Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

The role of multiple-choice tests in increasing access to difficult-to-retrieve information

Jeri L. Little

To cite this article: Jeri L. Little (2018): The role of multiple-choice tests in increasing access to difficult-to-retrieve information, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2018.1492581

To link to this article: https://doi.org/10.1080/20445911.2018.1492581

JOURNAL OF COGNITIVE PSYCHOLOGY
https://doi.org/10.1080/20445911.2018.1492581

The role of multiple-choice tests in increasing access to difficult-to-retrieve information

Jeri L. Little[a,b,c]

[a]Department of Psychology, California State University, East Bay, Hayward, CA, USA; [b]Department of Psychology, Hillsdale College, Hillsdale, MI, USA; [c]Department of Psychology, Washington University in St. Louis, Louis, MO, USA

ABSTRACT

Answering multiple-choice questions improves access to otherwise difficult-to-retrieve knowledge tested by those questions. Here, I examine whether multiple-choice questions can also improve accessibility to related knowledge that is not explicitly tested. In two experiments, participants first answered challenging general knowledge (trivia) multiple-choice questions containing competitive incorrect alternatives and then took a final cued-recall test with those previously tested questions and new related questions for which a previously incorrect answer was the correct answer. In Experiment 1, participants correctly answered related questions more often and faster when they had taken a multiple-choice test than when they had not. In Experiment 2, I showed that the more accurate and faster responses were not simply a result of previous exposure to those alternatives. These findings have practical implications for potential benefits of multiple-choice testing and implications for the processes that occur when individuals answer multiple-choice questions.

ARTICLE HISTORY
Received 19 January 2018
Accepted 17 June 2018

KEYWORDS
Testing effects; learning; multiple-choice; retrieval-practice; marginal knowledge

Most students breathe a sigh of relief when told that their exam will be multiple-choice. It is no surprise that if allowed to choose between recalling the capital of country (i.e. answering a cued-recall question) and selecting the answer from among several alternatives (i.e. answering a multiple-choice question), students will choose the multiple-choice version of the question. Having choices often makes correctly answering the question easier. An underappreciated effect of test-taking, however, is that people can learn from answering test questions (e.g. Roediger & Karpicke, 2006), and what people learn may depend upon the format of the question.

Multiple-choice testing can improve accessibility to information that might otherwise be too inaccessible to recall. For instance, Cantor, Eslick, Marsh, Bjork, and Bjork (2015) investigated the effect of multiple-choice testing as a way to assist in the recovery and stabilisation of *marginal knowledge* (i.e. information that a person has previously learned but may not be able to recall given a cued-recall question or fill-in-the-blank prompt; see also Berger, Hall, & Bahrick, 1999). For example, one may not be able to recall *petrification* to the question: *What is the long process by which a dead organism turns to stone?* Even if one had known that fact at some point in the past. However, given the question with the choices: *decomposition, ossification, rigour mortis,* and *petrification* as possible answers, it is much less difficult to answer the question correctly because correct responding does not depend upon retrieval of the answer. Pertaining directly to recovery and stability, providing individuals with a question in a multiple-choice – rather than a cued-recall – format not only increases their likelihood of answering the question correctly on *that* test, but also increases their likelihood of providing the correct answer to the cued-recall version of the question later – that is, when retrieval is required (Cantor et al., 2015). Participants can often increase access to correct answers more from taking multiple-choice tests than from taking cued-recall tests (assuming feedback is not provided). But is that learning restricted to the question and its answer? In the present paper, I examine the extent to which answering multiple-choice questions can increase accessibility to difficult-to-recall

CONTACT Jeri L. Little jerilittle@gmail.com
Supplemental data for this article can be accessed here https://doi.org/10.1080/20445911.2018.1492581.

information that is not itself tested but instead is related to those tested questions.

Related to this question, Little, Bjork, Bjork, and Angello (2012) showed that answering multiple-choice questions can improve retention of information *not* directly tested by the question – when answers to those related questions served as competitive incorrect alternatives on an initial multiple-choice test. They had participants study passages and then answer multiple-choice questions or cued-recall questions pertaining to those passages. They found that answering multiple-choice questions improved retention of related information, whereas answering cued-recall questions did not. Additionally, Little and Bjork (2015) showed that the incorrect alternatives had to be plausible competitive choices to obtain this benefit. When the incorrect alternatives were not competitive, there was no benefit for the later recall of related information. That is, simply exposing test-takers to the alternatives was not enough to increase their accessibility in a manner that would enhance performance on a later test. Indeed, it is likely that competitive alternatives induce deeper processing than do non-competitive alternatives. Further, Little and Bjork (2015) suggest that sometimes participants may spontaneously recall information pertaining to the incorrect alternatives, and these spontaneous retrievals lead to the improved retention of related information.

The finding that multiple-choice testing with competitive alternatives improves recall of those alternatives as correct answers to related questions has thus far only been shown when the initial multiple-choice test follows the study of information for which the questions (including the alternatives) are relevant. Following from the work by Cantor et al., it makes sense that an analogous recall benefit would emerge with a multiple-choice test that was not preceded by a study phase but instead based upon general but difficult-to-recall knowledge. For instance, when *petrification* is accompanied by competitive alternatives (e.g. *rigour mortis,* and *decomposition*, all of which are terms referring to processes that lead to changes in organisms following their death), it is likely that individuals would – at least some of the time – think deeply about these alternatives in order to reject them as answers to the *petrification* question. If this is the case, it follows that on a later test for which a question about *rigour mortis* is asked (e.g. *What process causes limbs of a corpse to stiffen after death?),* individuals would be more likely to answer that question correctly than they would had they never answered the multiple-choice question about petrification.

