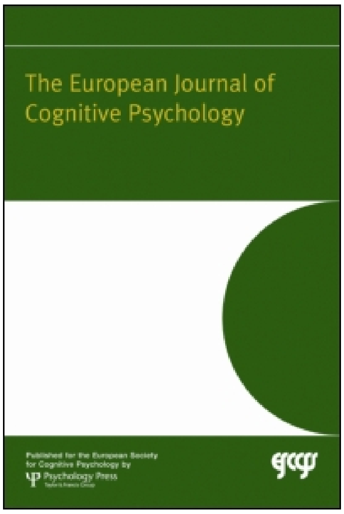
The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting

Article *in* European Journal of Cognitive Psychology · August 2009

DOI: 10.1080/09541440802413505

Sue Morris
UNSW Sydney

Kaaren J Watts



European Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713734596

The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting

Jacquelyn Cranney [a]; Mihyun Ahn [a]; Rachel McKinnon [a]; Sue Morris [b];Kaaren Watts [a]
[a] School of Psychology, University of New South Wales, Sydney, Australia [b] Learning and Teaching at UNSW, University of New South Wales, Sydney, Australia

First published on: 01 October 2008

To cite this Article Cranney, Jacquelyn , Ahn, Mihyun , McKinnon, Rachel , Morris, Sue andWatts, Kaaren(2009) 'The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting', European Journal of Cognitive Psychology, 21: 6, 919 — 940, First published on: 01 October 2008 (iFirst)

To link to this Article: DOI: 10.1080/09541440802413505

URL: http://dx.doi.org/10.1080/09541440802413505

EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY
2009, 21 (6), 919–940

Psychology Press
Taylor & Francis Group

The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting

Jacquelyn Cranney, Mihyun Ahn, and Rachel McKinnon
School of Psychology, University of New South Wales, Sydney, Australia

Sue Morris
Learning and Teaching at UNSW, University of New South Wales, Sydney, Australia

Kaaren Watts
School of Psychology, University of New South Wales, Sydney, Australia

Two studies were conducted to investigate aspects of the test effect in a tertiary education setting. During weekly tutorial sessions first year psychology students watched a psychobiology video (Phase 1), followed by different video-related activities (Phase 2). In the tutorial 1 week later, students took an unexpected test (Phase 3). In Phase 2 of Study 1, students completed a quiz in small groups (group quiz) or individually (individual quiz), highlighted the video transcript (re-study), or did nothing further (no-activity). Group quiz performance was superior to individual quiz in both Phase 2 and Phase 3. In Phase 3 individual quiz students performed better than no-activity students, but not better than restudy students. In exploring the individual testing effect further, Phase 2 of Study 2 included quiz (individual), restudy, and no-activity conditions. Quiz participants were presented with one (target) of two sets of questions, whereas restudy participants were presented with equivalent statements. During Phase 3, all participants answered both sets of questions (target and related). Quiz performance was superior to restudy and no-activity performance on both target and related material, supporting the retrieval-induced facilitation explanation of the testing effect. Implications of the current research for assessment practices in classroom settings are discussed, and directions for future research are indicated.

Keywords: Testing effect; Collaborative testing; Retrieval-induced facilitation; Educational setting.

Correspondence should be addressed to Jacquelyn Cranney, School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: j.cranney@unsw.edu.au

Study 1 in this paper was presented at the 29th annual conference of the National Institute on the Teaching of Psychology, St. Petersburg Beach, Florida, January 2007, as well as at the Australasian Experimental Psychology Conference, Canberra, April, 2007. The authors would like to acknowledge the participation and cooperation of all the PSYC1011 students and tutors in 2006 and 2007, as well as the financial support of the School of Psychology, UNSW.

http://www.psypress.com/ecp DOI: 10.1080/09541440802413505

THE TESTING EFFECT

Knowledge of an encroaching assessment usually prompts students to study; however, they do not always study in an effective manner. For example, "study" in the form of simple rereading of the material may not be the optimal means for enhancing memory for that material. Many studies within the cognitive learning literature suggest that repeated testing leads to greater retention of information than does repeated reading, a finding termed the "testing effect" (Glover, 1989; Spitzer, 1939). In particular, research has shown that repeated testing leads to greater retention of information than repeated rereading sessions when there is a delay of 24 hours or more between the retrieval practice and final recall stages (Glover, 1989; Karpicke & Roediger, 2007; Kuo & Hirshman, 1996; Spitzer, 1939).

The testing effect has been examined primarily in laboratory settings and the learning materials administered have commonly been dissimilar to real educational tasks (e.g., word lists or pairs of words; Roediger & Karpicke, 2006a). Recent research by Roediger and Karpicke (2006b), however, provides a demonstration of the testing effect using educational prose material covering general scientific topics. In their Experiment 2, all of the participants studied one of two prose passages in the initial learning block. In the subsequent three 5 minute blocks, participants in the repeated study condition (SSSS) reread the passage; participants in the SSST condition reread the passage for the next two periods, and then completed a free-recall test related to the studied information; and participants in the repeated test condition (STTT) took three consecutive free recall tests. Participants then received a final recall test either 5 min after the final learning block or 1 week later. As predicted, at the 5-min retention interval, participants in the SSSS condition had the highest rate of recall, but at the 1-week retention interval, participants in the STTT condition showed the greatest recall. In other words, although massed reading had short-term efficiency, repeated testing was superior to repeated reading in promoting long-term retention. A separate study by Butler and Roediger (2007) found that the testing effect can be generalised to different methods of presentation of the original learning material, including video stimuli.

