1-1-2005

The Testing Effect: Using Retrieval Practice in the Classroom

Karen S. Clifton

THE TESTING EFFECT: USING RETRIVAL PRACTICE IN THE CLASSROOM

Thesis submitted to
The Graduate College of
Marshall University

In partial fulfillment of the
Requirements for the degree of
Master of Arts
Psychology

by

Karen S. Clifton

Steven P. Mewaldt, Chair
David L. Trumpower
Marc Lindberg

Marshall University

Summer 2005

ABSTRACT

Repeated testing was applied in a college classroom setting to determine whether a single intervening test, which allowed for retrieval practice, would improve performance on a final test compared to a single structured rehearsal of the material. Performance was measured using multiple-choice exams and relatedness rating tests. The findings suggest that a test condition which requires retrieval, when compared to a read-only condition, improves performance on a final test of item-specific knowledge but not on a test of relational knowledge of the same material. The difficulty of retrieval was manipulated using hard and easy questions on the intervening test and did not appear to have an effect. The findings support the use of repeated testing in the classroom and supplies evidence of the beneficial use of retrieval practice in enhancing student learning of classroom material.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Steven Mewaldt, Dr. David Trumpower, the undergraduate research team, and all the instructors and students who participated in the study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

The testing effect: Using retrieval practice in the classroom

It has been shown that repeated testing can produce hypermnesia, an increase in the amount of information retrieved across multiple recall opportunities (Burns & Gold, 1999; Glover, 1989; McDaniel, Moore, & Whiteman, 1998; Mulligan, 2001, 2004; Spitzer, 1939; Wheeler & Roediger, 1992). In the laboratory there are robust findings that show repeated testing consistently increases recall of material by either aiding item recovery, slowing the rate of forgetting, or a combination of both. The testing effect has been found using many different types of procedures (e.g., recall, cued-recall, and recognition). These findings suggest repeated testing has a positive effect on memory that might be beneficial in a classroom setting. However, little literature was found that applied the concept of repeated testing to an educational setting. The present study attempted to apply the testing effect to the classroom to improve student's retention of course material.

Repeated testing can produce hypermnesia by increasing item recovery, i.e., the retrieval of new information not previously retrieved on a prior test, and decreasing item loss, i.e., the inability to recall information previously recalled on an earlier test. When item recovery is greater than item loss the result is hypermnesia (Mulligan, 2001). Item-specific processes should increase the discriminablity of an item increasing item recovery, while relational processes should increase the accessibility of an item decreasing item loss. Memory enhancement through repeated testing should be dependent on the discriminabilty of an item from other items and the accessibility of an item in memory (Burns & Gold, 1999).

Encoding

Relational encoding. Tulving and Donaldson (1972) proposed that relational information about an item is organized around the common features the item shares with the other items in the learning episode thereby increasing the accessibility of the item and protecting the item against being forgotten (Hunt & Einstein, 1981; Mulligan, 2001). The organization of relational information in the memory system encourages one memory trace to trigger a number of related memory traces. This facilitates item recovery. During early tests relational processes benefit retrieval by making the items more accessible and less likely to be forgotten through the use of effective search strategies.

Item-specific encoding. When item-specific processes are used to encode an item, attention is paid to the distinctiveness of the unique features of the item (Hunt & Einstein, 1981).

The levels-of-processing approach explains how item-specific encoding increases item recovery (Craik & Lockhart, 1972). During encoding when attention is paid to the distinctiveness of an item it is presumed that a "deeper", more thorough, encoding of the items' attributes occurs. If retrieval of an item is dependent on the retrieval of a certain number of those attributes, it is possible that on an initial test, retrieval of the item does not occur because not enough of the item's attributes were retrieved. However, on later tests retrieval of the item may occur because additional attributes or a different critical set of attributes may be retrieved (Mulligan, 2001). Klein, Loftus, and Schell (1994) argued that item-specific processes get in the way of creating a consistent search strategy for the item in memory. The inconsistency in the search strategy allows for differences in the retrieval of an item's attributes, and consequently creates variability in the likelihood of an item's recovery at each attempt to remember. However, McDaniel, Moore, and Whitman (1998) found that after an item-specific processing task, when the consistency in the retrieval plan was controlled by giving category cues on the test, no difference occurred for item recovery compared to when no category cues were given and the retrieval plan was presumably less consistent (Experiment 1 & 2). However, when they controlled for the number of attributes (1-3) generated in the item-specific processing task their findings were inconclusive. When three attributes were generated the number of items recalled by the end of the test cycle was greater than when only one attribute was generated during the item-specific task, but the number of attributes generated did not increase item-recovery between tests. This could have occurred because by the end of the test cycle recall approached ceiling levels for items in the three-attribute condition (Experiment 3).

Retrieval

While the relational/item-specific framework emphasizes how an item is encoded, the retrieval of an item during repeated tests may be just as important. Roediger & Payne (1987) suggested that the benefits of repeated testing occurs because of the additional retrieval time created by the repeated tests. They suggested that the retrieval of items on an early test makes later retrieval of that item less difficult creating additional time for the retrieval of additional items on later tests. Similarly, Erdelyi and Becker (1974) proposed that repeated tests create additional time to generate items, which may result in an increase in the number of items retrieved across tests.

Bjork and Bjork (1992) proposed that there are two components that influence the recallability of an item, its storage strength (i.e., encoding) and its retrieval strength. They suggested that repeated testing forces repeated retrieval of the to-be-remembered items and therefore increases recall of the items by increasing the item's retrieval strength. Kazen and Solis-Macias (1999) have similarly argued that an increase in recall across testing is due to repeated retrieval opportunities.

Although the relational/item-specific framework focuses on encoding, it does not discount the importance of retrieval. For example, McDaniel et al. (1998) proposed that after an item is retrieved, the retrieval path is strengthened regardless of how the information was initially encoded.

Repeated Testing in the Classroom

Many have debated the effectiveness of repeated testing in education and other applied pursuits (Glover, 1989; Glover & Krug, 1990; Roediger, Karpicke, & Marsh, 2003; Wheeler, Ewers, & Buonanno, 2003). The debate stems from the fact that while researchers have consistently been able to produce hypermnesia using the recall of lists of words or pictures (Burns & Gold, 1999; Burns & Schoff, 1998; Glover, 1989; Klein et al. 1989; McDaniel et al., 1998; Wheeler et al., 2003; Wheeler & Roediger, 1992), studies using prose or recognition testing, which more closely resembles the type of learning and testing that take place in the classroom, have been less consistent in producing hypermnesia.

An argument against the ability of prose material to produce hypermnesia is that one is more likely to confuse prior false retrieval on early tests with the original encoding of the prose material (Fritz, Morris, Bjork, Gelman, & Wickens 2000). Bartlett (as cited in Wheeler & Roediger, 1992) found that participants who made errors when tested on prose material continued to make the same types of errors on later tests. Barlett argued that this was likely due to the fact that the participants had no further contact with the prose and were thus consistent in their recall of the material. However, later studies found that even when participants receive additional time post recall with the prose material, there was little change in performance across tests (Fritz, et al., 2000; Howe, 1970; Kay, 1955). After a participant had made an error in retrieval they continued to make the same type of error even after the additional study time with the material. Fritz et al. suggested two plausible explanations for why the additional study time with the prose did not facilitate additional learning. They argued that because later tests are more similar to earlier tests than to the additional presentation(s) of the material, participants during late tests were more likely to refer to their previous recall, as opposed to the later presentation(s) of the material. If source confusion occurred during recall, then the participants would be unable to determine the source of the material during recall, i.e., they might confuse prior retrieval with the prior presentation of the material. Frost, Ingraham, and Wilson (2002) agreed and found that when participants were confused about the source of the to-be-remembered material, they typically attributed information from prior acts of retrieval to information in the text.