That the findings documented by Little and colleagues would expand to trivia questions provided without a preceding study session is not a foregone conclusion, however. Critically, Little and colleagues always provided multiple-choice tests immediately following the reading of passages for which those questions would be relevant. Indeed, all of the incorrect alternatives and relevant information pertaining to them were provided in the passages. Additionally, these readings often contained information that participants did not already know. These differences in methodology are important to consider. First, if participants recognise the alternatives from prior study of a passage, they may process those alternatives differently. Perhaps they would be more likely to think about information pertaining to them. Second, it is possible that new learning would benefit from multiple-choice questions in a way that differs from earlier learned material. Finally, the fact that tested and related information was presented in a coherent fashion in prose passages might have made the relationship between them clear, and this may have led to the benefits in retention of non-tested related information. In fact, study in the form of a coherent text, as compared to randomised facts, has been shown to promote the positive effects of initial testing (with cued-recall questions) on later recall of related information (see Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan, 2009; Little, Storm, & Bjork, 2011). Both Chan (2009) and Little et al. (2011) found that when that same information was presented as a set of random facts (i.e. sentences from the prose passage were randomised) rather than in a coherent prose format, no facilitation for the recall of previously nontested related information occurred; and with a short delay between the initial and final test, impairment occurred instead. Although the research discussed above used cued-recall questions, one might speculate that the patterns shown by Little and colleagues – in terms of multiple-choice questions facilitating recall of related information – might have also depended upon the related information being provided in a specific context like a coherent prose passage. Thus, without control over the coherence and integration of the tested and related information, as well as what participants know about the questions or

alternatives in general, benefits in later recall of related information may not occur when using trivia questions without a prior study session.

For the reasons presented above, the present studies have clear methodological differences from those reported earlier, and it is possible that the outcomes would be different. In the present paper, the main goal was to assess the extent to which answering trivia questions increases accessibility to previously incorrect alternatives when they are later queried by related questions. Accessibility was measured in two ways: what people recall and how fast they recall it. Experiment 2 replicated and extended the research to gain some leverage on the mechanisms influencing any observed benefit to retention of related information. To foreshadow, multiple-choice testing did improve accessibility of related information, and Experiment 2 demonstrated that this benefit could not be attributed to exposure to the correct answers as incorrect alternatives in the previous multiple-choice test.

Experiment 1

Method

Participants and design

Thirty-five students at Washington University in St. Louis participated for partial credit in a psychology course. Prior research indicated that this sample size would likely be sufficient to observe a benefit for related information. Participants were tested one at a time on individual computers. The experiment utilised a one-way within-subjects design (topic: tested, nontested) with two types of questions for the tested topic: previously tested and related. Two dependent variables of interest were accuracy and response times.

Materials

Twenty-eight pairs of trivia questions from a variety of topics (e.g. mythology, sports, literature, science, history) were created for this study (56 questions total). Question pairs were created such that the same four alternatives would be plausible choices for the two questions in a given pair, but each question in the pair would have a different alternative as the correct answer. For example, the answer to one question in a pair about mythical creatures (i.e. *In classical mythology, a creature that is half human and half horse is called a _____. Minotaur, Saytr, Centaur, Chimera*) is Centaur, and the other (i.e. *In classical mythology, a creature that is half human and half bull is called a _____. Minotaur, Saytr, Centaur, Chimera*) is Minotaur. Questions were provided with alternatives on the initial multiple-choice test, but on the final cued-recall test, only the stems were provided. Several additional examples of questions pairs used in Experiments 1 and 2 are provided in the Appendix.

Procedure

After providing consent to participant in the experiment, all participants took an initial test with fourteen of the multiple-choice questions. Participants were given 20 s to answer each question, and after typing in their answer they continued to view the question until the 20 s elapsed. No feedback was provided. Response times were recorded. After a 3 min delay during which they solved mazes, participants took a final 56-question cued-recall test. Of critical interest was performance for related items, so related items and their comparable control items were tested first to control for output interference (see Roediger & Schmidt, 1980; Smith, 1971). Thus, fourteen questions were related to the questions on the initial test (i.e. the previously non-tested paired question), and these were presented in the first half of the test along with 14 unrelated previously non-tested questions (i.e. control questions, one question from each of the 14 non-tested pairs). The fourteen previously tested questions were tested in the second half of the final test along with the remaining 14 questions from pairs that were not tested in the initial test. Which items served as tested items or control items was counterbalanced across participants, as was which item from a pair of questions was tested. When pairs served as control questions, which item was presented in the first versus second half of the test was also counterbalanced.

Results

Initial test performance

Overall, participants chose the correct answer to 64% ($SE = 3\%$) of the multiple-choice questions on the initial test.

