American Psychological Association

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/19/$12.00

2019, Vol. 45, No. 8, 1473–1485
http://dx.doi.org/10.1037/xlm0000638

The Role of Retrieval in Answering Multiple-Choice Questions

Jeri L. Little
California State University, East Bay, and Hillsdale College

Elise A. Frickey
Hillsdale College

Alexandra K. Fung
California State University, East Bay

Taking a test improves memory for that tested information, a finding referred to as the testing effect. Multiple-choice tests tend to produce smaller testing effects than do cued-recall tests, and this result is largely attributed to the different processing that the two formats are assumed to induce. Specifically, it is generally assumed that the multiple-choice format bypasses the need to retrieve information. Research suggests, however, that multiple-choice questions can be constructed to induce retrieval of information pertaining to the incorrect alternatives. In the present research, we investigated the processes that individuals use to answer multiple-choice questions and how those processes relate to later memory, particularly for information pertaining to the incorrect alternatives. Most critically, we found that participants sometimes spontaneously recall information pertaining to incorrect alternatives, and these spontaneous retrievals are associated with retention of those alternatives as correct answers to related questions later. Although multiple-choice questions can be constructed so as to bypass retrieval, they can also be constructed to induce retrieval, and when they are, learning benefits are likely to occur. The present work has practical implications for how instructors can create multiple-choice questions to induce processes that facilitate learning.

Keywords: testing effects, learning, multiple choice, retrieval

Test questions can serve not just as assessments, but also as powerful learning events: answering a test question correctly increases the likelihood of retaining that information (Roediger & Karpicke, 2006)—and can sometimes improve retention of related nontested information. One theory for these findings is that retrieval modifies memories associated with that tested knowledge (Bjork, 1975). But to what extent can multiple-choice tests induce retrieval? The aim of the present paper is to examine how multiple-choice tests can induce retrieval that improves retention.

Tests involving more retrieval are better for learning than are tests relying on less retrieval or on recognition (e.g., Carpenter & DeLosh, 2006; Foos & Fisher, 1988). Multiple-choice tests, at least those assessing fact knowledge, are commonly believed to avoid the need for retrieval, instead relying on recognition. Consistent with that idea, multiple-choice tests generally fare less well than more open-ended tests (e.g., cued-recall) in terms of their ability to promote learning (Anderson & Biddle, 1975; Hamaker, 1986). However, in some circumstances multiple-choice tests are better for learning than are cued-recall tests (Little, Bjork, Bjork, & Angello, 2012), suggesting that processing may depend upon factors other than the format of the question itself.

When asked how they answer multiple-choice questions, test-takers report a variety of strategies (Skakun, Maguire, & Cook, 1994), only some of which circumvent retrieval (e.g., guessing, relying on familiarity). Sometimes test-takers report trying to recall the answer before examining the choices. Additionally, test-takers sometimes report an elimination strategy. That is, especially when they do not immediately know the answer to a multiple-choice question, but they have some knowledge of the topic, they report eliminating choices that they know are wrong (Embretson & Wetzel, 1987; Skakun et al., 1994).

When using an *elimination strategy*, participants will at the least consider—and possibly deeply process—the alternative choices. Critically, this elimination strategy may promote retrieval. For example, Little et al. (2012) postulated that when the alternatives are plausible choices, test-takers may recall information pertaining to those incorrect alternatives in order to reject them. Consistent with this idea, they found improved retention for the previously incorrect competitive alternatives when they later served as answers to related questions. Aiming to provide additional evidence for this *retrieval hypothesis*, Little and Bjork (2015) reasoned that plausible incorrect alternatives would induce retrieval more often

This article was published Online First August 16, 2018.

Jeri L. Little, Department of Psychology, California State University, East Bay, and Department of Psychology, Hillsdale College; Elise A. Frickey, Department of Psychology, Hillsdale College; Alexandra K. Fung, Department of Psychology, California State University, East Bay.

Elise A. Frickey is now at the Department of Psychology, Iowa State University.

Experiment 1 was conducted as a research project sponsored by Hillsdale College's LAUREATES (Laboratory for Advanced Undergraduate Research Education Adapted for Talented and Extraordinary Students) program. We thank Chris Carroll for comments on this paper.

Correspondence concerning this article should be addressed to Jeri L. Little, Department of Psychology, California State University, East Bay, 25800 Carlos Bee Boulevard, Hayward, CA 94542. E-mail: jeri.little@csueastbay.edu

than less plausible incorrect alternatives would because test-takers would not need to recall specific information about less plausible alternatives in order to reject them. Consistent with this reasoning, Little and Bjork (2015) showed that incorrect alternatives were more likely to be recalled as the correct answer to a related question if they had been more plausible answer choices than if they had been less plausible answer choices. That is, when answering multiple-choice questions with plausible incorrect alternatives, participants should sometimes use an elimination strategy that involves the retrieval of information pertaining to the incorrect alternatives.

Although consistent with an elimination strategy involving retrieval, Little and colleagues' evidence for such a strategy is indirect. That is, although the strategies that people report using while answering multiple-choice questions and the retention of related information following the answering of multiple-choice questions have been explored separately, no explicit connection has been made between the strategies and retention. Additionally, the idea that an elimination strategy would be associated with recall of information pertaining to incorrect alternatives is far from certain (e.g., people may eliminate an alternative without explicit recall), and retrieval processes used while answering multiple-choice questions is largely unexplored. In the present paper, our goal was to provide more direct evidence for the *retrieval hypothesis*—that is, that participants sometimes recall information pertaining to the incorrect alternatives and such retrieval is associated with correct recall of those answers to related questions.

In the present experiments, participants took a multiple-choice general knowledge (trivia) test while also explaining what came to mind about the choices (Exps. 1 and 2) or why they made the choice in general (Exp. 2). After a short delay, participants took a cued-recall test with previously tested, related, and nontested (control) questions. We expected to replicate findings that answering multiple-choice questions improves access to previously tested and related information (Little et al., 2012). More critically, however, we wanted to examine what participants recalled during the initial multiple-choice test and how that recall predicted performance on related questions. Specifically, in Experiment 1, we predicted that participants would report information pertaining to the incorrect alternative that would later serve as the answer to the related question and that recalling this information during the multiple-choice test would be associated with correct performance for related information later.

