relevant passages, performance decreased even further when the number of alternatives was increased from four to six, $t(23) = 3.48$, $SE = .03$. In summary, subjects did worse on the multiple-choice test when tested with more alternatives and with unfamiliar (more difficult) material.

Cued-Recall Test: Correct Answers

Of greatest interest was whether the prior multiple-choice format affected cued-recall performance even when subjects were strictly warned against guessing on the final test. Despite the warning, as shown in the top portion of Table 2, the number of prior multiple-choice alternatives had two different, opposite effects on later recall.

There was a large positive testing effect, $F(1, 23) = 61.52$, $MSE = .01$, and this was larger for read passages, $F(1, 23) = 8.79$, $MSE = .01$. On average, subjects answered 28% of questions correctly for nontested items but were able to answer 46% correctly when the questions had been tested previously.

Not all forms of prior testing were equally beneficial. Prior testing with two alternatives led to 51% correct on the cued-recall test; this dropped to 45% following testing with four alternatives, and 43% following testing with six alternatives. The effect of number of prior alternatives remained significant even after the never-tested items were removed from the analysis, $F(2, 46) = 6.35$, $MSE = .01$, and there was no interaction between reading and number of prior alternatives ($F_s < 1$). That is, performance decreased linearly with number of alternatives previously read for both passages that had been read, $F(1, 23) = 4.05$, $MSE = .01$, $p = .056$, and for those that had not been read, $F(1, 23) = 6.25$, $MSE = .01$. All pairwise comparisons were significant except the difference between having been tested with four ($M = .45$) versus six ($M = .43$) alternatives.

Cued-Recall Test: Production of Multiple-Choice Lures

The critical prediction concerned whether subjects would report lures from the multiple-choice test as correct answers on the

cued-recall test. Even though subjects received a strong warning against guessing, they did report prior multiple-choice lures as answers on the cued-recall test. The relevant data are shown in the bottom portion of Table 2. Production of wrong answers increased linearly with number of previously read alternatives for both read, $F(1, 23) = 7.58$, $MSE = .01$, and nonread passages, $F(1, 23) = 9.96$, $MSE = .01$. Put another way, prior testing increased production of multiple-choice lures on the cued-recall tests, $F(3, 69) = 5.86$, $MSE = .01$, and this effect did not interact with whether the passage had been read. Because reading did not interact with prior number of alternatives, we collapsed over the reading variable to compare performance in each of the prior multiple-choice conditions (0, 2, 4, 6) with each other. Of the six pairwise comparisons, four reached traditional levels of significance. The difference between zero ($M = .05$) and two ($M = .07$) was marginally significant, $t = 1.57$, $SE = .01$, and the difference between four and six was not significant.

Finally, we examined whether the misleading lures produced on the final test were the same or different from the options that subjects had selected on the first (multiple-choice) test. Of the errors produced on the final cued-recall test, 75% had been selected on the initial test. The effect of number of prior multiple-choice alternatives remained significant when the analysis was restricted to errors that had also been made on the earlier multiple-choice test, $F(2, 46) = 4.94$, $MSE = .01$. In contrast, the linear trend disappeared when the analysis was restricted to intrusions of lures that had *not* been chosen on the earlier test ($F_s < 1$). Thus, production of the lures on the final test was largely due to subjects reproducing the incorrect lures that they had chosen on the earlier multiple-choice test. Multiple-choice lures that were read but not selected were not produced on the final test, although perhaps the familiarity accrued to these items would be expressed on other forms of testing, such as true-false.

Cued-Recall Test: Production of Other Errors and Items Left Unanswered

Given that response categories are not independent, it is not surprising that fewer questions were left unanswered when the

Table 2
Performance on the Cued Recall Test as a Function of Passage Reading Status and Number of Alternatives (Including the Correct Answer) on the Prior Multiple-Choice Test

| Passage reading status | No. of previous multiple-choice alternatives | | | |
|------------------------|--|-----|------|-----|
| | Zero (not tested) | Two | Four | Six |
| Proportion correct | | | | |
| Read passages | .40 | .67 | .61 | .61 |
| Not read passages | .16 | .34 | .28 | .26 |
| Proportion lure errors | | | | |
| Read passages | .04 | .06 | .08 | .09 |
| Not read passages | .06 | .09 | .13 | .15 |

Note. For correct responses, the standard error was .022 in the read condition and .015 in the nonread condition. For multiple-choice lure responses, the standard error was .009 in the read condition and .018 in the nonread condition.

relevant passages had been read, $F(1, 23) = 168.75$, $MSE = .03$. Similarly, prior testing reduced unanswered items on the final test, $F(3, 69) = 25.16$, $MSE = .02$. Both reading and prior testing increased correct answers and, thus, reduced items left unanswered.