Final test performance

Performance on the final cued-recall test is shown in Figure 1. Previously tested questions were answered with greater accuracy ($M = 55\%$, $SE = 3\%$) than were comparable questions from the control condition

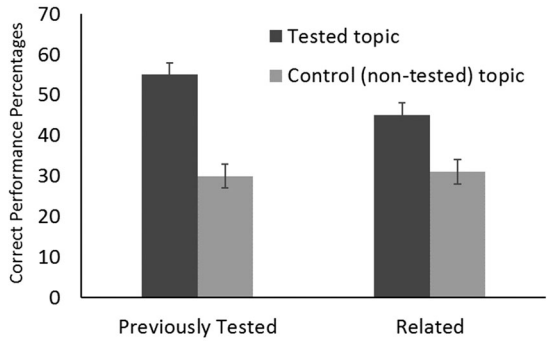

Figure 1. Performance on the final cued-recall test in Experiment 1 for previously tested and related questions from the tested topics and comparable questions from the control condition. Error bars represent +/− SE.

($M = 30\%$, $SE = 3\%$), $t(34) = 7.88$, $p < .001$, $d = 1.34$. Similarly, previously non-tested related questions were answered with greater accuracy ($M = 45\%$, $SE = 3\%$) than were comparable items from the control condition ($M = 31\%$, $SE = 3\%$), $t(34) = 6.16$, $p < .001$, $d = 1.05$.

A potential concern regarding the interpretation of the benefit for related information is that participants, realising that they were receiving a related but new question on the final test, may simply have thought back to the incorrect alternatives and chosen one they could remember. That is, they may have been more likely to recall a previously incorrect alternative – but not necessarily the right one – to the related question following the multiple-choice test, as compared to the control condition, a strategy that should increase correct responses but should also increase intrusions of incorrect alternatives. Indeed, following a multiple-choice test, participants intruded a previously exposed alternative that was not correct for the related question 15% ($SE = 2\%$) of the time, which was significantly more than in the control condition ($M = 10\%$, $SE = 2\%$), $t(34) = 2.19$, $p < .05$, $d = 0.39$. Critically, however, a 2×2 ANOVA revealed that the response increase, as compared to control, was significantly larger for correct answers than for intrusions, $F(1, 34) = 10.47$, $p < .05$, $\eta_p^2 = .24$; that is, the increase in correct responses trumped the increase in intrusions.

On the other hand, a competing concern is that participants did not realise, when they received the previously untested related questions, that the questions were different; and they simply provided their previous response to the new question, which would sometimes be correct for related items. Or, one alternative was particularly salient to the participants, and they simply recalled that response as the answer to all questions for which it could reasonably fit. In both situations, participants may have incorrectly chosen the answer to the related question on the initial test and then provided it as the answer to the related question on the final test. This was not the case. Conditional upon choosing an answer *other* than the answer to the related question (e.g. correctly choosing *Minotaur* or incorrectly choosing *Satyr* or *Chimera* for the *Minotaur* multiple-choice question), participants correctly answered the related questions more often ($M = 48\%$, $SE = 3\%$) than the comparable questions from the control condition ($M = 31\%$, $SE = 3\%$), $t(34) = 6.81$, $p < .001$, $d = 1.16$.

Response times

Response times were recorded for initial key presses and answer submissions (in milliseconds). Below, I only consider response times for correct responses. Table 1 shows response times for submissions, but initial key presses followed the same pattern.

Response times were faster for correctly answered previously tested questions than for correctly answered control questions, $t(33) = 3.72$, $p < .01$, $d = 0.64$. Response times were also faster for correctly answered non-tested related questions than for correctly answered comparable control questions, $t(34) = 3.23$, $p < .01$, $d = 0.55$.

Discussion

In Experiment 1, answering multiple-choice trivia questions increased recall of that information as well as related information on a later cued-recall test. The improvement for previously tested information replicated other work investigating the benefit of answering multiple-choice questions

Table 1. Mean response times (ms) on the Final cued-recall test in Experiments 1 and 2

Experiment	Item Type	Tested	Control (non-tested)
Exp. 1	Previously Tested	5642 (275)	7665 (659)
	Related	7365 (477)	8860 (577)
Exp. 2	Previously Correct Answer		
	Trivia (Previously Tested)	6118 (288)	7502 (381)
	Exposure-control	6973 (338)	7611 (297)
	Previously Incorrect Alternative		
	Trivia (Related)	6598 (274)	7502 (381)
	Exposure-control	7816 (350)	7611 (297)

Note: Standard errors in parentheses.

(e.g. Cantor et al., 2015; see also Butler & Roediger, 2008). More interestingly here, this benefit of multiple-choice testing extends to related information, even without immediate prior exposure to the material.

A potential criticism is that incorrect alternatives may simply have received an increase in accessibility by virtue of being presented in the question and were then outputted as an answer to a new related question. Indeed, participants did sometimes intrude other previous alternatives as answers to related questions; however, this could not explain the full benefit in recall for related information. The increase in correct responses as compared to control was nine percentage points larger than the increase in intrusions as compared to control. The hypothesis is that this benefit for the retention of related information occurs because participants deeply process the incorrect alternatives during the initial multiple-choice test. Perhaps they sometimes recall information about the incorrect alternatives in order to reject them, and in so doing, improve their ability to provide that alternative as the answer to a related question later (see Little & Bjork, 2015). Although the present finding is consistent with these ideas, it is possible that related information was better recalled not because participants processed that information in a meaningful way, but because they were exposed to it.