Experiment 1

Method

Participants. Thirty-eight participants (24 males, 14 females) were recruited from Amazon's Mechanical Turk and paid $2.75 for their participation. Participants' age ranged 21 to 60 years of age ($M = 34.5$, $SD = 10.5$). Previous studies with similar materials yielded effect sizes ranging from $d = 0.65$ to 1.05 (see Little, 2018). Given that the design was within-subjects, a sample size of approximately 23 was large enough for 80% power ($d = .85$). The experiment was deemed exempt from institutional review (45 CFR 46.101(b).2).

Materials. Twenty pairs of trivia questions from a variety of topics (e.g., literature, pop culture, mythology) were constructed for this study (see Appendix). For a given question-pair, each multiple-choice version of the question had the same four alternatives, with one of the alternatives being the answer for one question in the pair, and the other alternative being the answer for the other question in the pair. For example, Figure 1 shows a question about a half-man, half-horse creature (Answer: *Centaur*), which included *Minotaur*, *Lytra*, and *Chimera* as incorrect alternatives. The paired multiple-choice question would have the same four alternatives, but *Minotaur* would serve as the correct answer (i.e., to a question about a half-man, half-bull creature). From these multiple-choice questions, cued-recall questions (same stem, but without the alternatives) were created. Several of these question pairs were modified from those used by Little (2018). Each question pair was assigned to one of two groups (A or B).

One alternative for each question pair was a fictitious option created to resemble a plausible answer (e.g., *Lytra*). These fictitious items were included to examine the extent to which participants would (a) recall information pertaining to them during the test and (b) intrude them as answers on the test. Our plausible but fictitious alternatives and questions were drawn from or inspired by the work by Kornell, Hays, and Bjork (2009; see also Berger, Hall, & Bahrick, 1999), who showed that such fictitious informa-

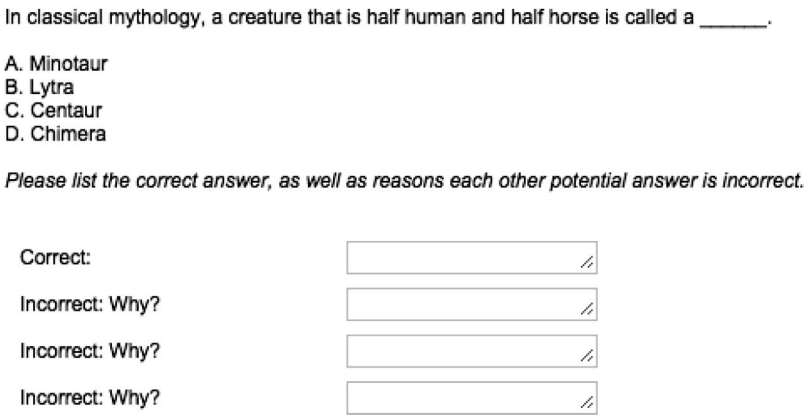

Figure 1. An example of a multiple-choice question, as shown to participants in Experiment 1.

tion was often believed to be plausible, albeit obscure, and could be learned.

Finally, 20 fictitious cued-recall questions (one for each question pair described above) were created (e.g., "*What mythical creature has the face of a lion and body of a snake*?") These fictitious questions appeared only on the final test and enabled us to examine the extent to which participants would incorrectly provide alternatives (fictitious or real) as answers to any question for which they seemed reasonable. These questions served as a control for guessing.

Procedure. The experiment, conducted online via the survey platform Qualtrics, began with brief instructions pertaining to the nature of the experiment. Before beginning, participants were required to agree that they would not look up any answers, and they were told that payment did not depend on their performance.

Before beginning the initial test, participants received an example question pertaining to religious texts (i.e., *For what religion is the Holy Book called the Torah? Catholicism, Islam, or Judaism*), provided with three text boxes. Participants were instructed to type any information that came to mind about the incorrect alternatives when answering the question, as shown in Figure 1, but they were not required to provide responses. After answering the practice question, participants were provided with an example of how they would use the text boxes (i.e., "Correct: Judaism, Incorrect: Why? Catholic = Bible, not Torah, Incorrect: Why? Islam = Koran"). Participants were informed that they would have unlimited time to answer each multiple-choice question, but they would need to use at least 20 s to think about the question. For the initial multiple-choice test, participants answered 10 multiple-choice questions, with one question coming from each of the 10 question-pairs from one of the two sets (A or B).

Following the initial multiple-choice test, participants completed a 2-min visual search distractor task.

Finally, participants took a 60-question cued-recall test. The final test contained 10 previously tested questions, 10 related questions for which a previously incorrect alternative was the correct answer (i.e., the question in a question-pair that had not been tested on the initial test), 10 fictitious questions related to the questions tested on the initial test, and the 30 questions from nontested control topics for these three types of questions, respectively. The first two thirds of the test cycled among related questions, comparable questions from nontested control topics, fictitious questions related to tested questions, and their comparable fictitious control questions. In the last third of the test, participants answered the previously tested questions and comparable control questions. Related and fictitious questions were tested before the previously tested questions because we were primarily interested in the effect of multiple-choice testing on the recall of information pertaining to previously incorrect alternatives, and we wanted to avoid interference that could arise from answering previously tested items first (i.e., Roediger & Schmidt, 1980). Finally, participants answered demographic questions and questions pertaining to their experiment experience including whether they had looked up any answers.

Items were counterbalanced such that they served in the tested and control conditions equally often, and when in the tested condition, items in the question-pairs served as previously tested or related items equally often.

Results and Discussion

First, we describe performance on the initial and final tests. Then, we discuss results pertaining to the primary questions of interest: what participants recalled during the initial test and how it related to their performance on the final test. Finally, we discuss results pertaining to the fictitious alternatives and intrusions.

The data from six participants were omitted from the following analysis because they reported technical difficulties or looking up answers.

Initial test performance. Participants correctly answered 65% (SE = 4%) of the items on the initial test.

Final test performance. In order to assess whether answering multiple-choice questions improved performance for related information, performance for related items which were tested in the first two thirds of the final test were compared with performance for control items which were also tested in the first two thirds of the final test. As shown in Figure 2, and assessed with a paired-samples t test, answers to related questions were recalled with greater accuracy (M = 48%, SE = 4%) than were answers to questions in the control condition (M = 36%, SE = 4%), t (31) = 3.95, $p < .001$, d = 0.70. To assess whether answering multiple-choice questions improved performance for previously tested information, performance for previously tested items which were tested in the last third of the final test were compared with performance for comparable control items which were also tested in the last third of the final test. Previously tested questions were also recalled with greater accuracy (M = 50%, SE = 5%) than were answers to questions in the control condition (M = 32%, SE = 4%), t (31) = 5.30, $p < .001$, d = 0.94.