Subjects did intrude wrong answers other than the multiple-choice lures, albeit at a relatively low rate ($M = .05$), which was expected given the warning against guessing. These extraneous errors were more frequent when subjects had not read the passages ($M = .06$) than when they had ($M = .04$), $F(1, 23) = 13.63$, $MSE = .01$, and they also decreased as a function of the number of alternatives on the initial test, $F(3, 69) = 11.29$, $MSE = .01$. Errors decreased after testing with two ($M = .04$), four ($M = .04$), or six ($M = .03$) alternatives, as compared with the nontested condition ($M = .09$). The drop in extraneous errors in the tested conditions was accompanied by increases in production of both correct and multiple-choice lure answers, as described earlier.

The last analysis considered errors as a function of opportunities to produce errors. That is, as learning increased in the tested conditions, opportunities to make an error were reduced. A 2 (passage status: read or not) \times 2 (number of alternatives: 0, 2, 4, 6) analysis of variance on total errors conditional on the failure to answer correctly revealed a strong tendency for prior testing to increase errors, $F(3, 69) = 3.82$, $MSE = .03$. When subjects did not give the correct answer, they produced a wrong answer on 20% of trials in the nontested condition, as compared with 26% of items previously tested with two alternatives, and 30% of items previously tested with four or six alternatives.

Discussion

Taking a multiple-choice test caused subjects to answer later cued-recall tests with incorrect information. The effect was greater for the nonstudied material and also increased with the number of lures on the multiple-choice test. This negative suggestion effect occurred despite the fact that subjects were strongly warned against guessing. However, there was also a positive effect of testing: Subjects answered more questions correctly on the final cued-recall test when they had been tested previously on a multiple-choice test relative to when they had not. This positive effect of testing occurred even though no feedback was given about student's selections on the multiple-choice test. This outcome confirms that the testing effect frequently observed in list learning experiments (e.g., Tulving, 1967) applies to educationally relevant materials (see Roediger & Karpicke, in press, for further evidence with prose material and free-recall tests). Relatively few prior studies have used such materials (however, see Glover, 1989; McDaniel & Fisher, 1991; Spitzer, 1939), and none used a design involving transfer from multiple-choice to cued-recall tests, as we did.

Prior reading of multiple-choice alternatives may have aided later performance via several different mechanisms. Reading the option on the test serves as an additional study opportunity. As such, it may have reminded subjects of previous knowledge that they would not have been able to retrieve otherwise. Reading the alternatives may also have taught subjects answers by a process of deduction, if they used knowledge about the other options to select the correct answer by a process of elimination. Or reading the

items may simply have increased the familiarity of the correct answer, especially if subjects guessed it was the correct answer.

The most important finding from our study is that the lures on multiple-choice tests can be a source of interference in learning from tests. The greater the number of alternatives provided on the multiple-choice test, the less benefit subjects received from taking the test. Besides reducing correct responding, increasing the number of alternatives also increased the probability that subjects answered cued-recall questions with lures from the prior multiple-choice test. Thus, multiple-choice tests with many lures can actually create false knowledge or beliefs in students—false beliefs that may be carried from the classroom. We are currently investigating the effects of delay and whether negative effects of testing persist if students receive feedback on the multiple-choice responses.

How does a multiple-choice test impair performance on a later test? Not only does reading the multiple-choice question serve as a study trial for the correct answer but it also exposes the subject to one or more incorrect answers, similar to retroactive interference (McGeoch, 1932), a misinformation effect (Loftus & Palmer, 1974), or a fan effect (Lewis & Anderson, 1976). As such, the test may remind subjects of prior wrong beliefs that they otherwise would not retrieve, or it may also teach subjects incorrect answers as they incorrectly use knowledge to eliminate the other answers (including the correct one). Reading the lures also increases the fluency or familiarity of the incorrect answers and may result in the “mere truth” effect from statement repetition (e.g., Hasher et al., 1977). In short, the same mechanisms that may help a subject to correctly produce an answer later on may also lead to production of the incorrect target lures.