As with prose recall many attempts have been made to produce hypermnesia using recognition testing with inconsistent results. While some studies have been successful in demonstrating hypermnesia using recognition tests (Erdelyi & Stein, 1981; Kazen & Solis-Marcias, 1999; Roediger et al., 2003), others have not (Payne & Roediger, 1987; Otani & Hodge, 1991; Otani & Stinson, 1994). Roediger, et al. (2003) proposed that one reason for these inconsistencies could be that a miss on a recognition test could create a memory trace back to the wrong information and on a repeated test, participants would repeat the mistake, thereby strengthening a link to the wrong information. Otani and Stinson (1994) attempted to alleviate this problem by creating an overtly easy test where a miss on the early test was almost impossible. However, using this strategy with repeated testing created a different problem. Using easy recognition tests early in the test cycle produced a ceiling effect on the earlier tests which resulted in little item recovery or retrieval practice, and hypermnesia was not found (Otani & Stinson, 1994). A solution to prevent the student from learning the wrong information without creating too easy a task might be to give immediate feedback during the test cycle. Feedback given within a test cycle would likely be recalled on later tests because of the similarity between the test cycle with feedback and the later test cycle. Immediate feedback could also affect the retrieval link. If feedback is given during the same episode in which retrieval occurred, it is possible that the corrective feedback could block the consolidation of the incorrect links created through the retrieval process. Another possibility is that the feedback would make the retrieval link to the wrong information distinct in memory, so that on subsequent tests one might remember that the link was incorrect and be motivated to correct the previous mistake.

A review of the literature suggests that repeated testing could be applied to a classroom, even if some minor alterations are needed to assure successful learning. The present study was an attempt to determine whether repeated testing would be more beneficial than repeated study in enhancing the retention of course material in a college classroom. Few studies have compared repeated testing with repeated study as a means of enhancing memory. Wheeler et al. (2003) used repeated test and study trials to measure differences in the rate of forgetting between the two encoding conditions. Participants listened to a list of unrelated nouns with the instructions that they should try to memorize the lists. Participants were then either given repeated opportunities to recall the list by writing the words on paper, or to re-listen to the words on a cassette tape. Participants in the repeated study condition demonstrated a higher degree of learning when tested within minutes of the initial learning episode, but when the retention intervals were increased to 48 hours in Experiment 1 and one week in Experiment 2, the relationship was reversed as recall was now greater in the repeated test condition. Wheeler et al. hypothesized that retrieval in the repeated test condition protected the information from being forgotten. Mulligan (2001) performed a similar study comparing generated items with read items. Mulligan was looking specifically for the presence of hypermnesia and found that in Experiments 1-3 hypermnesia occurred in the generated condition but did not occur in the read condition. In Experiment 4 Mulligan did find hypermnesia in the read condition but to a lesser degree than was found in the generated condition.

The goal of the present study was to determine whether repeated testing could be applied to increase the retention of classroom material. Although the present study does not assess hypermnesia, per se, it was hypothesized that if repeated testing can create hypermnesia in the laboratory, then repeated testing should also improve retention in the classroom. In the present study the to-be-remembered information was presented to students in the learning phase through lecture and text, and then the students completed a computer exercise in which they either read a statement repeating a concept in the course, answered an easy or hard question testing previous learning of the concept, or as a control condition were not given any additional presentation of the material. Immediately after reading the study statement or answering the hard or easy questions students were given feedback by the repeating of the study statement or the question stem accompanied by the correct answer to the question. Later, students were tested for their retention of the items on a multiple-choice unit exam. The procedure in which participants were tested with easy or hard questions and given feedback before the unit exam was referred to as a study-test-study-test (STST) method. The procedure in which participants were presented material as a statement on the computer and then were re-presented the same material during the same computer exercise before the unit exam was referred to as a study-study-study-test (SSST) method. We evaluated the benefits of repeated testing by comparing STST and SSST. In addition, in the STST condition we employed both hard and easy questions to evaluate which type of practice items would be more beneficial in the classroom; hard questions which should have produced a retrieval challenge but opened the door to source monitoring errors, or easy questions which might have produced minimal retrieval practice during the computer testing phase, but should have reduced retrieval errors.

We hypothesized a testing effect would occur and that students would perform better on the unit exam on material presented through STST than presented through SSST, and that performance would be greatest for items in the hard question condition. It was thought that the possibility of a ceiling effect with the easy questions on the quiz would have a negative influence causing retrieval practice with the easy question items to be minimal, thereby reducing the testing effect. It was also thought that immediate feedback given during each quiz would either block the consolidation of the retrieval link for missed items or would sensitize the student to the missed items in the hard condition, keeping source monitoring errors to a minimum.

We felt that it was also important to understand what type of learning was occurring in the classroom, and if the different strategies affected the type of learning that occurred. The present study, therefore, evaluated the testing effect for both item-specific and relational knowledge of the target items.

Method

Participants

Participants were 283 undergraduate students enrolled in introductory psychology classes at Marshall University. Students participated in exchange for extra credit in the course. Participants received extra credit for each quiz or relatedness rating test they took regardless of the student's performance. The unit exams, of course, were a normal part of the class and counted as half of their overall grade for the course. Eight introductory psychology classes, each containing approximately 40 students, were tested. Six of the eight classes met on a three-day-per-week schedule, while two met twice per week. While most students completed all parts of the experiment, some students missed parts of the experiment, as would be expected in an experiment involving close to 300 students and approximately 12 weeks of testing.

Materials

The target items consisted of 72 psychological concepts introduced in three separate units during a 15-week semester. In each unit, participants were presented with many novel concepts, including 24 target concepts. The concepts were presented at least once during class lecture and were available to students throughout the course in the textbook, Psychology 6th ed., by David Myers (2002).

WebCT 3.8 software was used to present the target items in six online quizzes. Each course was divided into four units, which were approximately three to four weeks in duration, and each unit ended with a unit exam. During each of the first three units, one quiz was made available to the students early in the unit and one was available late in the unit. The fourth unit was not included as part of the study. The format of the online quizzes was consistent across the semester. Each quiz contained nine items: three were presented in hard test questions, three were presented in easy test questions, and three were presented as study statements. In addition, there were three items for each quiz which were not presented on the quiz but were tested on the unit exam (no study items). When a hard test question was used to present a target item, the question was similar to what the student would later see on the subsequent unit exam. A fill-in-the-blank question was presented and the target item was the correct choice among three other likely, but wrong, answer choices. The hard test foils were derived from other items presented in class and were chosen to make answering the question difficult. The easy form of a test question used the same fill-in-the-blank question as in the hard test condition. In addition, the correct answer choice appeared in the same answer slot as in the hard test question. The only difference between the hard and easy test question was in the foils. In the easy test, the foils consisted of illogical choices and non-psychology items. In the study condition students were again presented with the same fill-in-the-blank question, but this time the target item was inserted into the question statement and the target item was presented in bold-type. Incorrect foils were not included. Items in the no study condition were not included in the quiz (see Appendix A for a complete quiz).

For each quiz, encoding method (hard test, easy test, study, no study) was arranged, so that on a quiz a set of three target items was presented in each of the four encoding methods to each participant. Across participants, items were rotated through encoding methods so that each set of target items appeared in all four encoding methods on a quiz. This created four forms of each quiz created to correspond with the four groups of participants. In addition, within each set of items the items were randomized to create three distinct orders in which items could appear on each quiz. Items appeared in all three orders on each of the four forms of the quiz creating 12 version of each quiz.

Item-specific knowledge of the 72 target items was measured through three multiple-choice unit exams. Each of the three course units included in the study ended with a 50-question multiple-choice exam which tested 12 target items from the early quiz, 12 target items from the late quiz, and 26 items from the unit that were not part of the study. The multiple-choice exams consisted of conceptual and definitional type questions that were taken from the test banks that accompanied the text. It was important to assure that on the unit exam it was the target item that was being recognized and not just the target question from the quiz. Therefore, each of the 24 target items per exam the question stems on the quiz and the question stems on the multiple-choice exam were worded differently (see Appendix B for question comparisons).