Although Little and Bjork (2015) provided evidence inconsistent with this type of explanation when the initial test followed the reading of a passage, it is possible that such an explanation would be a valid here. Little and Bjork used passages that contained all of the alternatives, so all of those potential alternatives were to some degree strengthened during the experimental procedure, even if they did not appear in the context of the multiple-choice question. In the case of trivia questions, the alternatives could only be strengthened if they had been provided in the context of a multiple-choice question. It is possible that simply seeing items that are related to a common theme (e.g. mythical creatures) may serve as a reminding that would increase accessibility, and providing a trivia question may not be necessary. Or, participants may realise the connection between the initial and final test when taking the final test and then strategically think back to the earlier test to come up with possible answers when their relevance becomes clear on the final test. Note, however, that although this specific test-taking strategy might increase recall, it should not improve response times because it would presumably take time for participants to think back to the earlier test.

Experiment 2 was conducted to examine whether exposure of the alternatives was enough to provide a benefit for the retention of related information. In Experiment 2, the trivia condition in Experiment 1 was compared to a condition (henceforth called the *exposure-control* condition) in which all of the same alternatives were provided with a question that, although leading to the choice of one answer, would not be expected to induce the type of processing that would foster enhanced recall of the alternatives. For example, in the exposure-control condition, the question, *Alphabetically, which word would come first? A. Minotaur, B., Saytr, C. Centaur, D. Chimera* replaced the trivia condition question, *In classical mythology, a creature that is half human and half horse is called a _____ A. Minotaur, B. Saytr, C. Centaur, and D. Chimera*. In both cases, Centaur is the correct answer, and Minotaur would be the answer to a new question on the final test.

Critically, although I predicted that both conditions would lead to the enhanced recall of the answer that had previously been correct, the incorrect alternatives would only be correctly recalled as answers to new (related) questions in the trivia condition (although the specific test-taking strategy I referenced above might lead to a small numerical boost in the exposure-control condition). Additionally, although I expected an increase in speed in the trivia condition when participants were answering related questions, I did not expect an increase in speed in the exposure-control condition.

A final change made to the procedure of Experiment 2 was to present questions in the final test in a randomised order in order to examine whether the effect obtained in Experiment 1 would occur when related questions were not necessarily tested before previously tested items.

Experiment 2

Method

Participants and design

Ninety-four students at Hillsdale College participated for partial credit in a psychology course. The sample size was more than doubled from Experiment 1 to Experiment 2 because of (a) the addition of the exposure-control condition (between-subjects) and (b) the expectation that the exposure-

control condition might evidence a small numerical advantage for prior testing. Participants were tested one at a time on individual computers. The experiment utilised a 2 (topic: tested, nontested) × 2 (question type: trivia, exposure-control) mixed-subjects design. For the tested topics, I examined questions for which the answer was a previously correct answer and questions for which the answer was a previously incorrect alternative. I examined accuracy and response times.

Materials

For the trivia condition, the materials used in Experiment 2 were the same as those used in Experiment 1. For the exposure-control condition, the questions were modified such that the alternatives and correct answer would be the same as those in the trivia condition, but the questions would not induce spontaneous recall of the information pertaining to the incorrect alternatives that would be later tested. For example, for the trivia condition, a question was, *In classical mythology, a creature that is half human and half horse is called a _____ A. Minotaur, B. Saytr, C. Centaur, and D. Chimera*), with Centaur being the correct answer and Minotaur being an incorrect alternative that would later serve as a correct answer to another question. For the exposure-control condition, the question was *Alphabetically, which word would come first? A. Minotaur, B., Saytr, C. Centaur, D. Chimera.* See the Appendix for other examples.

Procedure

After providing their consent to participate in the experiment, participants were randomly assigned to the trivia or exposure-control condition. Additionally, on the final test, the questions were randomised rather than being presented in blocks as in Experiment 1. Finally, at the end of the experiment, participants answered questions about their experience including whether they noticed the connection between the initial test and the final test and if so, how they think the initial test influenced their performance. The procedure was otherwise the same as in Experiment 1.

Results

Initial test performance

Overall, participants chose the correct answers to 62% ($SE = 3\%$) of the multiple-choice questions on the initial test in the trivia condition and 90% ($SE = 1\%$) in the exposure-control condition. Response times (in milliseconds) were marginally faster in the trivia condition ($M = 10117$, $SE = 308$) than in the exposure-control condition ($M = 10818$, $SE = 215$), $t(74) = 1.9$, $p = .06$, $d = 0.44$.[1]

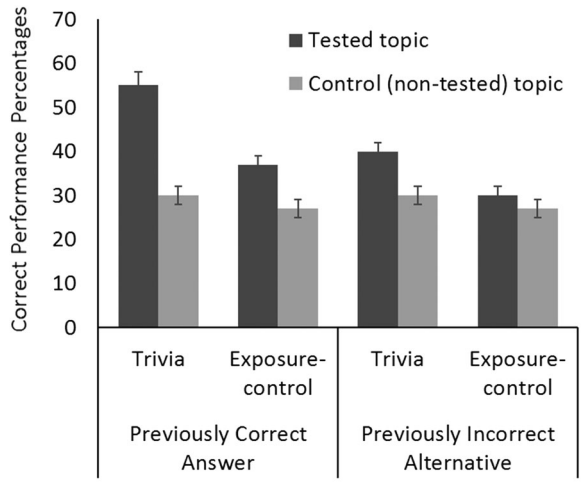

Figure 2. Performance on the final cued-recall test in Experiment 2 for questions for which answers are previously correct answers (previously tested in the trivia condition) and questions for which answers are previously incorrect alternatives (related in the trivia condition) as well as non-tested control questions, as a function of condition (trivia versus exposure-control). Error bars represent +/− 1SE.