Information recalled on the initial test and its relation to final test performance. The results above supported our hypothesis that multiple-choice testing improves recall of previously tested and related information on a final cued-recall test. But how was performance for related questions on the final test related to what participants recalled during the initial test?

Scoring. Recall of information pertaining to the multiple-choice test was coded for three attributes. We coded whether participants provided information pertaining to the nonfictitious alternatives that they rejected as the correct answer (both accurate and inaccurate) and whether participants recalled correct informa-

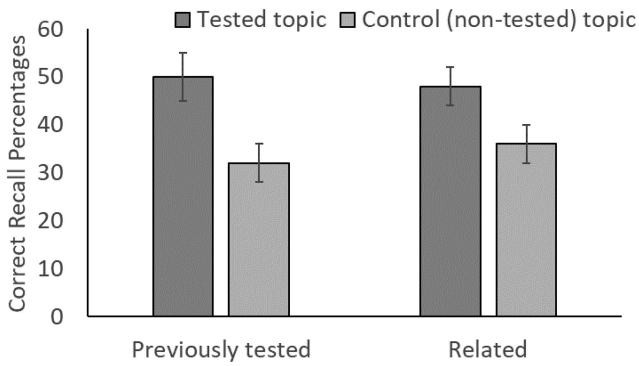

Figure 2. Performance on the final cued-recall test in Experiment 1 for previously tested and related questions from the tested topics and comparable questions from the control condition. Error bars represent ±1 *SE*.

tion about the alternative that would later serve as the answer to the related question (i.e., critical alternative). We also coded whether participants provided information pertaining to fictitious alternatives or chose them as the correct answer. Recall of information pertaining to the alternatives was scored by two raters, and interrater reliability was high ($\kappa = .84$, $p < .001$, agreement = 92%). Scores obtained from one rater were used.

Analyses. Overall, individuals reported information (both accurate and inaccurate) for 46% ($SE = 5\%$) of the incorrect alternatives on the initial test. They recalled accurate information pertaining to critical alternative 32% ($SE = 2\%$) of the time. Often, this information pertained directly to the related question. Pertaining to the fictitious alternatives, participants provided information (e.g., attempting to glean contextual information about the fictitious item, such as "Nefaru = goddess of water?") or occasionally mistook them for correct answers (combined, 11% of the items, $SE = 2\%$), suggesting that many participants believed them to be plausible choices.

How did correct recall relate to performance for related items on the final test? When participants recalled information pertaining to the critical incorrect alternative on the initial multiple-choice test, they answered the related questions correctly 75% ($SE = 6\%$) of the time, which was reliably higher than when they did not recall information pertaining to the incorrect alternative during the initial test, 35% ($SE = 4\%$), $t(28) = 6.20$, $p < .001$, $d = 1.19$.[1]

Intrusions. Participants might remember alternatives from the earlier test and simply provide them as an answer to any question that seems reasonable—that is, make an educated guess—and this strategy could explain a significant amount of the benefit for related information. We intended our fictitious alternatives and questions to be one way to examine this possibility.

First, fictitious alternatives were almost never provided as an answer to a fictitious question ($M = 0.3\%$, $SE = .3\%$). Because the fictitious alternatives were novel, it is possible that participants could not have remembered them even if they wanted to, so we next examined whether participants intruded the answer to the related question as the answer to a fictitious question. Our expectation was that participants might do this, but the benefit in correct performance on related questions should outweigh the increase in intrusions on fictitious questions. Participants were marginally more likely to intrude the critical incorrect alternative (e.g., *Minotaur*, when the answer to the previously tested question had been *Centaur*) as a response to the fictitious questions when they had seen that answer as a multiple-choice alternative ($M = 5\%$, $SE = 2\%$) than when they had not (i.e., in the control condition; $M = 2\%$, $SE = 1\%$), $t(31) = 1.83$, $p = .08$. However, a 2×2 ANOVA confirmed that the benefit that we observed for multiple-choice testing on related questions was larger than the increase in intrusions of those same answers to fictitious items, $F(1, 31) = 7.22$, $p = .01$, $\eta_p^2 = .19$.

Summary and Motivation for Experiment 2

In Experiment 1, we replicated the finding that answering multiple-choice questions can improve later recall of related and previously tested information. More critically, however, in this experiment, participants documented the information that they were recalling about the incorrect alternatives during the multiple-choice test, and they often recalled accurate information about these alternatives that could help them to answer related questions later. Finally, we demonstrated that participants were more likely to correctly answer related questions when they had recalled correct information pertaining to those answers while answering a multiple-choice question than when they had not, suggesting that improved performance was associated with meaningful retrieval earlier.

The procedure in the present experiment may have induced retrieval during the multiple-choice test that would not have occurred naturally. That is, it is possible that the procedure used in Experiment 1 would increase later recall of related information as compared to a procedure that did not explicitly cue participants to recall information pertaining to the incorrect alternatives. Note however, that using similar trivia multiple-choice questions, Little (2018) showed comparable improvements in performance for related information without a "report retrieval" procedure ($d = 1.05$ and $d = 0.65$ in Exps. 1 and 2, respectively).

More critical to our main question of interest, it is possible that if test-takers were not cued to provide information about the incorrect alternatives, but instead queried more generally about their reasoning, they would never report information pertaining to the incorrect alternatives. Thus, in Experiment 2, we introduced a condition in which participants would explain their reasoning without being cued to recall information about the incorrect alternatives (henceforth called the *reasoning* condition), which we compared to a condition using Experiment 1's procedure (henceforth called the *recall* condition). In addition to providing a more naturalistic multiple-choice condition, this condition allowed us to examine more generally what participants self-report about their question-answering processes (e.g., elimination with or without recall of information pertaining to the incorrect alternatives, guessing, familiarity, simply knowing the answer) and how those strategies relate to later performance.

We expected that participants would recall more information about the incorrect alternatives in the recall condition than in the reasoning condition, but we expected that participants would sometimes recall information pertaining to the incorrect alternatives in the reasoning condition, and these recalls would be positively associated with performance on related questions.