Relational knowledge of the 72 target items was measured using six relatedness-rating tests. Each multiple-choice exam was followed by one of two relatedness-rating tests. The two relatedness rating tests for each unit contained either the 12 items from the early quiz or the 12 items from the late quiz. Within each of the relatedness rating tests there was one subtest containing items from each of the four encoding methods: hard test, easy test, study, and no study. Within each subtest, in addition to the three target items from the quiz, two highly related words per target item were taken from the course material and were included for a total of nine items per subtest. The nine items were listed in a random order at the top of the page. The items were then paired with every other item creating a total of 36 unique pairs of items that students were asked to rate on a 5-point scale, with 1 = highly unrelated and 5 = highly related. Each subtest contained two columns of items with each line containing an item pair on the left of the page followed by the rating scale to the right. The four subtests, which represented the four encoding methods, were combined to create one relatedness rating test with 144 unique pairs of items. The order in which the item pairs appeared on the page and the left-right order of the items was randomized using the Knot (Knowledge Network Organizing Tool) program. Two versions of each subtest were created with the item pairs in two distinct random orders. For each complete test, the four subtests were counterbalanced so that the subtests would appear in eight different orders. This created 16 versions of each relatedness rating test.

Design

The experiment employed a 4 (Encoding Method) x 2 (Time of Presentation) within-subject design. Encoding method consisted of hard test, easy test, study statement, or a no study condition, and the time of presentation variable consisted of the early and late quiz conditions. The dependent variables were performance on the multiple-choice unit exams and the ratings on the relatedness rating tests.

Procedure

During each of the three units, which lasted an average of three and a half weeks each, participants were asked to complete two online quizzes. One quiz was made available to students midway through the unit and covered material presented in the early part of the unit. The second quiz was made available to students just prior to the multiple-choice exam date and covered material presented in the second half of the unit. For each quiz students were given on average a four-day period of time to complete the quiz. Students were able to complete the quiz at any time within the time period. Students were advised to take the quizzes in a quiet location away from distractions and to read each question carefully before selecting an answer. Each quiz question appeared on the screen one at a time. Using the mouse, students had to click on an answer choice and to click a button to save the answer choice before moving on to the next question. At the end of the quiz students clicked a button to submit the entire quiz for grading. Immediately, after the quiz was submitted students were given feedback on their performance. Students were able to review the quiz in its entirety with both their answer choices and the correct answer choices highlighted for their consideration. Students were allowed only one opportunity to access the quiz during the allotted time period. After starting the quiz, students had to complete the quiz during the same computer session, but they had unlimited time during the computer session to complete the quiz.

The two quizzes per unit were administered to create an early and late condition. Therefore the retention interval between the early quiz and the multiple-choice unit exam was longer than the retention interval between the late quiz and multiple-choice unit exam. However, the nature of the study, as it was applied to an actual classroom setting, did not allow for precise control of the retention interval in the early and late conditions. The retention interval in the early condition varied slightly in accordance with the length of each unit and when the student opted to take the quiz, but it was approximately two weeks in length. The retention interval in the late condition did not vary according to the length of the unit and was always less than five days.

The multiple-choice unit exams were given in the classroom and were paper and pencil exams. Students were instructed to read each item carefully and answer the questions to the best of their ability. Students indicated their answer choice by filling in circles on an answer sheet and were given approximately one hour to complete each exam. During the next class session after completion of each multiple-choice exam, students received a relatedness rating test. The relatedness rating test was also given in a paper and pencil format. Half the students completed a relatedness rating test that was based on the early quiz items and half completed a test that was based on the late quiz items in each unit.

Students were asked to rate each item pair according to their level of relatedness, using the 5-point scale described above. Students were instructed that the purpose of this test was to record their initial impressions of relatedness between the pairs and that they should not spend more than a few seconds on any particular pair. It took students approximately 20 min to complete the rating task.

Results

The validity of the difficulty manipulation on the quizzes was checked by calculating the number correct out of three for both the easy and hard quiz items for each participant. A paired t-test confirmed a significant difference between performance on the hard and easy quiz items, $t(1166) = -24.67$, $MSe = .027$, $p < .001$. Responses to hard quiz items averaged 2.1 items correct compared to 2.8 items correct for the easy quiz items.

Item-Specific Knowledge

A restriction was placed on the unit exam data, such that the inclusion of a response in the data set for each participant was conditional on that participant’s completion of the corresponding quiz. This procedure was necessary because not all 283 participants took all 6 quizzes. For each participant, the total number correct on the three unit exams for the target items in each encoding condition was tallied for both the early and late quizzes. The sums were then divided by the corresponding total number of early or late quizzes completed for that participant. This resulted in the maximum mean of three correct in each encoding condition for each quiz time (early or late). Performance on the target items on the multiple choice unit exams was then analyzed using a 4 (Encoding Condition) x 2 (Time of Presentation) within-subjects ANOVA. A main effect (Figure 1) was found for encoding condition, F (3, 741) = 8.25, MSe = .360, $p < .001$. Examination of the means revealed better performance on items in the hard ($\bar{X} = 2.17$) and easy ($\bar{X} = 2.17$) conditions than on those in the study ($\bar{X} = 2.06$) and no study ($\bar{X} = 2.02$) conditions. Follow-up analyses indicated performance on hard and easy items was significantly better than the performance on items in the study condition, $p < .01$, for both. Performance on the hard and easy items was also better than performance on items in the no study control condition, $p < .001$, for both. Overall, performance in the STST conditions was 5.1 % better than in the SSST conditions and 6.9% better than in the no study condition. There were no significant differences between performance on items in the hard and easy conditions or between the study and no study conditions, $p > .05$.

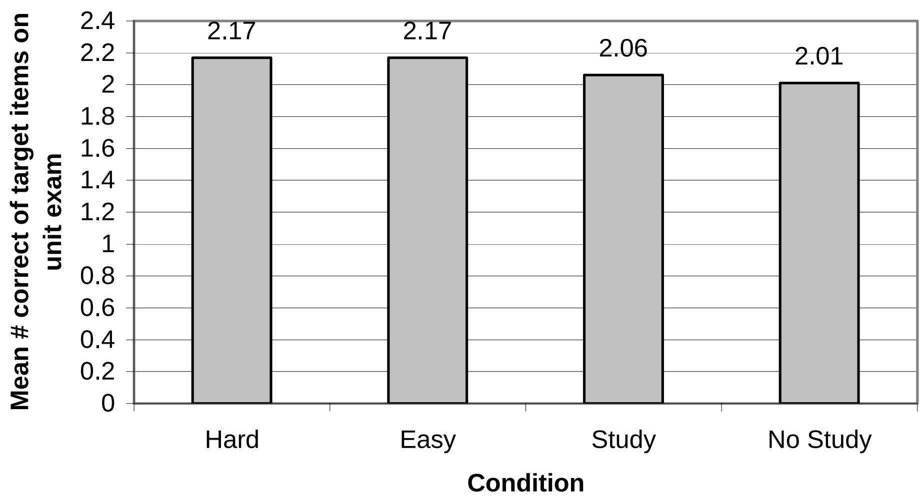

Figure 1. Main effect for encoding condition for the multiple-choice unit exam

No main effect occurred for time of presentation, F (1, 247) =1.85, ns. However an interaction between encoding condition and time of presentation was obtained, F (3, 741) = 4.85, MSe= .322, $p < .01$ (see Figure 2). Performance on items in the hard and easy conditions was better than performance on items in the study and no study conditions when the quiz was presented midway (early condition) through the unit, but performance on items in the study and no study conditions did not differ from performance on items in the hard and easy conditions when the quiz was presented just prior (late condition) to the unit exam. In the early quiz condition a simple effect was found, F (3, 741) = 15.03, MSe =. 276, $p < .001$, here performance on items which followed hard ($\overline{X}$ = 2.21) and easy ($\overline{X}$ = 2.17) quiz items were better than the performance on items which followed study ($\overline{X}$ = 1.98) and no study ($\overline{X}$ = 1.96) quiz items. For the early quiz the hard quiz items produced an 10.4% improvement in performance over the study condition and a 11.3% improvement over the no study condition, while the easy quiz questions improved performance 8.8% and 9.7 % respectively. In the late quiz condition no effect was found for encoding condition, $F < 1$.