Final-test performance

Previously correct answers. Overall performance on the final cued-recall test is shown in Figure 2. The pattern of results suggests that the benefit of testing on the later ability to recall previously correct answers was larger in the trivia condition than in the exposure-control condition. Indeed, a 2 × 2 ANOVA revealed a significant interaction between question type (trivia vs. exposure-control) and topic (tested vs. non-tested control) on the recall of responses that served as correct answers on the earlier test, $F(1,92) = 19.84$, $p < .001$, $\eta_p^2 = .18$. Specifically, in the trivia condition, previously tested questions were answered with greater accuracy ($M = 55\%$, $SE = 3\%$) than were comparable questions from the non-tested control condition ($M = 30\%$, $SE = 2\%$), $t(47) = 9.69$, $p < .001$, $d = 1.43$. Questions were also answered with greater accuracy when the answer had been the correct choice in the exposure-control condition ($M = 37\%$, $SE = 2\%$) as

[1]One computer failed to record response times.

compared to control ($M = 27\%$, $SE = 2\%$), $t(45) = 5.58$, $p < .001$, $d = 0.81$, but the effect was smaller.

Previously incorrect alternatives. As can also be seen in Figure 2, it appears that the benefit of testing on the later correct recall of a previously incorrect alternative was larger in the trivia condition than in the exposure-control condition. A 2×2 ANOVA revealed a significant interaction between question type (trivia vs. exposure-control) and topic (tested vs. control) on the recall of answers that were previously incorrect alternatives and comparable control information, $F(1,92) = 5.50$, $p < .05$, $\eta_p^2 = .06$. Specifically, in the trivia condition, questions related to previously tested questions were answered with greater accuracy ($M = 40\%$, $SE = 2\%$) than were comparable questions from the control condition ($M = 30\%$, $SE = 2\%$), $t(47) = 4.16$, $p < .001$, $d = 0.65$. In the exposure-control condition, there was no benefit of having been exposed to the answer as an incorrect alternative on the earlier test ($M = 30\%$, $SE = 2\%$ and $M = 27\%$, $SE = 2\%$ for previously incorrect alternative and control, respectively), $t(45) = 1.74$, $p > .05$.

Response times

Response times were recorded for initial key presses and answer submissions (in milliseconds), but here, I only include analyses for answer submissions. Initial key presses and submissions followed the same pattern of results. Again, I only consider response times for correct responses. Table 1 shows response times for submissions.

Previously correct answers. A 2×2 ANOVA failed to reveal an interaction between question type (trivia vs. exposure-control) and topic (tested vs. control) on response times for questions that had previously correct answers or comparable control information, $F(1,74) = 2.20$, $p > .05$, but there was a main effect of testing, such that tested items were recalled faster ($M = 6545$, $SE = 225$) than were comparable control items ($M = 7557$, $SE = 239$), $F(1,74) = 16.16$, $p < .001$, $\eta_p^2 = .18$.

Previously incorrect alternatives. In examining the response times for questions having a previously incorrect alternative as the answer, a 2×2 ANOVA revealed a significant interaction between question type (trivia vs. exposure control) and topic (tested vs. control) on response times for related or comparable information, $F(1,74) = 4.03$, $p < .05$, $\eta_p^2 = .05$. In the trivia condition, response times for submissions were faster for correctly answered related questions than for correctly answered control questions, $t(35) = 2.62$, $p < .05$, $d = 0.46$, but this was not the case in the exposure-control condition, $t(39) = .49$, $p > .05$. In fact, response times were numerically slower for questions whose answer was a previously incorrect alternative than for control questions in the exposure-condition condition. Additionally, response times were faster for related items in the trivia condition than for comparable questions in the exposure-control condition, $t(74) = 2.70$, $p < .01$, $d = 0.63$.

Discussion

The goal of Experiment 2 was to examine whether exposure (to the answer to a related question in the form of an incorrect alternative on an earlier multiple-choice test) could explain the increased accessibility to related information demonstrated in Experiment 1. To do so, I compared the condition in Experiment 1 to an exposure-control condition in which participants answered questions that contained the same alternatives but did not involve the answering of the trivia questions provided in Experiment 1. I predicted that trivia questions would provide a benefit for the later recall of related information, but this benefit would not occur in the exposure-control condition. Additionally, I predicted that on the final cued-recall test, speeded response times for related items would manifest in the trivia condition but not in the exposure-control condition. All of these predictions were confirmed.