A number of different direct and indirect mechanisms have been proposed to underlie the testing effect, and although some have been discarded (Fitch, Drucker, & Norton, 1951; Roediger & Karpicke, 2006a; Thompson, Wegner, & Bartling, 1978), it is likely that more than one mechanism involving both Transfer Appropriate Processing (TAP; Karpicke & Roediger, 2007; McDaniel, Friedman, & Bourne, 1978; Morris, Bransford, & Franks, 1977) and retrieval induced facilitation (e.g., Dempster, 1996; Roediger & Karpicke, 2006a) is operative. The retrieval hypothesis, for example, states that the level of retrieval effort can impact upon memory retention, such that elaboration during the retrieval process can increase the strength of a memory trace and increase the number of retrieval routes (Anderson, Bjork, & Bjork, 1994; Butler & Roediger, 2007; Roediger & Karpicke, 2006a), enabling the information to be better remembered at a later time. A test administered following the initial encoding stage is a more effortful process of continued learning than rereading because testing involves cognitive engagement, elaboration, and manipulation of concepts rather than relatively passive reprocessing of information. Thus, practice testing increases later memory performance (Roediger & Karpicke, 2006a).

An important issue that has arisen in this research is the appropriate comparison condition for the testing effect. In Roediger and Karpicke (2006b, Exp. 1), the testing effect was demonstrated by the superior performance of the test condition compared to the restudy condition on the final (delayed) criterion test. It is argued that this is a strict criterion testing effect because it occurs between two groups of participants who are given the same context and time frame in which to learn the material and only the method of review varies: testing or rereading. A lenient criterion testing effect would be that the test condition leads to better performance than a no-activity condition where there is no exposure to the material during the review phase (that is, neither rereading nor testing). Although the lenient criterion (cf. the strict criterion) application is of less interest theoretically, it is practically interesting to examine whether the test condition leads to better, and not worse, performance than no study at all.

The role of feedback in promoting the testing effect is yet to be defined. The testing effect has been demonstrated in many studies without feedback (e.g., Roediger & Karpicke, 2006b; Wheeler & Roediger, 1992). Nevertheless, feedback is generally considered to enhance the testing effect (e.g., Kulhavy & Stock, 1989; McDaniel & Fisher, 1991), and immediate feedback is widely recognised as optimal for maximising students' memory for the tested information in applied classroom settings (see Kulik & Kulik, 1988). Hence, in the current classroom research, corrective feedback immediately after testing was provided.

COLLABORATIVE TESTING

The benefits of collaborative learning in classroom settings have begun to be empirically documented, particularly in undergraduate psychology. Collaborative learning situations, for instance, promote "in-depth, process-oriented research" (Miyake & Shirouzu, 2006, p. 102). In traditional top-down teaching approaches, teachers or lecturers impart knowledge and are perceived as the "knowledge authorities". In contrast, collaborative learning encourages the development of "learning communities" and promotes the perception among novice learners that "knowledge is common property rather than primarily the property of authorities" (Cockrell, Caplow, & Donaldson, 2000, p. 359). Most importantly, collaborative learning promotes the development of strategies (e.g., cooperation, communication, and negotiation), which closely approximate skills required in the "real world" (p. 348). The current research will explore the effectiveness of collaborative test taking on later individual test performance.

In laboratory experiments it has been shown that during initial group recall of study material, group performance is superior to individual but not to nominal group performance; however, the subsequent individual memory performance is better for those who initially recalled in a collaborative than in a nominal group (e.g., Basden, Basden, & Henry, 2000; Weldon & Bellinger, 1997). Between- rather than within-session collaborative test-taking effects have not been rigorously explored.

The primary aim of the current quasi-experimental research was to examine the generalisability of the testing effect to classroom settings. The preliminary Study 1 focused on collaborative testing effects, whereas Study 2 improved the methodology in search of stronger individual testing effects on memory for the whole learning episode. In Study 1 it was predicted that memory performance for "old" test items (presented in the initial test, i.e., Phase 2) would be superior for those participants who completed the initial test compared to those who did not (lenient criterion). It was also expected that memory performance for "old" test items would be better for those who reviewed the material in an initial test compared to participants who simply restudied the material (strict criterion). It was anticipated that collaborative testing would produce superior memory retention for "old" test items relative to individual testing. In addition, an exploratory analysis of "new" test items, that is, test items from the study material but not included in the initial test, was also undertaken.

STUDY 1

Method

Participants and design

Seventy-five students (25 males, 50 females), who were enrolled in an introductory psychology course at the University of New South Wales, participated in this study. The age range was 15 to 28 years ($M = 19.42$, $SD = 1.90$). This introductory course was comprised of 4 hours of large-class lectures and 1 hour of a smaller group tutorial (which focused on lecture-related information and activities) each week. Data was analysed only from students who attended tutorials on Mondays because it was reasoned that students with tutorials on subsequent days may have become privy to information about the procedure from other students, potentially contaminating the results (i.e., the communication problem; Meltzoff, 1998).

There were three experimental phases: initial learning, review, and final memory test. The first two phases were presented consecutively in the first tutorial, and the final phase was presented a week later in the next tutorial. In a quasi-experimental approach, tutorials were randomly allocated to one of four conditions (one tutorial group per condition). The conditions were differentiated by the activity students engaged in during the review phase, after viewing a video recording in the initial phase. The four activities were: group quiz, individual quiz, restudy, and no-activity control. The dependent variable was percentage of items correct on the Phase 3 final memory test.

Materials

The video viewed by participants was called *Discovering Psychology: The Behaving Brain* (WGBH Boston & American Psychological Association, 2001). The content of the video was an introduction to the way in which electrical and chemical messages are transmitted in the brain, and the basic functions of several of the major brain structures.

Two quizzes were developed based upon the information presented in the video. The quiz (Phase 2) consisted of 10 questions (5 multiple choice and 5 fill-in-the-blanks, tapping into different concept units from the video). The final memory test or pop quiz (Phase 3) consisted of 13 items: Three of these items were replicated exactly from Quiz 1 ("old") (two multiple choice, one fill-in-the-blank); seven were new questions, derived from the video shown during the first phase ("new") (three multiple choice, four fill-in-the-blanks). Two of the "new" multiple choice questions partly overlapped with the concept units of two of the Phase 2 short answer questions (e.g., Phase 2: "the hippocampus is part of the [limbic] system"; Phase 3: "Which structures are part of the limbic system?"; correct answer: amygdala, hippocampus, hypothalamus, and thalamus). In addition, there were three new multiple choice questions based on information presented during the lectures ("unrelated material"). It is important to note that only participants in the group quiz and individual quiz groups completed the quiz and, therefore, participants in the restudy and no-activity groups were *not* exposed to the "old" questions prior to completing the final memory test. All materials in Study 1 and 2 are available from the first author upon request.