We suggest that participants would only sometimes recall information about the incorrect alternatives in the reasoning condition because there are many strategies that people report using when answering multiple-choice questions. We predicted that an elimination strategy would most likely occur when participants could not recall the correct answer but had some knowledge of the topic, and these conditions would only be met some of the time. To examine this prediction, we added the following components to our procedure: Before answering each multiple-choice question in either the recall or reasoning conditions, participants would try to answer a cued-recall version of that question (i.e., the stem without the choices) and would provide a confidence estimate for their ability to correctly answer the question if three choices were then provided. We expected that participants would be most likely to use an elimination strategy and recall information pertaining to the incorrect alternatives if they answered the initial cued-recall ver-

[1] Paired-samples *t*-tests only accounted for participants who provided observations in both of the conditionalized cells.

sion of the question incorrectly, but gave a confidence rating higher than chance.

Experiment 2

Method

Participants and design. Ninety-six participants (50 males, 46 females) were recruited from Amazon's Mechanical Turk and paid $3 for their participation. Ages ranged 21 to 63 years (M = 35.8, SD = 10.3).

The design was a 2 (reporting condition: recall, reasoning) × 3 (item type: previously tested, related, control), with reporting condition manipulated between subjects and item type manipulated within subjects. Given that the present experiment included a condition manipulated between subjects, we more than doubled our sample size from Experiment 1. The experiment was deemed exempt from institutional review (45 CFR 46.101(b).2).

Materials. The materials were the same as those used in Experiment 1, with two modifications. The fictitious alternatives were removed, leaving each question with three rather than four alternatives, and fictitious questions were not used.

Procedure. The procedure was the same as that used in Experiment 1, with the following exceptions. Participants were randomly assigned to recall or reasoning conditions. In the reasoning condition, rather than being told to provide information pertaining to the incorrect alternatives, participants were told that in addition to answering the question, they should briefly explain why they chose the answer they chose. They were provided with two textboxes, one labeled "Correct:" and one labeled "Explain your Reasoning." They received the practice item pertaining to sacred texts used in Experiment 1; after they answered the question, they were told that their reasoning might have included "I just knew it, I completely guessed, I was able to eliminate the other two choices, I knew that Catholicism uses the Bible and Islam uses the Koran, or other responses."

In addition to introducing the reasoning condition, we modified the procedure for answering questions during the initial test. In both the recall and reasoning conditions, for each question in the initial test, participants first attempted to answer a cued-recall version of the question (i.e., same stem, no choices) and then provided a confidence rating estimating their likelihood of answering the question correctly if they were provided with three alternatives. Participants were told that their ratings should range between 33% and 100%. They were given examples of confidence ratings, specifically told that a confidence rating of 33% would be appropriate for a complete guess, 70% would be appropriate if they thought they would probably guess correctly, and 100% would be appropriate if they knew that they would answer the multiple-choice version of the question correctly. Participants had up to 30 s to both provide their answer to the cued-recall question and provide their confidence, but they had to spend at least 10 s before they could move on to the multiple-choice version of the question.

Immediately after answering the cued-recall question and providing a confidence estimate for a given question, participants had up to 45 s to provide an answer to the multiple-choice question as well as to provide additional information about their thought processes, and they had to spend at least 20 s before they could move on to the next cued-recall question.

The final 40-question cued-recall test included 10 previously tested, 10 related, and 20 nontested control questions, and the questions in the final test were randomized with the constraint that one question from a given pair was presented in the first half of the test and the other question was presented in the second half.

Results and Discussion

First, we describe performance on the initial and final tests. Then, we discuss results pertaining to the primary questions of interest: what participants recalled during the initial test or their explanation of reasoning and how it related to their performance on the final test. Finally, we discuss results pertaining to how initial cued-recall performance and confidence predicted what people recalled during the initial test. The data from four participants were omitted from the analyses because they reported technical difficulties, looking up answers, or being distracted.

Modified degrees of freedom and p values are provided when variance between groups was not equal.

Initial test performance. Participants in the recall condition recalled the correct answer to the cued-recall version of the question at a rate (M = 36%, SE = 3%) that was comparable to that of those in the reasoning condition (M = 33%, SE = 4%), $t(90)$ = 0.49, $p > .05$. Participants in the recall condition also answered the subsequent multiple-choice questions correctly at a rate (M = 68%, SE = 3%) that was comparable to that of those in the reasoning condition (M = 66%, SE = 3%), $t(90)$ = 0.54, $p > .05$.

Final test performance. Performance on the final test for previously tested, related, and control items in the recall and reasoning conditions is shown in Figure 3, and as shown there, it appears that both previously tested and related information were recalled better than was information in the control condition, and whether participants were in the recall or reasoning condition did not appear to matter. Indeed, a 2 × 3 mixed-factors ANOVA used to assess the effect of reporting condition (recall, reasoning) and item type (previously tested, related, control) on performance on the final test revealed a main effect of item type, $F(2, 180) = 91.34$, $p < .001$, $\eta_p^2 = .50$, but no main effect of condition and no interaction (F's < 1). Post hoc LSD pairwise comparisons revealed performance for related questions (M = 44%, SE = 2%) to be better than performance for nontested

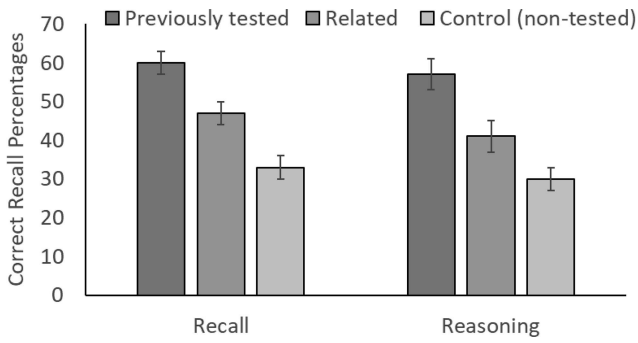

Figure 3. Performance on the final cued-recall test in Experiment 2 for previously tested and related questions from the tested topics and nontested control items as a function of reporting condition (recall or reasoning). Error bars represent ±1 *SE*.

control questions ($M = 32\%$, $SE = 2\%$)[2] and performance for previously tested questions ($M = 58\%$, $SE = 3\%$) to be better than performance for related questions (p's $< .001$). These results suggest that participants did not gain more as a consequence of being instructed to provide information about the incorrect alternatives than being instructed to explain reasoning in general.

Information recalled on the initial test and its relation to final test performance.

Scoring. The scoring procedure was modified from that used in Experiment 1. For both conditions, we coded for responses that indicated elimination (i.e., explicitly referencing "elimination," stating that one or both incorrect alternatives could not be correct, or providing specific information to explain why an alternative was wrong). If participants showed evidence of an elimination strategy, we recorded whether participants recalled information about the critical incorrect alternative. This information did not have to be correct, but it had to be substantial (e.g., simply stating that an answer is not correct was not sufficient). Finally, we coded for whether the participant recalled information that was accurate and directly pertained to the related question. (This was more conservative than our coding in Experiment 1.)