It should be noted that the method used to create an exposure-control condition in the present study is distinct from that used by Little and Bjork (2015). Little and Bjork manipulated the competitiveness of the alternatives, and showed that a benefit for related information only emerged when the answer to the related question was a previously competitive alternative, not when it was a previously non-competitive alternative. In their study, when the alternatives were non-competitive (i.e. used, in part, as an exposure-control), they were less semantically related to the correct answer than they were when they were competitive. In the present study, the semantic relatedness of alternatives was kept constant across conditions; that is, the same alternatives were used for the questions in the trivia condition and the questions in the exposure-control condition.

Despite controlling semantic relatedness of the alternatives, in general, the questions asked in the exposure-control condition would be expected to induce more shallow processing than those in the trivia condition (Craik & Tulving, 1975). For example, the questions often drew upon physical characteristics of the words, like letters and sounds. Such shallow processing is expected to not improve memory as much as the processing used in the trivia condition (Craik & Tulving, 1975). However, about a quarter of the questions drew upon more meaningful content (e.g. *Which item has the name of a city in Italy in the title? Merchant of Venice, Romeo & Juliet, Hamlet, Julius Caesar; Which of the following has a body of water in its title? A. Yosemite National Park, B. Yellowstone National Park, C. Crater Lake National Park, D. Olympia National Park; Which of the following has the same first name as the famous actor, _____ Redford? A. Robert Fulton, B. Guglielmo Marconi, C. Thomas Edison, D. Eli Whitney; Which of the following has a last name that goes with the word "lock"? A. Francis Scott Key, B. Robert Schumann, C. Stephen Foster, D. George M. Cohan*). Because I did not plan to assess the effect of depth of processing in the exposure-control condition on later recall, proper controls to enable that assessment (i.e. counterbalancing, equal numbers of shallow and deep items) were not included in the present study. On one hand, it is intuitive that questions involving deeper processing would increase performance on a later cued-recall test, and it might be the case that had I only used questions involving deep processing, performance may not have differed between the trivia and exposure-control conditions. On the other hand, however, the extent to which the associations activated during the initial test differ from the associations necessary in the final test, retrieval during the final test may fail (i.e. encoding specificity, Thompson & Tulving, 1970; Tulving, 1979). Thus, explanations for the difference between the trivia condition and exposure-control conditions could include differences in processing depth or the fact that contextual cues were much more different in the latter than in the former condition.

Despite these open questions, it is worth mention that although the majority of the question prompts used in the exposure-control condition would be expected to induce shallow processing and the prompts themselves differed from those used on the later test, the questions in the exposure-control condition included alternatives grouped in a manner that might be expected to increase their semantic processing and possibly provide significant cue overlap between initial and final tests. For example, seeing Minotaur, Centaur, Satyr, and Chimera together should lead participants to notice their common connection of mythological creatures, processing that could be semantic/deep in nature and that would bring to mind ideas about mythical creatures more broadly. In fact, almost all participants reported noticing that alternatives were grouped in this manner, and many said that they tried to use this information during the final test to answer the questions.

General discussion

Names of ancient goddesses, terms for scientific processes, and capitals of countries are initially studied in primary and secondary school, but they are often difficult for young adults to recall. Providing alternatives can make answering questions about this information easier. Critically, the present two experiments showed that answering trivia questions in a multiple-choice format can improve accessibility to related information pertaining to the incorrect alternatives, as evidenced by participants' ability to answer related questions more quickly and accurately than comparable control questions, findings that have not been demonstrated until now. Experiment 2 showed that this increase in recall and speed of recall was not simply the result of previous exposure to the answers to the related questions.

Although performance accuracy is commonly used to assess how testing improves accessibility, response times are not. Pertaining to response times, related questions were answered correctly faster than comparable correctly answered control questions in both Experiments 1 and 2. In Experiment 2, the response times for correct responses to nontested related questions were not only provided faster than correct responses to questions in the control condition; they were faster than response times to related questions in the exposure-control condition which also primed participants with the answers, suggesting that exposure alone is not enough to increase speed of later recall. In fact, in the exposure-control condition, response times for correctly answered related questions were numerically longer than for correctly answered comparable control questions, suggesting that when they could remember the correct answer, participants had to spend at least as much time

retrieving that answer when it was previously primed in the exposure-control condition as when it was not primed at all.

Although I compare this finding to prior research examining multiple-choice testing generally (Butler & Roediger, 2008; Cantor et al., 2015), it is worth noting that the methodology used here has a distinction from this earlier research that could be important. Most studies investigating the role of multiple-choice testing have used such testing following the learning of new information, but we did not. On the other hand, Cantor et al. did investigate the role of multiple-choice tests in improving accessibility to trivia-like information, but they were examining the recovery of *marginal information*, which they defined as information one cannot recall but could answer correctly on a multiple-choice test. Their methodology involved having participants try to answer a cued-recall version of the question before answering a multiple-choice version, and they focused analysis on items that participants failed to retrieve on a cued-recall version but successfully answered on a multiple-choice version. In our research, we did not use this methodology, so although I believe that these questions often tap into marginal knowledge, caution should be taken when seeing this research as a direct extension to the research conducted by Cantor et al.