Procedure

The study was integrated into the normal activities of the first year psychology tutorials; therefore, the time and place for each of the phases was the same throughout the experiment. There were three phases in this study, conducted over two sessions.

Phase 1: Initial learning. Psychobiology was introduced as a new topic in lectures during the week of Phase 1. In their tutorials, all students viewed an 8-min portion of the video about psychobiology. Students were instructed not to take notes because they would receive a transcript of the video prior to the exam.

Phase 2: Review. Immediately following the video, group quiz students collaborated with their team of 4 to 5 people (teams were formed during the previous week of tutorials) to complete the quiz. Individual quiz students completed the same quiz individually and without discussion. Students were given 10 min to complete the quiz; the quizzes were then collected. The quizzes were then randomly allocated to another group/individual to mark. Feedback was provided such that the answers were presented one at a time and at a reasonably rapid rate on an overhead projector, and students could ask questions. The quiz was then collected again, and in the group condition, the winning (highest score) group was announced. This feedback process took 5 min. Although students could check their score, they could not take their quiz away.

Restudy students read through a transcript of the video for 8 min and were instructed to "highlight or underline all the words and phrases that you consider to be important or relevant to this course". Then, they had 2 min to ask any questions about the material. Their transcripts were then collected. Students in the no-activity condition did not re-engage with the video information. In order to ensure equity of information access for further study within the course, the video transcript was made available to all students on the WebCT Vista course site 1 week after the completion of testing in this study (i.e., after Phase 3).

Phase 3: Final memory test. In tutorials exactly 1 week after Phase 2, all students were presented with an unexpected psychobiology "pop" quiz. They were told that although this quiz would not count toward their final grade, it would be an indication of how much they knew in preparation for a test in 3 weeks' time. Prior to completing the final memory test, participants judged the grade they thought they might receive (1 = "fail" to 5 = "high distinction"), and also indicated the number of hours they had spent independently studying the topic of psychobiology during the previous week.

Results and discussion

All of the analyses conducted were one-way analyses of variance (ANOVAs), unless otherwise specified. The significance level throughout was an alpha of .05.

Sample characteristics. The uniformity of conditions in terms of their demographic composition was checked using chi-square analyses, which indicated that there were no differences between conditions in terms of gender, $\chi^2(3, N = 75) = 6.298$, $p = .098$, phi = .29, or English/non-English native language, $\chi^2(3, N = 75) = 2.184$, $p = .535$, phi = .17.

Testing effect. Evidence for the testing effect in this education setting was determined by responses to "old" items, that is, those previously seen by group quiz and individual quiz participants in the Phase 2 quiz. The mean percentage correct (old items), as a function of postvideo condition, is provided in Figure 1.

A one-way ANOVA on "old" items confirmed that there was a significant difference in test performance between groups, $F(3, 71) = 8.67$, $MSE = 995.10$, $p < .05$, $p\eta^2 = .27$. Individual quiz participants performed more accurately on old items than did no-activity participants, $t(71) = 2.12$, $SEM = 10.41$, $p < .05$, $d = .50$. This means that completion of an individual quiz, with feedback, facilitated more accurate responding for items that were presented on the Phase 2 quiz. There was, however, no difference between individual quiz and restudy conditions on old items, $t(71) = 1.63$, $SEM = 10.82$, $p = .11$. Thus, the strict testing effect criterion, which allows a conclusion that the testing process produced gains over simply revisiting the material, was not met.

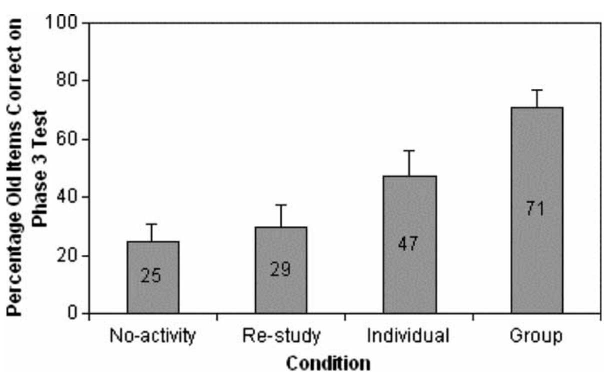


Figure 1. Mean percentage of "old" items correct on the Phase 3 test as a function of postvideo condition for Study 1. The error bars indicate the standard error of the mean.

Collaborative testing. Group quiz scores on the Phase 2 quiz ($M = 94\%$, $SD = 5.48\%$) were higher than the individual quiz scores ($M = 47.4\%$, $SD = 20.83\%$), $F(1, 24) = 23.93$, $MSE = 366.71$, $p\eta^2 = .50$. This means that collaborative input during the testing process resulted in greater accuracy. This is likely to be due to the benefits of pooled information.

Consistent with the expectation for the benefits of collaborative testing, those who completed the group quiz performed better than those who had completed the individual quiz on old items in Phase 3, $t(71) = 2.29$, $SEM = 10.29$, $p < .05$, $d = .54$. This suggests that a collaborative testing effect was maintained following a 1 week interval and reflects the impact of good performance in Phase 2 on the strength of the testing effect.

New items. An exploratory one-way ANOVA on new items revealed that there were significant differences in test performance between groups, $F(3, 71) = 6.39$, $MSE = 381.94$, $p < .05$, $p\eta^2 = .21$. The mean percentage correct on the new items, as a function of postvideo condition, is presented in Figure 2. It appears that the group quiz condition showed facilitated performance compared to all other conditions. No differences were observed between the individual quiz and restudy conditions, $t(71) = 0.25$, $SEM = 6.70$, $p = .80$, or between the individual quiz and No-activity conditions, $t(71) = -0.89$, $SEM = 6.45$, $p = .38$. That is, completion of the individual quiz did not lead to a significant enhancement of or detriment to memory for the new items. However, group quiz participants scored higher on new items than individual quiz participants, $t(71) = 3.57$, $SEM = 6.38$, $p < .05$, $d = .85$.