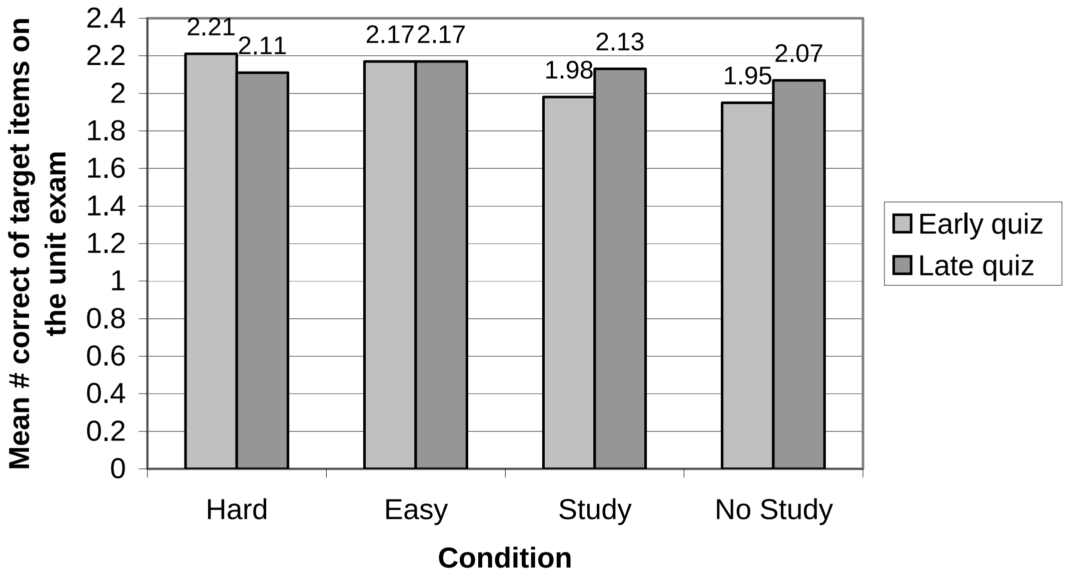

Figure 2. Interaction for encoding condition across time for the multiple-choice unit exam.

<u>Relational Knowledge</u>

The pair-wise ratings on the relatedness rating tests were converted into proximity matrices using the Knot Program. A proximity matrix was created for each subtest on the relatedness rating tests and contained all the pair-wise ratings between the items in each knowledge domain (the items within a condition and the added related terms). The Knot program uses the proximity matrix to read the distance between items, and assigns weights to relationships between the items. The goal is to find the shortest "path" between two items. The assigned weights are used to determine the shortest path and, hence, the most salient link between two items. If the shortest path between two items is a non-direct link then the direct link between those two items is deleted. Information about how each learner structured the knowledge was abstracted from the proximity matrix and represented graphically as a knowledge network which contained links and nodes. In the knowledge network the links represent the relationships between items and are weighted according to the strength of the relationship between the items. The nodes represent the target items. Similarity between the students' networks and an averaged expert network was obtained by dividing the number of links each network had in common by the number of links contained in either network. This measure of similarity is referred to as NETSIM.

Two faculty members in the psychology department and two graduate students teaching the introductory psychology course served as the experts for this study. Experts' ratings on the relatedness rating tests were converted into proximity matrices and measures of similarity were obtained among the experts using the procedure described above. Expert ratings with a NETSIM of .61 to .83 were used to create the averaged expert network for each set of items. Expert ratings which fell outside of this range were not included in the averaged expert ratings. For each item in the matrix an indirect measure of relatedness was obtained by correlating each item's rating with the rating of other items in the matrix. The indirect measure of relatedness was then correlated with original proximity data resulting in a measure of internal consistency called coherence, C. Coherence ratings close to one indicate a stable knowledge network, while coherence ratings further away from one indicate a less stable network and a decrease in the reliability of NETSIM. Coherence ratings for averaged expert networks ranged from C = .58 to .89 with a mean of C =. 78, suggesting there was internal consistency within the expert networks.

The NETSIM results were analyzed using a 4 (Encoding Condition) X 2 (Time of Presentation) split-plot ANOVA with encoding condition as the within-subject factor. Five of the six relatedness rating tests were analyzed. One relatedness rating test was left out of the analysis due to an error in the test construction, (some of the terms on the relatedness rating test did not match the target items from the corresponding quiz). As the semester progressed some students began to not take the relatedness rating tests seriously, as their answer patterns appeared random. Therefore, we dropped from the analysis any tests in which the coherence rating on three or more of the subtests were negative. Negative coherence ratings on the subtests suggested random responses. It was acknowledged, however, that a negative coherence rating could occur by chance, so any relatedness rating test with at least two positive coherence ratings was included in the analysis.

Using the NETSIM correlations between the students and experts a main effect was found for encoding condition F (3, 1653) = 4.41, MSe= 1.40, $p<.05$ (see Figure 3). NETSIM correlations for the item ratings which followed the study ($\overline{X}$ =. 354) and no study ($\overline{X}$ = .345) conditions on the quiz were the most like the experts, followed by the NETSIM correlations for item ratings which followed the hard ($\overline{X}$ = .334) and easy ($\overline{X}$ = .331) conditions on the quiz. Follow-up comparisions indicated that the NETSIM correlations for item ratings following the study condition were significantly different from the NETSIM correlations for item ratings which followed the hard and easy conditions on the quiz, while the NETSIM correlation for item ratings which followed the no study condition was significantly different from the NETSIM correlations for the items which followed the easy condition, but not significantly different than the NETSIM correlations for the items which followed the hard condition.

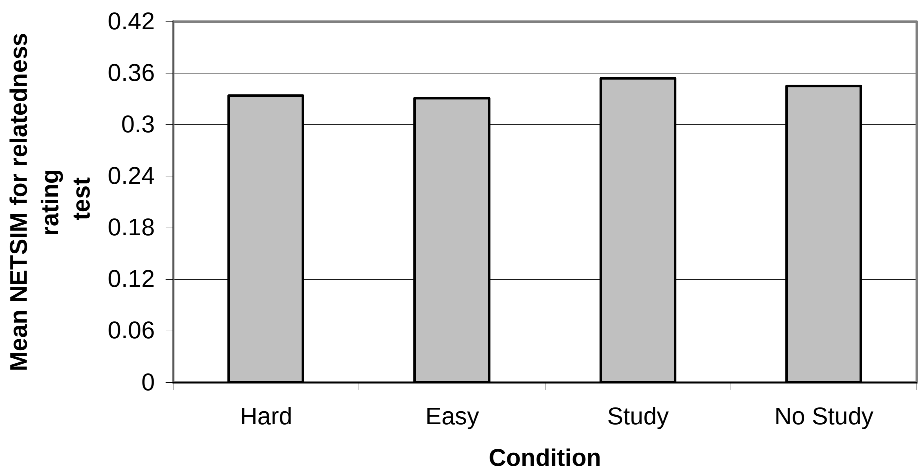

Figure 3. Main effect for encoding condition for the relatedness rating test

There was also a main effect for time of presentation F (1, 551) =149.05, MSe = 2.59, $p<.001$, with NETSIM correlations being greater for item ratings which followed the early quiz condition ($\overline{X}$ = .375) compared to the NETSIM correlations of item ratings which followed the late quiz condition ($\overline{X}$ = .307). Figure 4 shows the interaction between encoding condition and time of presentation, F (3, 1653) = 2.88, MSe = 4.06, $p < .05$. In the early quiz condition a simple effect was found for encoding condition, F (1, 762) = 3.94, MSe = 1.76, $p< .01$. The NETSIM correlation for the item ratings which followed the easy condition (.353) was significantly lower than the NETSIM correlation for the item ratings that followed the hard

(.379), study (.389) and no study (.382) conditions, all of which did not differ significantly from one another. In the late quiz condition, encoding condition was not significant, F = (1, 687) = 1.72, ns.