It is also important that without a study session, we did not control what participants knew (or were previously exposed to) about the alternatives prior to the initial multiple-choice test; instead, I made assumptions about what participants would bring to mind when presented with competitive alternatives and created related questions based on those assumptions. The present findings can then be seen as compelling evidence that multiple-choice questions that have competitive alternatives naturally engender the type of processing that can lead to the enhanced retention of related information pertaining to the incorrect alternatives. This finding should not be taken for granted because previous research with initial cued-recall tests has shown that improved recall of related information is most likely to occur when the connection between the tested and related material is made explicitly, for example, by presenting the tested and related information in a prose text immediately before taking the initial test (Carroll et al., 2007; Chan, 2009; Little et al., 2011). Of course, the multiple-choice questions I used here presented the answer to related questions, whereas cued-recall questions do not; however, the mere presence of the alternatives cannot explain the benefits observed in these two experiments, as evidenced by Experiment 2. One can also infer from examining performance for control items in Experiment 1 that answering cued-recall trivia questions would not facilitate recall of related trivia information. That is, in Experiment 1, half of the cued-recall questions were tested in the first half of the test and compared to the related information tested in the first half of the test, and the other half were tested in the second half of the test and compared to the previously tested items tested there. Thus, it is conceivable that control trivia items tested in the first half of the test would facilitate performance for the related control items tested in the second half of the test. This was not the case. In fact, performance was numerically impaired for control items that were tested second, as compared to those tested first.

One possible reason for why answering multiple-choice questions could improve recall of related information is that when answering questions, participants sometimes think about why the incorrect choices are wrong in order to eliminate them (Embretson & Wetzel, 1987; Skakun, Maguire, & Cook, 1994), and in doing so, they may spontaneously recall information pertaining to them (*retrieval hypothesis*, Little & Bjork, 2015). Although the present findings are consistent with the retrieval hypothesis, a simpler depth of processing account cannot be ruled out on the basis of these experiments because the exposure-control condition tended to have questions that would be expected to induce more shallow processing than that induced by the trivia questions. Furthermore, although it is commonsensical that people sometimes answer multiple-choice questions by thinking about why incorrect alternatives are wrong, they do not always do so; individuals can use a variety of strategies (Skakun et al., 1994). I suggest here that people probably only engage in this type of retrieval processing when (a) the answer is not readily accessible (because in that case, they would not consider the alternatives at all), (b) they know something about the incorrect alternatives, and (c) they are internally or externally motivated to try to answer the question correctly. Such circumstances are not going to be present for all people and for all questions. Although the goals of the present work were (a) to provide evidence of increased accessibility of

related information following multiple-choice testing with novel trivia materials and (b) to rule out a simple exposure explanation, future research should aim to clarify the processes that lead to this increased accessibility.

Additional considerations

Educators are often critical of multiple-choice questions when used for assessment (Frederiksen, 1984), and the present research does not address the use of multiple-choice questions in this way. Instead, the present research addresses the use of multiple-choice questions as a tool for learning, and when used in this way, multiple-choice testing improves access to previously tested information as well as some related information. This finding has implications for educational practice. Evidence suggests that the use of multiple-choice questions such as the ones used here is beneficial when students are learning potentially confusable information (Little et al., 2012; Little & Bjork, 2015). A clear limitation of the present research, however, is that the trivia questions used here (as well as materials used in most of the other research discussed in this paper) are simplistic, based solely on fact knowledge. Such questions target the lowest level of Bloom's Taxonomy (Bloom, 1956). For summative assessment, and often for formative assessment (learning), instructors would often aim to target their tests at higher levels within the taxonomy (e.g. understanding, application). Nevertheless, it is important to acknowledge that fact knowledge often underlies more complex ways of using knowledge, and in some domains, particularly in introductory courses, learning terms, names, and other facts is necessary. It is in these contexts where the present research has the clearest applications.

It should be noted that the increases in recall on a final cued-recall test occurred without the provision of feedback. In research examining the influence of multiple-choice testing on the retention of tested information, Butler and Roediger (2008) showed that the provision of feedback (and increased study prior to testing) improves performance on a final cued-recall test, so I would expect feedback to provide the same benefit here, at least for previously tested information. Whether the provision of feedback would improve recall of related information is uncertain with the present materials. Little et al. (2012) manipulated feedback in their second experiment (using the study-plus-multiple-choice test procedure) and found that feedback improved performance for previously tested information but not for related information.

Another question of interest is the persistence of the effect of intervening multiple-choice testing on long-term retention, as the present experiments used very short delays. Cantor et al. showed that the benefit in retention of marginal knowledge persisted for a week, and Little and Bjork (2012) showed that the benefit in retention of related information (in their paradigm in which testing followed the study of a prose passage) persisted for 48-hours (i.e. the longest time-frame that they tested). See also Bjork, Little, & Storm, 2014, for an examination of multiple-choice quizzing in a research methods course. The expectation, therefore, is that these effects would persist as well, but this is an empirical question.

A final issue worth considering is the possibility that the multiple-choice tests produce "misinformation." Much work (Butler & Roediger, 2008; Butler, Marsh, Goode, & Roediger, 2006; Marsh, Roediger, Bjork, & Bjork, 2007; Roediger & Marsh, 2005) has shown that answering multiple-choice questions can increase the likelihood of participants intruding incorrect alternatives on a later test of the previously tested information. That is, by virtue of answering a multiple-choice question, participants are more likely to intrude a previously incorrect alternative as an answer to that question later. This effect is well established, and although I did not present data regarding misinformation (in regard to previously tested information) in the present paper, analysis of a subset of the data presented here suggests that these experiments replicate those patterns of results.[2] I argue, however, that although misinformation effects are a valid concern in terms of multiple-choice testing, their presence does not negate the benefits that multiple-choice tests offer, especially those described here. Furthermore, misinformation effects are often eliminated when feedback is provided (Butler & Roediger, 2008).