For the reasoning condition, in addition to the coding scheme described above, we also coded for "know," "familiar," and "guess" responses. Coding of "know" occurred when participants stated that they knew the right answer, that they had learned the information in the past, that the answer was the first thing that came to mind, or that they thought that their chosen answer was correct. Coding of "familiar" occurred when participants said that the answer "sounded" or "seemed" correct, when they stated that it was the only choice that they recognized, or when they provided evidence of an educated guess that did not necessarily imply elimination (e.g., "Aphrodite sounds like aphrodisiac, which is related to sex" when the person chose Aphrodite). Finally, coding of "guess" occurred when participants stated that they "completely guessed" or they "had no idea." Occasionally, a response failed to meet any of the categories or met more than one category. For example, "not sure" was not coded into any category. Sometimes participants did not provide any information about their reasoning.

These responses were coded by two raters. Interrater reliability was very high ($\kappa = .88$, $p < .001$, rater agreement = 96%), and disagreements were reconciled by both raters for use in the final analyses.

Analyses. The percentage of responses within each classification are presented in Table 1, as are performance percentages for related and previously tested items conditional upon those classifications.[3]

As shown in Table 1, in both conditions, there appears to be a general increase in one's ability to answer related questions as one recalls more specific details about the incorrect alternatives during the initial multiple-choice test.

In the recall condition, while answering a multiple-choice question, 85% of participants recalled correct information that would later appear in a related question. Conditional upon recalling this specific information pertaining to the critical incorrect alternative, participants answered the related questions correctly 76% ($SE = 6\%$) of the time, which was higher than when they did not recall that information during the initial multiple-choice test ($M = 36\%$, $SE = 3\%$), $t(38) = 6.06$, $p < .001$, $d = .97$.

In the reasoning condition, 22% of participants recalled correct information that would appear in a related question while they were answering a multiple-choice question during the initial test. Conditional upon recalling this specific information pertaining to the critical incorrect alternative, participants answered the related questions correctly 92% ($SE = 5\%$) of the time, which was higher than when they did not recall that information during the initial multiple-choice test ($M = 44\%$, $SE = 6\%$), $t(9) = 8.21$, $p < .001$, $d = 2.64$.

Examining the table, in the reasoning condition, the pattern revealed a general decrease in one's ability to answer the related questions as participants moved from elimination to knowing to familiarity to guessing. For previously tested information, the trend was a bit different: Knowing was associated with the highest recall, with elimination following. This pattern of results for performance for related and previously tested information conditional upon self-reported strategy is consistent with the idea that correct recall of related information is most likely to occur with an elimination strategy, whereas correct recall of previously tested information is most likely to occur if people just report knowing that information. Performance is reasonably high for related information when participants reported knowing the answer to the question on the initial multiple-choice test, but we think that this performance is probably more the consequence of simply knowing this information than having accessibility to this information boosted by the earlier test. For example, if people "just know" that a *Centaur* is a half-man, half-horse creature, the likelihood is high that they would also know that a *Minotaur* is a half-man, half-bull creature.

How initial cued-recall and confidence predicts recall during the initial multiple-choice test. Finally, we had participants answer a cued-recall version of the question and then provide their confidence in their ability to answer a multiple-choice version of that question. We predicted that participants would be most likely to use an elimination strategy or to recall information pertaining to the alternatives during the multiple-choice test when they answered the initial cued-recall version of the question incorrectly, but gave a confidence rating higher than chance (33%). (It may seem strange to compare 33% to everything above 33%, but when participants could not answer the cued-recall question correctly, they provided 33% as their confidence estimate about twice as often as they provided all other confidence estimates combined, suggesting that 33% is a distinct category of confidence.) Looking only at the reasoning condition, which provides the fairest test of this prediction, when participants could not answer the cued-recall version of the question correctly, but gave a confidence estimate greater than 33%, they were numerically (but not reliably) more likely to use an elimination strategy ($M = 26\%$, $SE = 7\%$) than when they correctly recalled the answer to the cued-recall version of

[2] The effect sizes for related items compared to control items were $d = .82$ and $d = .65$ in the recall and reasoning conditions, respectively, which are comparable to those reported in Experiment 1 and by Little (2018).

[3] Means in the table were obtained from aggregated data (i.e., across participants) and should be considered descriptive. They may differ from means represented in the text which, when used for pairwise comparisons, only account for participants who provided observations in both of the conditionalized cells.

Table 1
Percentage of Responses During the Initial Test Based on Classification of Those Responses and Performance Percentages for Related and Previously Tested Items Conditional Upon Those Classifications

Condition	Resp. Classification	% of Resp.	Cond. % Corr. Related	Cond. % Corr. Prev. Tested
Recall	Eliminate	67	56	67
	Recall	39	67	79
	Recall correct	23	80	79
Reasoning	Eliminate	13	63	68
	Recall	4	76	65
	Recall correct	3	86	71
	Know	38	59	79
	Familiar	10	38	46
	Guess	30	17	34

Note. Resp. = Response. % of Resp. = Percentage of responses. Cond. % Corr: Related and Cond. % Corr: Prev. Tested represent percent correct performance conditional upon the given classification for related and previously tested information, respectively. "Recall" refers to instances in which participants recalled information about critical alternative. "Recall correct" refers to instances in which participants recalled information about the critical alternative that would serve as the answer to the related question. "Recall" is a subset of "Eliminate," and "Recall correct" is a subset of "Recall."

the question or answered incorrectly but provided a confidence estimate of 33% ($M = 16\%$, $SE = 4\%$), $t(33) = 1.44$, $p > .05$.

When participants could not answer the cued-recall version of the question correctly, but gave a confidence estimate greater than 33%, they were numerically (but not reliably) more likely to recall information pertaining to the incorrect alternatives ($M = 7\%$, $SE = 7\%$) than when they correctly recalled the answer to the cued-recall version of the question or answered incorrectly but provided a confidence estimate of 33% ($M = 5\%$, $SE = 3\%$), $t(33) = .44$, $p > .05$.

Summary. The present experiment replicated the findings from Experiment 1 and showed that the benefits for retention of related information could not be attributed to being explicitly prompted to recall information about incorrect alternatives. That is, even when not prompted to do so, participants sometimes reported the recall of information pertaining to incorrect alternatives; and recall of critical information during this task predicted later performance on related questions for which the alternatives were the answers.