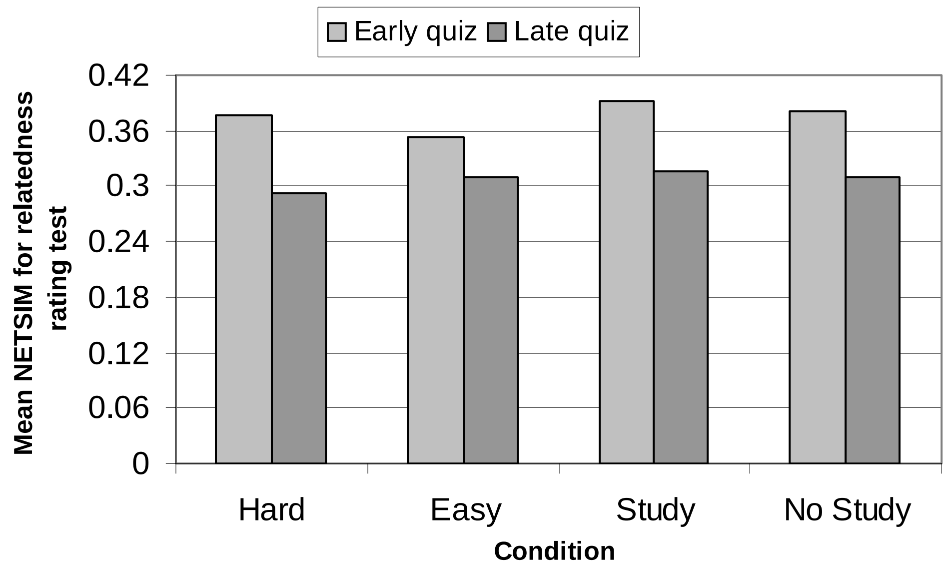

Figure 4. Interaction for encoding condition across time for the relatedness rating test

The NETSIM correlations for the item ratings from the relatedness rating tests were also compared with the number correct on the multiple-choice unit exams. Table 1 shows that across the early and late quiz conditions the NETSIM correlations of the item ratings and performance on the target items on the multiple-choice exams were significantly correlated. While all correlations between the two types of testing were significant, the strength of the correlations were weak.

Table 1. Correlations for NETSIM correlation from the relatedness rating tests and performance on target items on the unit exams

Encoding Condition	Hard	Easy	Study	No Study
Hard	**.278****	.198*	.146*	.150*
Easy	.093*	**.153****	.070	.110
Study	.169**	.093*	**.288****	.150*
No Study	.150**	.140**	.177**	**.139***

Note: Diagonals are correlations between NETSIM correlations and item performance on target items on the unit exam. The upper triangle contains correlations between encoding conditions for performance on target items on the unit exam. The lower triangle contains correlations between encoding conditions for NETSIM correlations from the relatedness rating test.
* $p < .05$, ** $p < .01$

Discussion

The purpose of the present study was to evaluate the practical application of repeated testing in a classroom setting. We hypothesized that a study method we termed the study-test-study-test (STST) method would increase retention of classroom material, relative to a rehearsal method, we termed study-study-study-test (SSST), by allowing for retrieval practice during the learning phase. We compared the STST method and the SSST method across time and measured the effect on multiple-choice unit exams (item-specific knowledge) and relatedness rating tests (relational knowledge).

Item-Specific Knowledge

The results from the item-specific measure support the use of testing during learning and provide evidence which suggests that repeated testing could be an effective tool in the classroom. It was predicted that the STST method would increase retention of classroom material and increase the performance on target items during a unit exam. This was supported by a main effect for encoding condition for the item-specific measure.

Test performance on target items was greater for items presented using the STST study method compared to those using the SSST method. Spitzer (1939) suggested that repeated testing increases recall because the recall on the repeated tests allows participants multiple study sessions with the material. He proposed similar results would be found if an equal amount of time was spent on the rehearsal of the material. The current findings did not support this claim. In the present study when participants used retrieval to answer quiz questions, they answered more corresponding multiple-choice questions on the unit exam correctly compared to when they rehearsed the material by reading a study statement on the quiz.

An interaction occurred between the encoding conditions and the time students received the quiz. When the quiz was given just prior to the multiple-choice unit exam there was no difference in the number correct among the encoding conditions, but when the quiz was given midway through a unit, approximately two weeks prior to the unit exam, retrieval on the quiz was associated with better performance on the target items on the unit exam. The current results are similar in part to findings by Wheeler et al. (2003) who, in testing free recall of word lists, found greater recall in a repeated study condition with a short (5 min) retention interval, but found greater recall in a repeated test condition with longer (48 hours and 7 days) retention intervals. The current results extend the demonstrated effective interval of repeated testing to at least two weeks and shows also the effect can be observed with recognition tests. However, we were unable to give overwhelming support to the benefits of repeated testing with retention intervals as short as 48 hours. This was due to a limitation with the present study concerning the timing of the quiz. It was important that the administration of the quizzes was practical for both the student and the instructor. This necessitated some variability regarding when the student took the quiz, and decreased control compared to the strict control possible in a laboratory. In the present study the participant was given a time period during which he/she could complete the quiz that could vary by as much as four days. It is unlikely that this had much effect on the early quizzes which were given approximately two weeks prior to the unit exam, since forgetting typically occurs rapidly and then becomes stable over time. However, we believe the variability in the retention interval may have had an effect on the late quiz which was given just prior to the unit exam. Wheeler et al. found that in a repeated study condition there was a higher degree of original learning, and a steeper rate of forgetting than in a repeated test condition. In the present study, students could complete the late quiz anytime between 4 days prior to the unit exam up to 5 minutes before they took the unit exam. The variability in when the student decided to take the late quiz, we believe, unintentionally measured students' retention at different critical points in the interaction between practice (retrieval vs. rehearsal) and test time. Students who took the quiz as soon as it was posted online likely benefited more from the retrieval in the repeated testing condition, which Wheeler et al. found slowed the rate of forgetting. Students who waited to take the quiz until right before the exam likely benefited more (at least on the exam) from the rehearsal in the repeated study condition, which Wheeler et al. found increased original learning but was more vulnerable to being forgotten. Unfortunately, our data does not allow us to determine when during the time window each quiz was taken.

Although no explicit encoding instructions were given to the students, past research suggested that the multiple-choice format of the quiz would encourage students to encode quiz material in an item-specific manner (Hunt & Einstein, 1981). Further, Mulligan (2004) found that when a read-only (i.e., study) condition was interspersed with either an item-specific or a relational encoding task, participants used the same encoding process to encode the read-only material as they did to encode the generated or retrieved material. Therefore, while the multiple-choice format likely facilitated item-specific encoding in the STST condition, the interspersing of the study statements in the SSST condition with the multiple-choice questions on the quiz may have facilitated item-specific encoding of the study statements also. Although not measured, past research on repeated testing suggest that the item-specific encoding task used in this experiment should have caused item recovery in the STST and SSST condition to be minimal. This is because students were given unlimited time to complete each quiz. While item-specific tasks should enhance item recovery, hypermnesic gains are thought to be dependent on the extent performance on the initial test exhausted the recovery of potential items. The unlimited time given on the quiz likely exhaust the pool of items available for recovery, making any differences in the number correct on the unit exam between the two encoding conditions the result of a different rate of forgetting between the conditions (Burns & Schoff, 1998; Klein, Loftus, Kihlstrom, & Aseron, 1989; McDaniel et al, 1989). This is in agreement with the findings reported by Wheeler et al. (2003) who used a recall test and measured the rate of forgetting for a repeated test and repeated study condition. The current study found similar results to those found by Wheeler et al. using a recognition test, but we did not measure the rate of forgetting to confirm the belief that the improved performance on the unit exam for the repeated test condition was due to a slower rate of forgetting. Previous research, however, suggest that this may be the case. Wheeler et al. argued that when items are retrieved they become more resistant to being forgotten, and this was true for the retrieved items in the item-specific measure. It is possible that the act of retrieval fixes an item in memory, strengthening the memory traces, and increasing the accessibility of the item in the memory system. It is assumed that the more an item is retrieved the stronger the trace will become and the more resistant the memory will become to being forgotten.

In the current study, students were given only one structured retrieval opportunity before the final retrieval attempt on the unit exam, although the material was available to the students throughout the test cycle. With only one structured retrieval prior to the unit exam, for items from the early quiz there was a 10.4% increase in exam performance on items presented in the hard condition over items presented in the study condition and a 11.3% increase in exam performance on items in the hard condition over the control condition, which received no structured study of the material. Likewise, for items in the easy condition on the early quiz, there was an increase in exam performance of 8.8% and 9.7% respectively. While these effects are somewhat small they might be considered robust given that they were the result of a single structured retrieval opportunity. No previous study was found which demonstrated hypermnesia where a single test trial was used prior to final testing. It is probable that additional structured retrieval opportunities would have created a stronger effect than was found in the current study.