Overall, the group quiz participants demonstrated superior test performance on new items compared to the other three groups (individual quiz, restudy, and no-activity), which provides strong evidence of a facilitation effect resulting from the collaborative testing procedure. This is likely to be

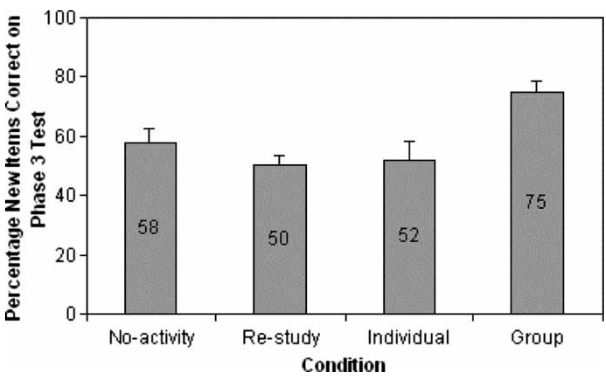


Figure 2. Mean percentage of "new" items correct on the Phase 3 test as a function of postvideo condition for Study 1. The error bars indicate the standard error of the mean.

the result of broader engagement with the video information as a result of group discussion in Phase 2.

Alternative explanations. The possibility of preexisting academic differences (such as superior general test performance or superior general comprehension of the psychobiology material) between the groups, or of some kind of tutor effect, was checked by comparing performance on the unrelated lecture-based questions. No differences were observed amongst the four conditions (no-activity, $M = 43.3\%$, SD = 28.8%; restudy, $M = 47.1\%$, $SD = 35.5\%$; individual quiz, $M = 47.1\%$, $SD = 35.5\%$; group quiz, $M = 49.2\%$, $SD = 22.7\%$), $F(3, 71) = 0.13$, $MSE = 933.86$, $p = .94$. The possibility of differences between groups on reported hours spent studying between Phases 2 and 3 was examined, and no significant differences existed amongst the conditions (no-activity, $M = 1.45$, $SD = 1.61$; restudy, $M = 3.24$, $SD = 3.42$; individual, $M = 1.41$, $SD = 1.57$; group, $M = 1.93$, $SD = 2.69$), $F(3, 71) = 2.14$, $MSE = 5.92$, $p = .10$.

An additional analysis assessed the potential benefit of the "study" engaged in by the restudy participants. Old and new items were combined in this analysis since neither restudy nor no-activity conditions had been previously exposed to the quiz questions. It was found that there was no difference in accuracy in the Phase 3 test between participants who highlighted the video material ($M = 44.1\%$, $SD = 12.7\%$) and the no-activity group ($M = 48\%$, $SD = 19.6\%$), $t(35) = -0.70$, $SEM = 5.54$, $p = .49$. Thus, the short session of highlighting produced no gains in memory that were maintained over the 1 week interval.

In summary, Study 1 demonstrated the testing effect in a real educational setting: that is, students who were tested in the initial phase performed better on "old" items in the final test than did those who did not spend any more time on the material. The effect, however, did not generalise to related "new" questions. The benefits of collaborative testing were demonstrated. Collaborative testing enhanced accuracy during an initial test and improved retention for both "old" and "new" questions on a final test; these effects are considered further in the General Discussion. It is unclear, however, why the individual testing condition did not yield better performance than for the restudy condition. Only one other study (McDaniel, Anderson, Derbish, & Morrisette, 2007), using a somewhat different methodology, has demonstrated a strict-criterion testing effect in a real educational setting. Is it possible that there are significant limitations to the generalisation of the testing effect from the laboratory to the applied setting? If so, then we need to acknowledge the limited practical value of such laboratory work, or determine the conditions under which findings reliably generalise to the classroom setting.

Indeed, the lack of replication of the strong testing effect to the educational setting in this study may be related to a number of methodological factors. First, there were very few "old" items, and so the measure of memory on the final test was limited in its range. Second, the "old" and "new" questions were not matched for difficulty, and it would appear from a comparison within the no-activity condition that the "new" questions were easier than the "old" questions, $t(19) = 6.31$, $SEM = 5.20$, $p < .05$. Third, some of the "new" questions tested similar concepts or information to some of the items in Phase 2. Fourth, there simply may not have been enough participants in each of the conditions to provide the power to detect what appears to be the superior performance of the individual quiz students compared to the restudy students (see Figure 1). Given the generalisabilty considerations discussed earlier, it was thought important to address these methodological limitations in Study 2 through careful selection and matching of test items. A related important educational issue that was not intentionally and directly tested in Study 1 is the opposing evidence in the literature that taking an initial test may either enhance memory for new, related material, termed "retrieval-induced facilitation" (Chan, McDermott, & Roediger, 2006), or, conversely, inhibit later memory for new, related material, termed "retrieval-induced forgetting" (Anderson, 2003). Study 2 addressed this issue within the context of the further exploration of the test effect with the individual condition.

STUDY 2

A retrieval induced facilitation effect is based on associative memory theories, such that activation of one concept produces facilitative effects for related concepts. For example, when a semantic node is activated, it also activates semantically related nodes through an associative network, a process called spreading activation (Collins & Quillian, 1972). This theory has also been applied to episodic memory (Roediger & McDermott, 1995). The adaptive control of thought–rational (ACT-R) and search of associative memory (SAM) theories have also been used to explain the underlying effects of retrieval induced facilitation. Both theories suggest that retrieval of a particular subset of studied material facilitates retrieval of the unstudied material when associative links exist between them (Anderson, 1996; Raaijmakers, 2003). Moreover, the stronger the associative links are between the two sets of material, the higher the probability of retrieval of the unstudied material. These theories, therefore, postulate that because the associative links between the tested and related materials are strengthened through repeated testing sessions, the retrieval of related material in a final memory test will be more likely than that of unrelated material.