General Discussion

People commonly assume that multiple-choice questions avoid the need for retrieval, instead relying on more shallow processing (e.g., recognition, familiarity, guessing). Our results clearly show, however, that although participants can answer multiple-choice questions by guessing or relying on familiarity, they can also rely on elimination strategies that include the recall of information pertaining to incorrect alternatives. The recall of such information during the multiple-choice test predicts one's later ability to answer related questions. What strategy is used probably depends more on the knowledge of the test-taker than on the format per se. This set of experiments provides direct evidence for the *retrieval hypothesis*.

Recalling Information Pertaining to the Incorrect Alternatives

When participants were not explicitly told to report information that came to mind about other alternatives, they reported recalling information about other alternatives 4% of the time; 3% of the time it was correct. These rates should be considered in light of baseline performance on the initial cued-recall versions of the questions. Cued-recall questions explicitly prompted retrieval of specific information, and correct performance was 33–36%. Recall of information pertaining to the incorrect alternatives was not explicitly prompted, so it makes sense that recall of this information would occur much less frequently. From the performance rates provided above, one could reason that when participants *could* recall information about the incorrect alternatives, they did so and reported it 8–9% of the time—without being prompted to do so.

The differences in rates of recalling information pertaining to the incorrect alternatives between the recall and reasoning conditions raises a potential concern. A main point of the present paper is that retrieval during the multiple-choice test underlies successful performance on related questions later, so why is performance on related questions so similar in the recall versus reasoning conditions when recall of correct information during the initial multiple-choice test is much higher in the recall versus reasoning condition? For one, we only know what came to mind through participants' self-reports, and what people self-report relies upon introspection and is affected by demand characteristics. Pertaining to introspection, participants may activate memories without conscious awareness (e.g., spreading activation, Collins & Loftus, 1975; see also Carpenter, 2011). Pertaining to demand characteristics, in the recall condition, the instructions probably prompted participants to recall information about the incorrect alternatives, and they did so for the critical alternative 39% of the time. In the reasoning condition, the demand characteristics were much lower (indeed, this is one reason that we included this condition in Experiment 2), and unsurprisingly, so are their reports of recall during the

multiple-choice test. However, they may have not thought to include such recall as part of their reasoning. In the reasoning condition, participants provided comments suggesting an elimination strategy 13% of the time (e.g., "Helsinki isn't in Norway"), and they may have brought more specific information to mind but neglected to report it.

In Experiment 2, we also explored the conditions under which participants would be likely to use an elimination strategy and recall information pertaining to the incorrect alternatives. Our hypothesis was that an elimination strategy and recall of information pertaining to the incorrect alternatives would be most likely to occur when participants could not recall the answer to a cued-recall version of the question but believed they had some knowledge of the topic. Although our results showed a numerical difference in the predicted direction, these differences were not reliable. An elimination strategy and recall of information pertaining to the incorrect alternatives was relatively rare in the reasoning condition, but perhaps under different conditions (e.g., many more items per participant), researchers could gain a better understanding of how accessibility and confidence interact to influence strategies on an initial test.

One might be concerned about how the addition of the cued-recall and confidence estimate aspects of the procedure affected performance. Importantly, performance for related information did not differ in the recall conditions in Experiments 1 and 2. However, the benefit for previously tested information appeared to be larger in Experiment 2 than in Experiment 1, suggesting that the cued-recall and confidence additions may have directed attention to the information that was explicitly tested. Thus, we believe that the new aspects of the procedure probably did not increase the likelihood of participants examining the alternatives or reporting retrieval.

How Multiple-Choice Tests Improve Recall of Related Information

In the present experiments, participants did not study anything before taking the initial test, which deviates from much prior work on the benefits of testing. However, answering multiple-choice questions has been shown to improve memory for tested content even without a study session (Butler & Roediger, 2008). Furthermore, Cantor, Eslick, Marsh, Bjork, and Bjork (2015) found that answering multiple-choice questions could help in the recovery and stabilization of information that participants were not able to recall before answering a multiple-choice question (i.e., *marginal knowledge*; Berger et al., 1999). Although they only considered recovery of tested information, the present experiments support the notion that multiple-choice tests can also aid in the recovery of related information. That is, perhaps exposing participants to the alternatives on the test jogs their memory for information pertaining to those alternatives, making the alternatives more accessible when they serve as answers to questions later. A question, then, is whether participants would need to answer a trivia question to obtain this benefit or whether that information would be strengthened simply as a consequence of exposure to those alternatives.

Using similar trivia materials, Little (2018) showed that increased performance on related items only occurred when participants answered trivia questions—not in a condition that controlled for exposure to the alternatives. In her Experiment 2, a trivia condition (like the ones we used in the present experiments) was compared to an exposure-control condition in which participants were exposed to the same alternatives, but instead of answering a trivia question, participants answered a question that focused on other aspects of the alternatives (e.g., *Which of the following is the shortest word? Oslo, Helsinki, Stockholm, Leningrad*). A limitation of this experiment was that condition (trivia vs. exposure-control) was confounded with depth of processing because the trivia condition likely induced deeper processing than did the exposure-control condition (Craik & Tulving, 1975). Although this seems problematic, it is worth noting that much of the literature on testing effects has compared testing to restudy, which although controlling for time on task, also confounds the manipulation with depth of processing. That is, the processes associated with testing are likely deeper than the processes associated with restudy.

Implications for Use of Multiple-Choice Questions in Educational Contexts

The procedures used have additional implications for how we are to interpret and apply these findings. Multiple-choice tests in the educational setting rarely, if at all, require that individuals write their reasons for rejecting each alternative or even have them explain their reasoning, as we instructed our participants to do. One could then argue that this procedure qualifies the findings; that is, the results apply only to a modified form of multiple-choice testing that instructs, as we did, individuals to reflect on reasons for rejecting each alternative. Note, however, that these results replicate findings using similar materials but without any instructions to record information about one's thought processes (Little, 2018). Additionally, we wager that when students are motivated to ascertain the correct answers to questions, as they are in educational settings, they are likely to be more careful to consider their rejection of incorrect alternatives than laboratory or online study participants would be.