Consistent with the belief that the quantity of separate retrieval events would increase retrieval strength is the related idea that the quality of retrieval would have an effect on future retrieval. It was predicted that the strength of the memory trace would be related to the level of effort needed to retrieve the material, i.e., more retrieval effort would create stronger memory traces. However, in the STST condition, unit exam performance did not differ significantly for items presented in the hard and easy condition on the quizzes. Perhaps our manipulation of easy and hard quiz questions did not influence the difficulty of the retrieval. The same information (the same memory trace) had to be retrieved in either case. Instead the difference in the easy and hard conditions was in difficulty in discriminating the retrieved trace from the alternative answers.

Like the items in the STST condition, no difference was found between the SSST condition and no study control condition. This was somewhat of a surprise, as it was assumed that there would be an exposure effect for the items in the study condition. One possibility is that the study items were only minimally processed. This could have happened if the learner believed that they had already learned the item prior to taking the quiz and thus, chose to ignore the study item when it was presented on the quiz.

Relatedness Knowledge

It was predicted that the STST method would also increase the relational knowledge of classroom material relative to the SSST method. Although we did not find support for this hypothesis, the relatedness rating measure did uncover some interesting insight in to how information is stored in memory.

Using the relatedness ratings measure a main effect was found for encoding condition, but the main effect did not correspond with what was found using the item-specific measure. Overall, the STST method did not increase relational knowledge of the material. Instead the relational knowledge of participants for items in the SSST condition was more similar in structure to the expert's knowledge than for items in the STST condition. It is possible that the quiz items did not improve performance on the relatedness ratings measure because the quizzes encouraged item-specific encoding. Item-specific encoding encourages that attention be placed on the specific attributes of an item, and not the shared qualities the item may have with other items in the learning episode. Attention to the item-specific attributes of an item may limit the amount of attention that can be placed on relational information. Although it was expected that the interspersing of the study statements among the multiple-choice questions on the quiz would have also encouraged item-specific processing of those items, (Mulligan, 2004), it is likely that the strength of the 'tendency to encode in an item-specific manner', although present, was not as strong as with the test statements. This allowed study only items to possibly benefit more from additional relational encoding processes. Therefore, how the information was structured in memory was influenced by whether the item was presented as a study statement which allowed for more relational processing or as a hard or a easy multiple-choice question which may have restricted the use of relational processes and forced attention on the item's item-specific attributes.

The interaction between encoding condition and time of presentation found in the present study suggests two things, first study items received more relational processing as discussed above and that items in the STST condition were influenced by the difference in the difficulty of retrieval between items first presented as hard or easy questions on the quiz. As suggested earlier, hard items may have encouraged more elaborate processing in order to discriminate the correct answer from the alternatives. This additional processing may have increased relational processing compared to that resulting from easy questions. We, however, remain puzzled by the relatively good relational processing in the no study condition.

As expected the time of presentation had a strong effect on the relational knowledge of the students. Students performed better on material on the relatedness rating test which corresponded with the early quiz given two weeks prior to the relatedness rating test than on material on the late quiz. The benefit of time was seen using both the STST and SSST study method and suggested that the structure of the knowledge was influenced by time. It is likely that the longer the information was stored in the memory system, the better organized the information became and the more chance new information could be tied to it, resulting in an increase in the relational knowledge for those items.

Conclusion

The present study took a critical look at the use of repeated testing in the classroom, and we believe that the evidence suggests that repeated testing can be effectively used in an education setting to increase learning. Repeated testing could be particularly helpful when the long-term retention of material is important. We found that with longer retention intervals repeated tests which challenged the learner increased performance on a multiple-choice exam and created knowledge structures that more closely resembled expert knowledge structures. The present study produced a small but significant improvement in performance by requiring just one retrieval of the material. Consistent with previously cited research it would be useful to confirm the prediction that additional intermediate tests could increase performance further.

Repeated testing using a multiple-choice test appears to strengthen the memory trace thereby increasing the likelihood that the item will be more available on a subsequent task. It is believed that this increase in availability is not dependent on the difficulty of the previous retrieval effort. However, we found that when the relational knowledge is assessed that difficult retrieval appears to be more beneficial than easy retrieval. It should be noted, however, that relational knowledge was not dependent on retrieval of the material. The rehearsal of material produced the same results as difficult retrieval with longer retention intervals.

A limitation of this study was that there was no explicit instructions on how to encode the items during the quiz, so we were only able to imply the type of encoding method used based on the type of item on the quiz. Additional studies will be needed to determine whether different types of encoding were occurring or if different degrees of the same type of encoding were occurring in the STST and the SSST methods. In addition, it will be important to further examine the question of why a difference occurred between the relational knowledge of items retrieved using difficult and easy retrieval.

The testing effect in the present study was significant albeit small, for reasons discussed earlier, we believe that the effect can be increased and that repeated testing can be a beneficial tool that instructors can use to increase retention of classroom material. Importantly, repeated testing which challenges the students appears to increase both the item-specific and relational understanding and retention of classroom material.

References

Bjork, R. A., & Bjork, E.L. (1992). A new theory of dissuse and an old theory of stimulus fluctuation. In A. Healy, S.M. Kosslyn, & R. M. Shiffrin (Eds.), *From Learning Processes to Cognitive Processes Essays in Honor of William K. Estes, 2*, 35-67.

Burns, D. J., & Gold, D. E. (1999). An analysis of item gains and losses in retroactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 978-985.

Burns, D.J., & Schoff K. M. (1998). Slow and steady often ties the race: Effects of item-specific and relatedness processing on cumulative recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24.* Retrieved November 25, 2003, from PsycARTICLES database.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633-642.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671-684.

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but nor for words in multiple recall trials. *Cognitive Psychology, 6*, 159-171.

Erdelyi, M. N., & Stein, J. B. (1981). Recognition hypermnesia: The growth of recognition memory over time with repeated testing. *Cognition, 9*, 23-33.

Hunt, R.R., & Einstein, G.O.(1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497-514.

Fritz, C.O., Morris, P. E., Bjork, R. A., Gelman, R., & Wickens, T. D. (2000). When further learning fails: Stability and change following repeated presentation of text. *British Journal of Psychology 91*, 493-511.

Frost, P., Ingraham, M., & Wilson, B. (2002). Why misinformation is more likely to be recognized over time: A source monitoring account. *Memory, 10*, 179-185.

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

Glover J. A., & Krug, D. (1990). The “testing” effect and restricted retrieval rehearsal. *Psychological Record, 40.* Retrieved November, 25, 2003 from Academic Search Premier database.

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

Kay, H. (1955). Learning and retaining verbal material. *British Journal of Psychology, 46,* 81-100.

Kazén, M., & Solís-Macías, V.M. (1999). Recognition hypermnesia with repeated trials: Initial evidence for the alternative retrieval pathways. *British Journal of Psychology, 90,* 405-424.

Klein, S. B., Loftus, J., & Schell, T. (1994). Repeated testing: A technique for assessing the roles of elaborative and organizational processing in the representation of social knowledge. *Journal of Personality & Social Psychology, 66,* 830-839.

Klein, S. B., Loftus, J. Kihlstrom, J. F., & Aseron, R. (1989). Effects of item-specific and relational information on hypermnesic recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15,* 1192-1197.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory,* (pp. 625-635). London: Academic Press.

McDaniel, M. A., Moore, B. A., & Whiteman, H. L. (1998). Dynamic changes in hypermnesia across early and late tests: A relatedness/item-specific account. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 173-185.

Mulligan, N.W. (2001). Generation and hypermnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 436-450.

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 838-855.

Myers, D. G. (2002). Psychology: Myers in modules. (6th Ed.) New York: Worth.

Otani, H., & Hodge M. H. (1991). Does hypermensia occur in recognition and cued recall? *American Journal of Psychology, 104,* 101-116.

Otani, H., & Stinson, M.J. (1994). A further attempt to demonstrate hypermnesia in recognition *Psychological Record, 44*. Retrieved November 23, 2003 from Academic Search Premier database.

Payne, D. G. (1987) Hypermnesia and Reminiscence in Recall: A Historical and Empirical Review. *Psychological Bulletin, 101,* 25-27.