Two studies by Chan et al. (2006) demonstrated that initial testing enhanced later memory for untested, related material. In Study 1, participants were allocated to one of three groups based on three different review conditions: test, restudy, and no-activity. In the review phase, the test group completed two tests on a subset of a target prose article while the restudy condition reread a subset of the target article, twice. The no-activity group did not receive any additional exposure to the material. The final memory test on the target article was administered 24 hours later. Consistent with the strict criterion, the test group had significantly better recall for the untested, related items on the final test than both the restudy and the no-activity groups.

In Chan et al.'s (2006) Study 2, participants were exposed to two sets of prose material (Stories A and B) in the initial learning stage in a within subjects design. This design did not allow for a test of the strict criterion. All participants received a test on a subset (A1) of one of the original stories as practice for the final memory test. A retrieval-induced facilitation effect was obtained such that performance on the related material (A2) was significantly better than that on the unrelated material (B). This provided a conceptual replication of their Study 1; testing significantly increased performance on a later memory test compared to no extra study. Hence, the testing effect was obtained in both experiments despite participants receiving no feedback about their performance.

In contrast to retrieval induced facilitation, retrieval induced forgetting is a phenomenon whereby remembering information can cause impairment in the ability to remember unpractised related information (Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007). It is based upon the notion that when two or more items of information are to be retrieved, we unconsciously forget the information considered irrelevant, in the process of retrieving the desired information (Anderson et al., 1994). This theory is consistent with the retrieval induced facilitation hypothesis that for the repeatedly retrieved target material (A1), the retrieval process itself is strengthened, leading to better "final" recall. However, during retrieval there is competition for activation so that semantically related material (A2) is unconsciously inhibited in order for the target material (A1) to be better recalled. Therefore, the more closely the material is related, the more likely a forgetting effect is to occur for the related material. Carroll et al. (2007), however, found a retrieval-induced forgetting effect only if the final test was immediately, rather than 24 hours after, the review session. In summary, the retrieval induced facilitation and forgetting effects appear to operate as opposing influences on learning through repeated retrieval of information.

The initial objective of Study 2 was to improve aspects of the Study 1 methodology and test again whether the individual test condition would produce better final memory test performance than the restudy condition, that is, the strict testing effect criterion not met in Study 1. A second objective of Study 2 was to determine whether recall for nontested related material was facilitated or inhibited in a real educational setting. Three study conditions were utilised: quiz (individual), restudy, and no-activity. Two sets of questions were developed to test discrete but related concepts from the video material. One set was utilised during Phase 2 (target material), while both sets (target, related) were administered during Phase 3. It was predicted that the testing effect would be demonstrated for the target material such that the quiz group would perform better than the control group (lenient criterion). It was also expected that the quiz group would outperform the restudy group (strict criterion). Given that the final test was 1 week after the review session, it was expected that a retrieval induced facilitation rather than forgetting effect would be found for the related material; specifically, the quiz group would perform better than the restudy and no-activity groups.

Method

Participants and design

Two hundred and eighteen first year psychology students (126 females and 92 males), enrolled in the Psychology 1B classes at the University of New South Wales participated in this experiment. The mean age was 19.9 years (range = 17–51). Data was analysed only for those students present for both sessions of the study.

As in Study 1, there were three experimental phases: initial learning, review, and the final memory test. The first two phases were presented consecutively in the one tutorial, and the final phase was presented 1 week later in the next tutorial. In a quasiexperimental approach, tutorials were randomly allocated to one of three conditions: quiz, restudy, and no-activity conditions. Data was analysed from the first six (two per condition) tutorials of the week where there were no procedural issues such as equipment malfunction. These occurred from Monday to Wednesday. The independent variable was the type of review method, which was operationalised by the presentation of the practice quiz, summary highlight sheet or nothing, during the review phase. The dependent variable was memory recall, which was measured in the final memory test.

Materials

The video learning material was identical to that presented in Study 1. All quizzes and summaries were presented on A4 sheets of paper as handouts. Two practice quizzes (I, II) were created for this experiment and each consisted of 10 cued recall item questions. During Phase 2 half of the quiz participants received Quiz I and half of them received Quiz II—this was the target material (A1). During Phase 3 all of the Quiz I participants received both the Quiz I questions (target A1) and the Quiz II questions (related A2). Similarly, two summary sets, which consisted of 10 summary points that were the completed items of the cued recall items, were also created. The final memory test (pop quiz) consisted of all the practice quiz questions (target and related), presented in item-by-item alternating order, and there were two counterbalanced orders. On the first page of the pop quiz students were asked to (a) indicate how many hours of psychobiology study they had undertaken in the past week and (b) make a percentage prediction of their own performance on the subsequent pop quiz (data not reported here).

Procedure

The study was integrated into the normal activities of the first year psychology tutorials; therefore, the time and place for each of the phases was the same throughout the experiment.

Phase 1: Initial learning. Participants were presented with the video segment in the first part of the tutorial. As in Study 1, they were instructed not to take notes during the video, as a transcript would be available later from the course website.

Phase 2: Review. Immediately after the video, participants received their specific review task, depending upon which tutorial group they were in. For the quiz group, one of the two practice tests (I, II; counterbalanced for each condition) was given to be completed within 8 min, in examination conditions. After the time had elapsed, the papers were collected and redistributed to another student in the same class for marking. Correct answers were briefly presented on an overhead projector by the tutor. Students marked the paper out of 12 (two questions required two answers each) before handing back the paper to the tutor. This process took approximately 5 min. If participants wanted to know their specific mark, they were instructed to ask the tutor at the end of the class. For the restudy group, participants were administered a summary set to be read for 8 min. Participants were instructed to highlight or underline what they considered the key points of each summary sentence, and that no additional writing was to be made. There was then a 2-min period where students were invited to ask any questions about the material. At the end of this phase of the experiment, their summary sets were handed back to the tutor. The no-activity group was given no task for this phase.