Concluding comments

Following from research on multiple-choice testing (e.g. Butler & Roediger, 2008; Cantor et al., 2015;

[2]In Experiment 1, following a multiple-choice test, participants intruded items that had been previously incorrect alternatives more often ($M = 24\%$, $SE = 3\%$) than in the control condition ($M = 7\%$, $SE = 2\%$), $t(34) = 5.51$, $p < .01$, $d = 1.40$.

Little et al., 2012; Little & Bjork, 2015), the present work demonstrates a new benefit that multiple-choice testing has for learning – the increased accessibility of difficult-to-recall knowledge that was not directly tested, measured both in terms of correct responses and response times. Answering a trivia question in a multiple-choice format engages deep processing that results in increased learning – not just of the tested content, but of related information. Results from the present studies and past research suggest that instructors in educational settings should incorporate regular multiple-choice testing and quizzing with competitive alternatives relevant to course materials to optimise learning.

Acknowledgement

I thank Mark A. McDaniel for conversations pertaining to Experiment 1 and for comments on an earlier version of this paper.

Disclosure statement

No potential conflict of interest was reported by the author.

References

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.

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, *3*, 165–170.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, handbook I: Cognitive domain*. New York: David McKay.

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, *20*, 941–956.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*, 604–616.

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.

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.

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.

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.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement*, *11*, 175–193.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, *39*(3), 193–202.

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.), *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.

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.

Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory (Hove, England)*, *19*, 346–359.

Marsh, E. J., RoedigerIII, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, *14*, 194–199.

RoedigerIII, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

RoedigerIII, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155–1159.

RoedigerIII, H. L., & 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.

Skakun, E. N., Maguire, T. O., & Cook, D. A. (1994). Strategy choices in multiple-choice items. *Academic Medicine*, *69*, S7–S9.

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*, 400–408.

Thompson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*, 255–262.

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Appendix

Following are seven example sets that show versions of the initial and final test questions from Experiments 1 and 2. The first question in each set is a trivia multiple-choice question that could have been used in an initial test in Experiment 1 and 2. The second question is a comparable exposure-control question used in Experiment 2. The third question is a final test question for which a previously incorrect alternative would be correct.

Set 1 (Russian Literature)

_____ is a novel by Fyodor Dostoevsky in which a man kills two old women because he believes that he is beyond the bounds of good and evil.
A. Crime and Punishment, B. War and Peace, C. Brothers Karamazov, D. Anna Karenina

Which of the following titles most clearly includes a cause and an effect? **A. Crime and Punishment,** B. War and Peace, C. Brothers Karamazov, D. Anna Karenina

_____ 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**

Set 2 (Foreign Language)

How do you say "thank you" in Japanese? **A. Arigato**, B. Sayonara, C. Xie xie, D. Obrigado

Which of the following is spelled Otagira backwards? **A. Arigato**, B. Sayonara, C. Xie xie, D. Obrigado

How do you say "good-bye" in Japanese?
Answer: **Sayonara**

Set 3 (Chemistry)

The spreading of atoms or molecules of one substance through those of another, especially into liquids and gases is known as _____. A. Entropy, B. Diffraction, **C. Diffusion**, D. Distillation

Noisuffid is which word spelled backwards? A. Entropy, B. Diffraction, **C. Diffusion**, D. 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**

Set 4 (Sports)

Who was the first college football player to win the Heisman Trophy as a sophomore? A. Earl Campbell, B. Archie Griffin, C. Jay Beranger, **D. Tim Tebow**

Alphabetically by last name, who would come last? A. Earl Campbell, B. Archie Griffin, C. Jay Beranger, **D. Tim Tebow**

Who was the only college football player to win the Heisman Trophy twice?
Answer: **Archie Griffin**

Set 5 (Geography)

What is the capital of Norway? **A. Oslo**, B. Helsinki, C. Stockholm, D. Leningrad

Which of the following is the shortest word? **A. Oslo**, B. Helsinki, C. Stockholm, D. Leningrad

What is the capital of Finland?
Answer: **Helsinki**

Set 6 (Clouds)

What is the term for large, white, puffy clouds that generally appear in fair weather, but that can also form thunderheads on hot days? A. Stratus, B. Nimbus, C. Cirrus, **D. Cumulus**

Which of the following has three syllables? A. Stratus, B. Nimbus, C. Cirrus, **D. Cumulus**

What is the term for lacy or wispy clouds that form at high altitudes, often before a change in the weather?
Answer: **Cirrus**

Set 7 (Shakespeare)

From what Shakespearian play comes the line, "This above all: to thine own self be true"? A. Julius Caesar, **B. Hamlet**, C. Romeo and Juliet, D. Merchant of Venice

Which of the following is a word that means "small settlement"? A. Julius Caesar, **B. Hamlet**, C. Romeo and Juliet, D. Merchant of Venice

From what Shakespearian play comes the phase "pound of flesh"?
Merchant of Venice