Payne, D. G., & Roediger, H. L. III. (1987). Hypermnesia occurs in recall but not in recognition. *American Journal of Psychology, 100*, 145-165.

Roediger, H. L., Gillespie, G., Lean, D. S., & Payne, D. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21*, 635-655.

Roediger, H. L., Karpicke, J., & Marsh, E. L. (November, 2003) Positive and Negative Effects of Testing in Retention of General Knowledge. Paper presented at the 44th Annual Meeting of the Psychonomic Society in Vancouver, British Columbia.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning*, 4, 11-18.

Schvaneveldt, R. (1990). The Knowledge Network Organizing Tool (KNOT) [Computer Software]. Retrieved from http://interlinkinc.net/index.html.

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.

Tulving, E., & Donaldson, W. (Eds.) (1972). *Organization of memory* New York: Academic Press.

Wheeler M.A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571-580.

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

APPENDIX A

Quiz One: As viewed online.
Participants were required to save their answer for each question before moving on to the next question. When participants saved each answer they received immediate feedback on their performance.

Question 1 (1 point)
The ________________ are bushy branching extension that receives messages from other cells.

- a. Cell Body
- b. Axon
- c. Dendrites
- d. Glial Cells

Save Answer

Question 2 (1 point)
Dr. Petro is testing a new drug for the treatment of depression. In an attempt to lessen bias in the study and the possibility of a placebo effect, Dr. Petro should use THE DOUBLE BLIND PROCEDURE.

- a. I understand this statement.
- b. I do not understand this statement.

Save Answer

Question 3 (1 point)
Hypotheses guide research because they help researchers make _________________.

- a. Money
- b. Predictions
- c. Deliberation
- d. Stagnation

Save Answer

Question 4 (1 point)
Marissa is a psychology student who collected data for a senior thesis. Describing her data to her professor Marissa gave the professor the standard deviation of the scores. Marissa used this statistic to describe the _____________ in the data.

- a. Partition
- b. Oscillation
- c. Fluctuation
- d. Variation

Save Answer

Question 5 (1 point)
Dr. Honeywell added all the scores in his data set and then divided the sum of scores with the number of scores. Dr. Honeywell was interested in knowing the ___________ of the data set.

- a. Genus
- b. Stage
- c. Mean
- d. Illumination

Save Answer

Question 6 (1 point)
A graduate student wanted to look at the effects of sugar consumption on hyperactivity in children. The student manipulated the amount of sugar each group of children received and then measured the activity level of the children. In the study the INDEPENDENT VARIABLE was sugar consumption because it was the factor being manipulated.

- a. I understand this statement.
- b. I do not understand this statement.

Save Answer

Question 7 (1 point)

Dr. Sivori led an experimental study on hunger. In the laboratory, Dr. Sivori compared hungry mice and well-fed mice and measured any difference in motivation to press a bar for food between the two groups of mice. Since Dr. Sivori measured motivation by how many times a mouse pressed the bar, motivation in the form of bar pressing was the DEPENDENT VARIABLE.

- a. I understand this statement.
- b. I do not understand this statement.

Save Answer

Question 8 (1 point)

The complexity of the neural information system is created from simplicity. ___________ are nerves cells that are units that make up the system.

- a. Dendrites
- b. Neurotransmitters
- c. Glial Cells
- d. Neurons

Save Answer

Question 9 (1 point)

The myelin sheath is a layer of fatty tissue that encases the axon of a neuron and increases the ___________________ of the neural activity.

- a. Intensity
- b. Speed
- c. Strength
- d. Concentration

Save Answer

Finish

Feedback

Question 1 (1 point)

The _______________ are bushy branching extension that receives messages from other cells.

Percent Value	Correct Response	Student Response	Answer Choices
0.0%			a. Cell Body
0.0%			b. Axon
100.0%	▶	▶	c. Dendrites
0.0%			d. Glial Cells

Score 1 / 1

Question 2 (1 point)

Dr. Petro is testing a new drug for the treatment of depression. In an attempt to lessen bias in the study and the possibility of a placebo effect, Dr. Petro should use THE DOUBLE BLIND PROCEDURE.

Percent Value	Correct Response	Student Response	Answer Choices
100.0%	▶	▶	a. I understand this statement.
0.0%			b. I do not understand this statement.

Score 1 / 1

Question 3 (1 point)

Hypotheses guide research because they help researchers make ________________.

Percent Value	Correct Response	Student Response	Answer Choices
0.0%			a. Money
100.0%	▶	▶	b. Predictions
0.0%			c. Deliberation
0.0%			d. Stagnation

Score 1 / 1

Question 4 (1 point)

Marissa is a psychology student who collected data for a senior thesis. Describing her data to her professor Marissa gave the professor the standard deviation of the scores. Marissa used this statistic to describe the _____________ in the data.

Percent Value	Correct Response	Student Response	Answer Choices
0.0%			a. Partition
0.0%			b. Oscillation
0.0%		▶	c. Fluctuation
100.0%	▶		d. Variation

Score 0 / 1

Question 5 (1 point)

Dr. Honeywell added all the scores in his data set and then divided the sum of scores with the number of scores. Dr. Honeywell was interested in knowing the ___________ of the data set.

Percent Value	Correct Response	Student Response	Answer Choices
0.0%			a. Genus
0.0%			b. Stage
100.0%	▶	▶	c. Mean
0.0%			d. Illumination

Score 1 / 1

Question 6 (1 point)

A graduate student wanted to look at the effects of sugar consumption on hyperactivity in children. The student manipulated the amount of sugar each group of children received and then measured the activity level of the children. In the study the INDEPENDENT VARIABLE was sugar consumption because it was the factor being manipulated.

Percent Value	Correct Response	Student Response	Answer Choices
100.0%	▶	▶	a. I understand this statement.
0.0%			b. I do not understand this statement.

Question 7 (1 point)

Dr. Sivori led an experimental study on hunger. In the laboratory, Dr. Sivori compared hungry mice and well-fed mice and measured any difference in motivation to press a bar for food between the two groups of mice. Since Dr. Sivori measured motivation by how many times a mouse pressed the bar, motivation in the form of bar pressing was the DEPENDENT VARIABLE.

Percent Value	Correct Response	Student Response	Answer Choices
100.0%	▶	▶	a. I understand this statement.
0.0%			b. I do not understand this statement.

Score 1 / 1

Question 8 (1 point)

The complexity of the neural information system is created from simplicity. __________ are nerves cells that are units that make up the system.

Percent Value	Correct Response	Student Response	Answer Choices
0.0%			a. Dendrites
0.0%			b. Neurotransmitters
0.0%			c. Glial Cells
100.0%	▶	▶	d. Neurons

Score 1 / 1

Question 9 (1 point)

The myelin sheath is a layer of fatty tissue that encases the axon of a neuron and increases the ________________ of the neural activity.

Percent Value	Correct Response	Student Response	Answer Choices
0.0%			a. Intensity
100.0%	▶	▶	b. Speed
0.0%			c. Strength
0.0%			d. Concentration

Score 1 / 1

Total score 8 / 9 = 88%

APPENDIX B

Question comparison between encoding condition for Quiz One and Exam One..

Question Group One

Target item: NEUROSCIENCE

(hard)
The psychological perspective that questions how blood circuitry is linked to abnormal behavior is a (n) ____________________ perspective.

a) Psychodynamic
b) Behavior genetics
c) Neuroscience
d) Developmental

(easy)
The psychological perspective that questions how blood circuitry is linked to abnormal behavior is a (n) ____________________ perspective.

a) Post modern
b) Social-Constructionist
c) Neuroscience
d) Essentialist

(study)
The psychological perspective that questions how blood circuitry is linked to abnormal behavior is a (n) NEUROSCIENCE perspective.

a) I understand this statement.
b) I do not understand this statement.