Phase 3: Memory test. During the following week's tutorial, participants were given a pop quiz, of 8 min duration, under examination conditions. Participants were unaware that this quiz was to be administered, to reduce the chances of additional external study of the material. When the time had elapsed, participants handed in their papers which were then redistributed across the class for marking. The answers were briefly presented on the overhead projector by the tutor. A final mark out of 24 was calculated by the student marker before the papers were collected by the tutor.

Results and discussion

A significance level of .05 was used for all statistical analyses. Pairwise comparisons were Bonferroni-corrected to the .05 level.

Sample characteristics and manipulation check. The uniformity of conditions in terms of their demographic composition was checked using chi-square analyses, which indicated that there were no differences between conditions in terms of gender, $\chi^2(2, N = 218) = 218$, $p = .329$, phi = .101, or English/non-English native language, $\chi^2(2, N = 212) = 3.351$, $p = .187$, phi = .126.

As two different sets of quiz/restudy items were used in this experiment for the target and related materials, quiz difficulty levels could have varied. To determine if this was the case, an initial omnibus 3 (group: test, highlight, control) × (2) (test material: target, related) × 2 (quiz type: I vs. II) mixed ANOVA was conducted. There was a significant Test material × Quiz type interaction, $F(1, 212) = 10.94$, $MSE = 138.35$, $p < .05$, $p\eta^2 = .05$, and therefore subsequent analyses included quiz type as a factor.

Testing effect. It was hypothesised that the quiz group would perform better than both the restudy and no-activity groups on the target material in the pop quiz. Figure 3 suggests that as expected, for the target material, the quiz group performed better than both the no-activity and restudy groups, and there was little difference between the latter two groups. A 3 (group) × 2 (quiz type: I, II) ANOVA indicated a main effect for group, $F(2, 212) = 50.79$, $MSE = 449.89$, $p < .05$, $p\eta^2 = .32$. Pairwise comparisons confirmed that the quiz group performed better than did both the no-activity group, $t(215) = 9.72$, $SEM = 3.48$, $p < .05$, $d = 1.33$, and the restudy group, $t(215) = 7.00$, $SEM = 3.45$, $p < .05$, $d = .95$. The restudy group also performed better than the no-activity group, $t(215) = 2.69$, $SEM = 3.59$, $p < .05$, $d = .37$. There were no significant main or interaction effects for quiz type. In summary, with the increased power afforded by more participants in

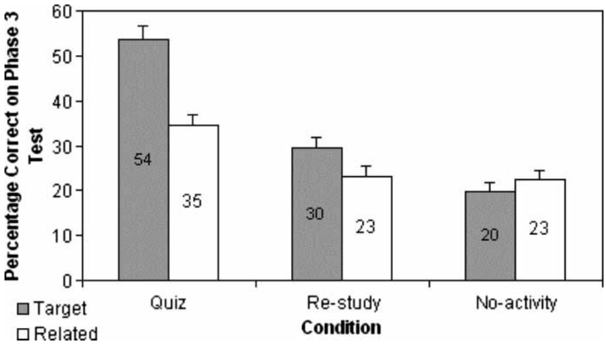


Figure 3. Mean percentage of target and related items correct on the Phase 3 test as a function of postvideo condition for Study 2. The error bars indicate the standard error of the mean.

each group, there was evidence for the individual testing effect according to both the lenient and strict criteria.

Retrieval induced facilitation versus retrieval induced forgetting. In line with the findings of Chan et al. (2006) a retrieval induced facilitation effect would be evident if performance on the related material is significantly better for the quiz group than both the restudy and no-activity groups with no difference between the latter two groups. Figure 3 suggests that, as expected for the related material, the quiz group performed better than the no-activity and restudy groups, with little difference between the latter two. A 3 (group) $\times$ 2 (quiz type: I, II) ANOVA on the related material indicated a main effect for group, $F(2, 212) = 9.40$, $MSE = 374.80$, $p < .05$, $p\eta^2 = .08$. Pairwise comparisons confirmed that the quiz group performed better than did the no-activity group, $t(215) = 3.72$, $SEM = 3.23$, $p < .05$, $d = .51$, and restudy group, $t(215) = 3.56$, $SEM = 3.21$, $p < .05$, $d = .49$, and there was no significant difference between the latter two groups. There was a main effect of quiz type, $F(1, 212) = 4.89$, $MSE = 374.80$, $p < .05$, $p\eta^2 = .02$ indicating that Quiz I performance was superior for related items. There was no significant interaction. These consistent patterns of findings suggest that the nonrandom assignment of participants to tutorial (and tutor) did not confound the results.

Alternative explanations. As for Study 1, it was possible that participants undertook additional study between Phases 2 and 3, which may have influenced final memory performance. A one-way ANOVA yielded no group differences (quiz $M = 2.02$, $SD = 2.67$; restudy $M = 2.61$, $SD = 3.71$; no-activity $M = 2.63$, $SD = 4.63$). Nevertheless, subsequent analyses of percentage items correct on the final test were also run with additional study hours as a covariate. All analyses yielded the same findings with or without the covariate included, so the covariate analyses are not reported here.

GENERAL DISCUSSION

Study 1 demonstrated the testing effect in a classroom setting in that individual quiz students performed better on the final test than did the no-activity group, who did not complete a Phase 2 quiz. However, the strict criterion for the testing effect was not met as the individual quiz students did not perform better than the restudy students. Given the importance of determining the potential limitations of generalisability of the test effect, Study 2 was designed to further investigate this issue by improving aspects of the methodology. In Study 1, the testing effect did not generalise to new (previously unseen) test items in that the individual quiz students did not perform better than the restudy or the no-activity students. Group quiz scores were superior to the individual quiz scores for both old and new items, as well as for the Phase 2 quiz, thus demonstrating a significant collaborative testing benefit.