In sum, the effect of taking a multiple-choice test has two opposing effects on students' knowledge. Despite an overall positive testing effect, students will sometimes come to believe that the distracter answers are correct and therefore leave the exam having acquired false knowledge. Our experiment and the few prior ones similar to it (e.g., Brown et al., 1999; Toppino & Luipersbeck, 1993) are just the beginnings of research that will determine effects of testing under conditions that are somewhat realistic for educational concerns. We mimicked the situation in which students study (or perhaps fail to study) class material and then take quizzes before a test, or take a series of tests before a final exam. In our opinion, the positive and negative effects of multiple-choice tests uncovered in our experiment deserve wide discussion at all levels of education in which multiple-choice tests are widely used.

References

- Ackil, J. K., & Zaragoza, M. S. (1998). Memorial consequences of forced confabulation: Age differences in susceptibility to false memories. *Developmental Psychology, 34*, 1358–1372.
- Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning & Memory, 5*, 241–252.
- Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioral Science, 17*, 199–214.
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 123–144). New York: Wiley.
- Brown, A. S. (1988). Encountering misspellings and spelling performance:

- Why wrong isn't right. *Journal of Educational Psychology*, 80, 488–494.
- Brown, A. S., Schilling, H. E. H., & Hockensmith, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology*, 91, 756–764.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633–642.
- Crowder, R. G. (1976). *Principles of learning and memory*. Oxford, England: Erlbaum.
- Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392–399.
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16, 107–112.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562–567.
- Izawa, C. (1970). Optimal potentiating effects and forgetting prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, 83, 340–344.
- Jacoby, L. L., & Hollingshead, A. (1990). Reading student essays may be hazardous to your spelling: Effects of reading incorrectly and correctly spelled words. *Canadian Journal of Psychology*, 44, 345–358.
- Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of notetaking and test expectancy on the learning of text material. *Journal of Educational Research*, 68, 363–365.
- Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, 109, 451–464.
- Lewis, C. H., & Anderson, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology*, 8, 311–335.
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 585–589.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, 16, 192–201.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 371–385.
- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, 39, 352–370.
- Meade, M. L., & Roediger, H. L., III. (in press). The effect of forced recall on illusory recollection in younger and older adults. *American Journal of Psychology*.
- Postman, L. (1962). Retention of first-list associations as a function of the conditions of transfer. *Journal of Experimental Psychology*, 64, 380–387.
- Rees, P. J. (1986). Do medical students learn from multiple-choice examinations? *Medical Education*, 20, 123–125.
- Remmers, H. H., & Remmers, E. M. (1926). The negative suggestion effect on true-false examination questions. *Journal of Educational Psychology*, 17, 52–56.
- Rickards, J. P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology*, 3, 136–143.
- Roediger, H. L., III, Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language*, 35, 300–318.
- Roediger, H. L., III, & Karpicke, J. D. (in press). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*.
- Roediger, H. L., III, Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 97–134). San Diego, CA: Academic Press.
- Runquist, W. N. (1986). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, 40, 65–76.
- Schooler, J. W., Foster, R. A., & Loftus, E. F. (1988). Some deleterious consequences of the act of recollection. *Memory & Cognition*, 16, 243–251.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30, 641–656.
- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 210–221.
- Toppino, T. C., & Brochin, H. A. (1989). Learning from tests: The case of true-false examinations. *Journal of Educational Research*, 83, 119–124.
- Toppino, T. C., & Luipersbeck, S. M. (1993). Generality of the negative suggestion effect in objective tests. *Journal of Educational Psychology*, 86, 357–362.
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175–184.
- Zaragoza, M. S., Payment, K. E., Ackil, J. K., Drivdahl, S. B., & Beck, M. (2001). Interviewing witnesses: Forced confabulation and confirmatory feedback increase false memory. *Psychological Science*, 12, 473–477.

Received July 26, 2004

Revision received March 14, 2005

Accepted March 16, 2005 ■