The results of the present studies have additional practical implications. As previously noted, we tested (via cued-recall questions on the final test) only one alternative for each multiple-choice question on the initial test. But during that initial multiple-choice test, individuals recalled information pertaining to other alternatives as well. It is reasonable to assume, then, that the results for retention of related information would also be present for questions pertaining to other alternatives—to the extent that individuals brought to mind related information pertaining to them in the initial multiple-choice test; and there is some evidence that they did. For example, when answering the question about *Minotaurs* in Experiment 1, 68% of participants who answered that question were able to recall some accurate information about *Chimeras* on the initial multiple-choice test. Because recall of related information on a multiple-choice question predicted performance on the related question on the cued-recall test, we can assume that many participants would have been able to answer a question about *Chimeras* later. That is, when multiple incorrect alternatives are competitive, the benefit of answering multiple-choice questions should extend to information pertaining to more than a single alternative.

Discussion of this point leads to the question of whether other related information not directly tied to the alternatives would be strengthened. Although the present experiment did not address this

issue, other research suggests that for related information to be strengthened, it has to be related to tested information in specific ways. For example, asking cued-recall questions can impair recall of related information (Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan, 2009; Little et al., 2012; Little, Storm, & Bjork, 2011) or it can improve it (Chan, McDermott, & Roediger, 2006). Whether testing helps or hurts retention of related information seems to rely on a variety of factors including: the nature of materials (Chan, 2009; Little et al., 2011), delay between the initial and final tests (Chan, 2009), the relationship between the tested and related information (Little et al., 2011), and expertise (Carroll et al., 2007).

Does this mean that our materials are too contrived to be generally meaningful? We argue no. In educational contexts in a variety of domains, learners are exposed to large amounts of confusable information (often facts), and instructors cannot or do not test every fact or concept. We argue here that multiple-choice questions can be created to pack several related concepts or facts into one question. If one provides learners with multiple-choice questions containing plausible alternatives, they may strengthen learners' access not just to the information directly being tested, but also to related information. It is possible that instructing students to answer questions by thinking about why the alternatives are wrong would also lead to benefits. Although the questions used in the present experiments are more straightforward and fact-based than questions often used in educational contexts, straightforward fact-based questions are used in many domains, especially when students are first learning terms and concepts.

Finally, would these effects persist over a delay of more than a few minutes? Following the study of passages, Little and Bjork (2012) showed that taking a multiple-choice test resulted a benefit for related information that persisted over a delay of 48 hr, so we have reason to believe that the effects would persist here as well, but this is an avenue for future research.

Conclusions

Although not typically associated with retrieval processes, multiple-choice questions can induce such processes, and these processes can be good for learning. When constructed with plausible alternatives, multiple-choice questions encourage an elimination strategy that induces retrieval. Especially when used as a tool to promote learning, multiple-choice tests can be used to induce students to think broadly, even bringing to mind information that was not directly tested by the question.

References

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9, pp. 89–132). New York, NY: Academic Press. http://dx.doi.org/10.1016/S0079-7421(08)60269-8

Berger, S. A., Hall, L. K., & Bahrick, H. P. (1999). Stabilizing access to marginal and submarginal knowledge. *Journal of Experimental Psychology: Applied, 5,* 438–447. http://dx.doi.org/10.1037/1076-898X.5.4.438

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Butler, A. C., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36,* 604–616. http://dx.doi.org/10.3758/MC.36.3.604

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition, 43,* 193–205. http://dx.doi.org/10.3758/s13421-014-0462-6

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. http://dx.doi.org/10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268–276. http://dx.doi.org/10.3758/BF03193405

Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology, 19,* 580–606. http://dx.doi.org/10.1080/09541440701326071

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61,* 153–170. http://dx.doi.org/10.1016/j.jml.2009.04.004

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571. http://dx.doi.org/10.1037/0096-3445.135.4.553

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428. http://dx.doi.org/10.1037/0033-295X.82.6.407

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. http://dx.doi.org/10.1037/0096-3445.104.3.268

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement, 11,* 175–193. http://dx.doi.org/10.1177/014662168701100207

Foos, P. W., & Fisher, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology, 80,* 179–183. http://dx.doi.org/10.1037/0022-0663.80.2.179

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56,* 212–242. http://dx.doi.org/10.3102/00346543056002212

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. http://dx.doi.org/10.1037/a0015729

Little, J. L. (2018). The role of multiple-choice tests in increasing access to difficult-to-retrieve information. *Journal of Cognitive Psychology*. Advance online publication. http://dx.doi.org/10.1080/20445911.2018.1492581

Little, J. L., & Bjork, E. L. (2012). The persisting benefits of using multiple-choice tests as learning events. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 683–688). Austin, TX: Cognitive Science Society.

Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition, 43,* 14–26. http://dx.doi.org/10.3758/s13421-014-0452-8

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science, 23,* 1337–1344. http://dx.doi.org/10.1177/0956797612443370

Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory, 19,* 346–359. http://dx.doi.org/10.1080/09658211.2011.569725

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., III, & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 91–105. http://dx.doi.org/10.1037/0278-7393.6.1.91

Skakun, E. N., Maguire, T. O., & Cook, D. A. (1994). Strategy choices in multiple-choice items. *Academic Medicine, 69*(Suppl.), S7–S9. http://dx.doi.org/10.1097/00001888-199410000-00025

Appendix

Materials Used in Experiments 1 and 2

Following are the 20 question pairs used in Experiments 1 and 2. The alternatives used in the initial multiple-choice questions are provided first (Experiment 2 did not use fictitious alternatives). The alternatives are followed by the two general knowledge questions in a given pair, one of which would be tested with the alternatives on the initial test and serve as the previously tested item on the later test, and the other which would serve as the related question on the final test. Finally, the third question in each group is a fictitious related question used in Experiment 1.