Study 2 yielded strong evidence for the testing effect in the classroom setting in that individual quiz students performed better than both the restudy and no-activity students on the target (previously tested) items in the final test, suggesting that the lack of a strict criterion testing effect in Study 1 was related to lack of power. Moreover, the retrieval induced facilitation hypothesis was supported. Memory for related (untested) material was significantly enhanced as a result of testing (quiz) but not following simple rereading of the material (restudy). This finding contrasts with the Study 1 result of no difference in performance on new items for restudy and individual quiz groups. In Study 1, old (conceptually equivalent to target) and new (conceptually equivalent to related) quiz items were not counterbalanced, and new items were easier to answer than old items. Further research is required to investigate whether item difficulty interacts with facilitation of memory performance with related (untested) material. The memory performance of the restudy group did not differ significantly from the no-activity group, who did not engage in any revision of the video material.

The current research demonstrated the benefits of testing on later memory retention of both target and related material that was presented in an authentic tertiary educational setting. Additional hours of study between review and the final test did not influence the results in either study, suggesting that the findings are consistent with "direct" theoretical accounts of the testing effect such as the retrieval hypothesis. This theory states that repeated elaborative retrieval of the target material (e.g., testing) produces better retention. The proposed mechanism is that the effortful retrieval processes used at the time of encoding increase the strength of a memory trace and increase the number of available retrieval routes (Anderson et al., 1994; Butler & Roediger, 2007; Roediger & Karpicke, 2006a). Therefore, it is predicted that the more effortful or elaborative retrieval processes that are engaged in during the review phase, the better the information will be remembered at a later time. A test administered during the review phase is a more effortful form of learning than rereading (e.g., passive rereading or highlighting of the information). As such, the elaborative processes used in a practice test increase later memory performance (Roediger & Karpicke, 2006a).

Study 2 yielded results in favour of the retrieval induced facilitation hypothesis. These findings suggest that as a direct result of testing the target material, thematically related information retention is also enhanced after a delay between the retrieval practice and final retrieval phases (Chan et al., 2006). The retrieval induced facilitation effect obtained in Study 2 is in direct opposition to the retrieval induced forgetting effects obtained by other researchers (e.g., Anderson et al., 1994; Carroll et al., 2007; Macrae & MacLeod, 1999). For example, the current findings diverge from those of Carroll et al. (2007), who obtained a forgetting effect using similar prose materials in two studies. There was, however, an important difference between Study 2 in the current research and Carroll et al.'s studies: the length of the delay between the review phase and the final test. This delay is one of several boundary conditions influencing the likelihood of retrieval induced forgetting. Other studies support the postulate that retrieval induced forgetting appears to decay relatively quickly such that it is eliminated at longer delays (e.g., 24 hours; MacLeod & Macrae, 2001).

Theoretically this suggests that practice tests not only significantly increase learning of the tested information, but memory for related information is also increased when practice tests are taken a week before the final exam. It appears that in relation to the testing effect, elaborative retrieval processes activate schemas and semantic processes for related information as well as for the target information. Therefore, the related information does not appear to compete for activation, as previous retrieval induced forgetting research suggests (Anderson et al., 1994; Carroll et al., 2007). In contrast, a spreading of the activation of the semantic nodes and processes appears to yield a retrieval induced facilitation effect over a longer interval (Chan et al., 2006; Collins & Quillian, 1972). Since Study 2 included a 1 week delay between Phase 2 and the final test, the relative influence of retrieval induced facilitation and retrieval induced forgetting could not be examined in the "immediate" time period. Future research should incorporate an immediate condition in a real world educational setting to examine any differences in performance on the final memory test as a function of time between the retrieval practice and final recall phases.

The current research represents an extension of the work of Chan et al. (2006) because the testing effect and retrieval induced facilitation were examined in conjunction with corrective feedback. That is, in both studies, corrective feedback was provided after the practice quiz. This raises the possibility of different outcomes if no feedback is provided. Previous studies on the role of feedback suggest that it can influence the testing effect both directly and indirectly (Black & Wiliam, 1998). One possible indirect effect of feedback is that participants may undertake additional study because they did not perform optimally on the practice test, and want to improve (Fitch et al., 1951; Roediger & Karpicke, 2006a). This, however, does not appear to be the case in the current research because hours of (self-reported) additional study did not differ significantly between the groups. A potential direct effect is that feedback could function as another retrieval practice opportunity, thus further enhancing memory retention (Black & Wiliam; 1998; Roediger & Karpicke, 2006a). In contrast, without feedback, incorrect responses during the initial test may be assumed to be correct by the respondent, and this effect of misinformation may appear in the final memory test (Roediger & Marsh, 2005; Rogers, Goldstein, & Benassi, 2006). To explore these notions further, future research on the testing effect in a classroom setting should directly compare feedback and no-feedback conditions (Butterfield & Metcalfe, 2006; Kang, McDermont, & Roediger, 2007).

The current research also explored collaborative testing effects in the classroom. A previous classroom study by Cortright, Collins, Rodenbaugh, and DiCarlo (2003) reported a positive collaborative testing effect on subsequent individual student memory of course material; however, the collaborative manipulation was confounded with retrieval opportunity (i.e., the individual testing comparison group had not received an equivalent number of test exposures). One way in which to conceptualise the collaborative testing experience is in terms of more elaborative retrieval processing afforded by the group interaction, thus leading to the better later retrieval of both old and new items in Study 1. This notion needs to be tested more rigorously with better controlled materials and conditions, such as those employed in Study 2.

An alternative approach is from the social cognition literature, which has reported the collaborative inhibition effect, whereby collaborative group performance is worse than nominal group performance (Basden, Basden, Bryner, & Thomas, 1997; Weldon & Bellinger, 1997; Wright & Klumpp, 2004). Studies by Meudell and colleagues (e.g., Meudell, Hitch, & Boyle, 1995) suggest that the collaborative inhibition effect is reduced if individuals first respond individually immediately prior to discussing and responding within a group testing situation. These and other issues, such as the inclusion of misinformation (Clark, Abbe, & Larson, 2006; Maki, Weigold, & Arellano, 2008), deserve further rigorous testing within the classroom setting. Overall, however, it can be concluded that the individual usually benefits in the long term from the pooling of information during an interim collaborative testing episode.