Minotaur, Lytra, Centaur, Chimera
In classical mythology, a creature that is half man and half bull is called a _____.
Answer: Minotaur
In classical mythology, a creature that is half human and half horse is called a _____.
Answer: Centaur
Fictitious question: What mythical creature has the face of a lion and the body of a snake?
Answer: Lytra

Rhumous, Nimbus, Cirrus, Cumulus
What is the term for large, white, puffy clouds that generally appear in fair weather, but that can also form thunderheads on hot days?
Answer: Cumulus
What is the term for lacy or wispy clouds that form at high altitudes, often before a change in the weather?
Answer: Cirrus
What cloud type is found exclusively around mountaintops?
Answer: Rhumous

Pulsca, Arthropods, Mollusks, Echinoderms
What is the name of the phylum within the animal kingdom that includes shrimp, centipedes, insects, and spiders?
Answer: Arthropods
What is the name of the phylum of invertebrates that contains snails, octopus, and squid?
Answer: Mollusks
What phylum contains the species Hemipneustia and Unocronda?
Answer: Pulsca

Sunset Boulevard, Casablanca, Courteous Rascals, Gone with the Wind
From what classic movie comes the line, "Here's looking at you, kid"?
Answer: *Casablanca*
From what classic movie comes the line, "Frankly my dear, I don't give a damn"?
Answer: *Gone with the Wind*
In what movie, set in the 1940s, do two unruly journalists from Scotland woo and wed two foreign princesses?
Answer: *Courteous Rascals*

Athena, Aphrodite, Venus, Nefaru
The Roman goddess of love is _____.
Answer: Venus
The Greek goddess of love is _____.
Answer: Aphrodite
Who is the African Goddess of love?
Answer: Nefaru

(*Appendix continues*)

s, cpl, kg, mol
The abbreviation for the SI base unit for time is _____.
Answer: s
The abbreviation for the SI base unit for amount of substance is _____.
Answer: mol
What is the SI base unit for frequency?
Answer: cpl

Ronny Triton, Archie Griffin, Jay Beranger, Tim Tebow
Who was the only college football player to win the Heisman Trophy twice?
Answer: Archie Griffin
Who was the first college football player to win the Heisman Trophy as a sophomore?
Answer: Tim Tebow
What baseball player in 1991 proposed to his girlfriend before the last pitch of the game?
Answer: Ronny Triton

Euclid, Lourdeis, Aristotle, Archimedes
Who was the ancient Greek mathematician that is considered the "Father of Geometry"?
Answer: Euclid
Who was the ancient Greek mathematician, scientist, and inventor best known for his investigations of buoyancy?
Answer: Archimedes
What ancient scientist theorized about "motion-based force" of the earth?
Answer: Lourdeis

Magemun, Femur, Tibia, Humerus
What is the name of the large bone in the upper arm?
Answer: Humerus
What is the name of the large bone in the upper leg?
Answer: Femur
What bone in your pelvis is the size of a pea?
Answer: Magemun

"Hey Jude," "Bridge Over Troubled Water," "Getting to Somewhere," "Imagine"
In which of his hit singles does John Lennon sing of a world at peace and free of religious and national boundaries?
Answer: "Imagine"
Which 7-min-long Beatles song was written to comfort a child after his parents' divorce?
Answer: "Hey Jude"
What '60s song by an English boy-band proposed that we all wander the earth aimlessly, with no purpose other than to exist?
Answer: "Getting to Somewhere"

William Tell, Sir Galahad, Robin Hood, Sir Alfred
In the tales of King Arthur, who was the young knight whose exceptional purity and virtue enabled him to see the Holy Grail in all its splendor, while many other knights who sought it could not see it at all?
Answer: Sir Galahad
Who was the legendary hero who, famous for his skill as an archer, was forced to shoot an apple off of his own son's head?
Answer: William Tell
What knight betrayed his king by fighting for his enemies in the bloody Battle of Tristony?
Answer: Sir Alfred

Aldous Huxley, Ralph Hekinburg, Kurt Vonnegut, George Orwell
Brave New World is a novel written by _____.
Answer: Aldous Huxley
Animal Farm is a novel written by _____.
Answer: George Orwell
Who wrote the propagandist novel *Fickle Freedoms,* which described the advantage of Communism over destructive capitalist nations?
Answer: Ralph Hekinburg

(*Appendix continues*)

Edmond Cowley, James Madison, John Adams, Andrew Jackson
Who was the second president of the United States?
Answer: John Adams
Who was the fourth president of the United States?
Answer: James Madison
The actions of what 20th-century American president strained public relations with Wales?
Answer: Edmond Cowley

ergo, al etre, et cetera, ad nauseam
The Latin translation of "therefore" is _____.
Answer: ergo
The Latin translation of "and so forth" is _____.
Answer: et cetera
Finish the Latin phrase: "nic ubic _____"
Answer: al etre

Oslo, Helsinki, Doukland, Leningrad
What is the capital of Finland?
Answer: Helsinki
What is the capital of Norway?
Answer: Oslo
In what Estonian city does the famed Parade of Nations take place each year?
Answer: Doukland

Ricardo Huyente, Francisco Goya, Pablo Picasso, Salvador Dalí
_____ was a Spanish painter of the 20th century, well known for cubism and his painting *Guernica.*
Answer: Pablo Picasso
_____ was a Spanish surrealist painter of the 20th century, known for his iconic use of melting clocks.
Answer: Salvador Dalí
What Peruvian painter is known for his masterpiece, *El Buhonero*?
Answer: Ricardo Huyente

Crime and Punishment, Conquering the Frye, The Brothers Karamazov, Anna Karenina
____ is a novel by Feodor Dostoevsky in which the plot concerns the trial of one of four brothers for the murder of his father.
Answer: *The Brothers Karamazov*
_____ is a novel by Leo Tolstoy in which a woman enters a tragic adulterous affair and commits suicide by throwing herself under a train.
Answer: *Anna Karenina*
What German novel told the story of a Jewish Nazi's complex experience in the time of Hitler?
Answer: *Conquering the Frye*

Arigato, Sayonara, Cui Chen, Obrigado
How do you say "good-bye" in Japanese?
Answer: Sayonara
How do you say "thank you" in Japanese?
Answer: Arigato
How do you say hello in the Xiajen dialect of Mandarin Chinese?
Answer: Cui Chen

(*Appendix continues*)

Entropy, Ventration, Diffusion, Distillation
In chemistry, the separating of the constituents of a liquid by boiling it and then condensing the vapor that results is called_____.
Answer: Distillation
The spreading of atoms or molecules of one substance through those of another, especially into liquids and gases, is known as _____.
Answer: Diffusion
The recent discovery of the process by which air is cycled rapidly to manipulate oxygen levels is called _____.
Answer: Ventration

The Canterbury Tales, The Arabian Nights, Aesop's Fables, *Tales of the Lovely Teller*
What collection of stories from the 14th century recounts tales about a group of pilgrims who meet at an inn near London?
Answer: *The Canterbury Tales*
What collection of stories recounts tales that, supposedly, queen Scheherazade told her husband?
Answer: *The Arabian Nights*
The novel by Gordon Peters in which an elderly woman tells fantastical love stories to her grandchild is called _____.
Answer: *Tales of the Lovely Teller*

Received June 20, 2017
Revision received April 16, 2018
Accepted May 4, 2018 ■