Practical and educational implications

The current research provides evidence that a practice test yields better performance on a later test than no additional study (Studies 1 and 2) and also compared to restudy of the original material (Study 2). Importantly, the benefits of testing in the review phase generalised to semantically related (but untested) material in Study 2. In other words, contrary to the retrieval induced forgetting hypothesis (Anderson, 2003; Anderson et al., 1994; Carroll et al., 2007), the findings of Study 2 suggest that studying does not adversely affect memory for the unstudied material when there is a significant delay between review and the final test. The findings of Carroll et al. (2007) suggest that forgetting is most likely to occur for the related, unstudied information when the final exam is taken immediately after retrieval practice. Hence, early, spaced (rather than massed) testing during review appears most likely to facilitate (rather than inhibit) memory for related material. Taken together, the findings highlight the efficacy of repeated testing as a learning tool and not just as a means of assessment (see also McDaniel et al., 2007). This indicates that regular testing would be beneficial to learning in undergraduate university courses.

The presence of feedback may aid in the learning process, but further research is needed to disentangle the effects of feedback and the testing effect. The current findings suggest that formative assessments should be used as practice tests with immediate corrective feedback in the lead up to a final summative exam. Moreover, the level of motivation of the students during a practice quiz may be important for their learning and subsequent performance on a final test (Velan & Kumar, 2007), such that higher motivation produces better performance on a final test. The practice tests, therefore, should appear necessary to students' learning and future final test performance in order to increase motivation to perform well in the practice tests and to learn from the feedback provided.

The benefits of collaborative testing compared to individual testing on later memory retention were evident in Study 1. Collaborative testing produced higher scores on the practice test, but more importantly, it produced superior performance on both "old" and "new" items in the final test. To our knowledge, this research is the first to examine the impact on the testing effect of collaborative testing. The findings are consistent with previous research that has demonstrated that collaborative learning situations enhance students' understanding of and later memory for psychological concepts (e.g., Miyake & Shirouzu, 2006). In a group situation, students have the opportunity to collaboratively read, explain, exchange, and discuss the material. This pooled information may enhance the integration of the material in memory and assist students to form a cohesive and abstracted view of the key concepts (Miyake & Shirouzu, 2006). The current findings highlight that substantial collaborative learning opportunities in school and university curricula will encourage in-depth, process-oriented learning. A fruitful avenue for future research will be to examine the mechanisms underlying the beneficial effects of collaborative learning on later memory retention.

CONCLUSIONS

The present study found evidence in support of the testing effect in a classroom setting. The substantial benefits of collaborative learning on later memory retention were also demonstrated. It appears that as a result of effortful retrieval processes from a practice test on a target subset of information, performance on a final memory test is improved not only for the target subset, but also for semantically related information. Corrective feedback appears to be important although future research is required to examine precisely how feedback influences memory retention. More research is required to further examine the retrieval processes operating during these different retrieval practice opportunities and how they influence performance on a subsequent final memory test.

Original manuscript received December 2007
Revised manuscript received May 2008
First published online October 2008

REFERENCES

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *The American Psychologist*, *51*, 355–365.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1063–1087.

Basden, B. H., Basden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, *14*, 497–507.

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, R. L., III. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 1176–1189.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, *5*, 7–74.

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*, 514–527.

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*, 69–84.

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., 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.

Clark, S. E., Abbe, A., & Larson, R. P. (2006). Collaboration in associative recognition memory: Using recalled information to defend "new" judgements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *6*, 1266–1273.

Cockrell, K. S., Caplow, J. H., & Donaldson, J. F. (2000). A context for learning: Collaborative groups in the problem-based learning environment. *Review of Higher Education*, *23*, 347–363.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory* (pp. 117–137). New York: Wiley.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiological Education*, *27*, 102–108.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. C. Carterette & M. P. Friedman (Series Eds.) and E. L. Bjork & R. A. Bjork (Vol. Eds.), *Handbook of perception and cognition: Vol. 10. Memory* (2nd ed., pp. 317–344). New York: Academic Press.

Fitch, M. L., Drucker, A. J., & Norton, J. A. (1951). Frequent testing as a motivating factor in large lecture classes. *Journal of Educational Psychology*, *42*, 1–20.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558.

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*, 279–308.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97.

Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, *109*, 451–464.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, *12*, 148–152.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*, 463–473.

Maki, R. H., Weigold, A., & Arellano, A. (2008). False memory for associated word lists in individuals and collaborating groups. *Memory and Cognition*, *36*, 598–603.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494–513.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192–201.

McDaniel, M. A., Friedman, A., & Bourne, L. E. (1978). Remembering the levels of information in words. *Memory and Cognition*, *6*, 156–164.

Meltzoff, J. (1998). *Critical thinking about research: Psychology and related fields*. Washington, DC: American Psychological Association.

Meudell, P. R., Hitch, G. J., & Boyle, M. M. (1995). Collaboration in recall: Do pairs of people cross-cue each other to produce new memories? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *48A*, 141–152.

Miyake, N., & Shirouzu, H. (2006). A collaborative approach to teaching cognitive science to undergraduates: The learning sciences as a means to study and enhance college student learning. *Psychologia*, *49*, 101–113.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behaviour*, *16*, 519–533.

Raaijmakers, J. G. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science*, *27*, 431–452.

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

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

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

Roediger, H. L., III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 803–814.

Rogers, R. A. ,Goldstein, G. D., & Benassi, V. A. (2006). *Using tests to promote retention of information acquired while watching a video*. Paper presented at the 29th annual conference of the National Institute on the Teaching of Psychology, St. Petersburg Beach, FL.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210–221.

Velan, G., & Kumar, R. (2007). *Integrated online formative assessment in the biomedical sciences*. Paper presented at the International Society for the Scholarship of Teaching and Learning, Sydney, Australia.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 1160–1175.

WGBH Boston (Producer), in association with the American Psychological Association. (2001). *Discovering psychology: The behaving brain* [Videotape]. Boston: American Psychological Association.

Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*, 240–245.

Wright, D. B., & Klumpp, A. (2004). Collaborative inhibition is due to the product, not the process, of recalling in groups. *Psychonomic Bulletin and Review*, *11*, 1080–1083.