(exam)
In class lecture, Professor Hampton emphasized the extent to which abnormal body chemistry can contribute to psychological disorders. The professor's lecture highlighted a ___________ perspective on psychological disorders.

a) Psychodynamic
b) Behavior genetics
c) Neuroscience
d) Social-cultural
e) Cognitive

Target item: INDUSTRIAL/ORGANIZATIONAL

(hard)
Dr. Wu is using applied research to help a company determine the best strategy to motivate their employees and increase morale. Dr. Wu's research suggests she is a/an ________________ psychologist.

a) Developmental
b) Industrial/organizational
c) Social
d) Experimental

(easy)
Dr. Wu is using applied research to help a company determine the best strategy to motivate their employees and increase morale. Dr. Wu's research suggests she is a/an ________________ psychologist.

a) Animal
b) Industrial/organizational
c) Popular Culture
d) Paranormal

(study)
Dr. Wu is using applied research to help a company determine the best strategy to motivate their employees and increase morale. Dr. Wu's research suggests she is a/an ________________ psychologist.

a) I understand this statement.
b) I do not understand this statement.

(exam)
Dr. Lewis is involved in an applied research study of customer satisfaction with a newly developed line of facial cosmetics and beauty aids. Dr. Lewis is most likely a(n)________________psychologist

a) Clinical
b) Developmental
c) Biological
d) Personality
e) Industrial/organizational

Target item: CLINICAL

(hard)
_________________ psychology is the branch of psychology that treats individuals with psychological disorders.

a) Developmental
b) Personality
c) Clinical
d) Cognitive

(easy)
_________________ psychology is the branch of psychology that treats individuals with psychological disorders.

a) Engineering
b) Managerial
c) Clinical
d) Literary

(study)
CLINICAL psychology is the branch of psychology that treats individuals with psychological disorders.

a) I understand this statement.
b) I do not understand this statement.

(exam)
Clinical psychologist are most likely to be involved in:
a) Assessing the linkages between biology and behavior
b) The experimental study of motivation and emotion
c) Providing therapy to troubled people
d) The systematic study of how people are influenced by enduring personality traits.

Question Group Two

Target item: DOUBLE-BLIND PROCEDURE

(hard)
Dr. Petro is testing a new drug for the treatment of depression. In an attempt to lessen bias in the study and the possibility of a placebo effect, Dr. Petro should use ____________________ .

a) Random sampling
b) The double blind procedure
c) An experimental condition
d) A dependent variable

(easy)
Dr. Petro is testing a new drug for the treatment of depression. In an attempt to lessen bias in the study and the possibility of a placebo effect, Dr. Petro should use ___________________ .

a) Organization
b) The double blind procedure
c) A diversified portfolio
d) Team building skills

(study)
Dr. Petro is testing a new drug for the treatment of depression. In an attempt to lessen bias in the study and the possibility of a placebo effect, Dr. Petro should use THE DOUBLE BLIND PROCEDURE.

a) I understand this statement.
b) I do not understand this statement.

(exam)
In order to minimize the placebo effect, researchers are likely to make use of:
a) A scatterplot
b) The double-blind procedure
c) Random sampling
d) Standard deviation

Target item: INDEPENDENT VARIALBE

(hard)
A graduate student wanted to look at the effects of sugar consumption on hyperactivity in children. The student manipulated the amount of sugar each group of children received and then measured the activity level of the children. In the study the ________________ was sugar consumption because it was the factor being manipulated.

a) Control group
b) Dependent variable
c) Independent variable
d) Placebo

(easy)
A graduate student wanted to look at the effects of sugar consumption on hyperactivity in children. The student manipulated the amount of sugar each group of children received and then measured the activity level of the children. In the study the ________________ was sugar consumption because it was the factor being manipulated.

a) Voltage
b) Philanthropist
c) Independent variable
d) Gradient

(study)
A graduate student wanted to look at the effects of sugar consumption on hyperactivity in children. The student manipulated the amount of sugar each group of children received and then measured the activity level of the children. In the study the INDEPENDENT VARIABLE was sugar consumption because it was the factor being manipulated.

a) I understand this statement.
b) I do not understand this statement.

(exam)
In order to study some effects of alcohol consumption, Dr. Chu tested the physical coordination skills of 21-year-old men who were first assigned to drink a beverage with either 4, 2,or 0, ounces of alcohol in the laboratory. In this study the independent variable consisted of:

a) The age of the research participants
b) The physical coordination skills of the research participants
c) The amount of alcohol consumed
d) All of the above

Target item: DEPENDENT VARIABLE

(hard)
Dr. Sivori led an experimental study on hunger. In the laboratory, Dr. Sivori compared hungry mice and well-fed mice and measured any difference in motivation to press a bar for food between the two groups of mice. Since Dr. Sivori measured motivation by how many times a mouse pressed the bar, motivation in the form of bar pressing was the _____________.

a) Independent variable
b) Dependent variable
c) Behavior
d) Theory

(easy)
Dr. Sivori led an experimental study on hunger. In the laboratory, Dr. Sivori compared hungry mice and well-fed mice and measured any difference in motivation to press a bar for food between the two groups of mice. Since Dr. Sivori measured motivation by how many times a mouse pressed the bar, motivation in the form of bar pressing was the _____________.

a) Concrete variable
b) Dependent variable
c) :Abstract variable
d) Moderate variable

(study)
Dr. Sivori led an experimental study on hunger. In the laboratory, Dr. Sivori compared hungry mice and well-fed mice and measured any difference in motivation to press a bar for food between the two groups of mice. Since Dr. Sivori measured motivation by how many times a mouse pressed the bar, motivation in the form of bar pressing was the DEPENDENT VARIABLE.

a) I understand this statement.
b) I do not understand this statement.

(exam)
In an experimental study of the effects of sleep deprivation on memory, memory would be the:

a) Control variable
b) Independent variable
c) Experimental condition
d) Dependent variable

Question Group Three

Target item: NEURONS
(hard)
The complexity of the neural information system is created from simplicity. __________ are nerves cells that are units that make up the system.

a) Dendrites
b) Neurotransmitters
c) Glial Cells
d) Neurons

(easy)
The complexity of the neural information system is created from simplicity. __________ are nerves cells that are units that make up the system.

a) Marrow
b) Free Radicals
c) Lysergic Acid
d) Neurons

(study)
The complexity of the neural information system is created from simplicity. NEURONS are nerves cells that are units that make up the system.

a) I understand this statement.
b) I do not understand this statement.

(exam)
The cells that serve as the basic building blocks of the body's information processing system are called:
a) Neurons
b) Neurotransmitters
c) Vesicles
d) Glial cells

Target item: DENDRITES

(hard)
The _______________ are bushy branching extension that receives messages from other cells.

a) Cell Body
b) Axon
c) Dendrites
d) Glial Cells

(easy)
The _______________ are bushy branching extension that receives messages from other cells.

a) Transducers
b) Receivers
c) Dendrites
d) Dominators

(study)
The DEDRITES are bushy branching extension that receives messages from other cells.

a) I understand this statement.
b) I do not understand this statement.

(exam)
Neurotransmitter receptor sites are located on the:

a) Dendrites
b) Mylelin sheath
c) Cell body
d) Axon

Target item: SPEED

(hard)
The myelin sheath is a layer of fatty tissue that encases the axon of a neuron and increases the __________________ of the neural activity.

a) Intensity
b) Speed
c) Strength
d) Concentration

(easy)
The myelin sheath is a layer of fatty tissue that encases the axon of a neuron and increases the __________________ of the neural activity.

a) Stretch
b) Speed
c) Elasticity
d) Placidity

(study)
The myelin sheath is a layer of fatty tissue that encases the axon of a neuron and increases the SPEED of the neural activity.

a) I understand this statement.
b) I do not understand this statement.

(exam)
The myelin sheath helps to increase the ______________ of neural impulses.

a) Frequency
b) Intensity
c) Threshold
d) Speed

<u>Question Group Four</u>
Target item: VARIATION

(hard)
Marissa is a psychology student who collected data for a senior thesis. Describing her data to her professor Marissa gave the professor the standard deviation of the scores. Marissa used this statistic to describe the ______________ in the data.

a) Skewness
b) Central tendency
c) Distribution
d) Variation

(easy)
Marissa is a psychology student who collected data for a senior thesis. Describing her data to her professor Marissa gave the professor the standard deviation of the scores. Marissa used this statistic to describe the ______________ in the data.

a) Partition
b) Oscillation
c) Fluctuation
d) Variation

(study)
Marissa is a psychology student who collected data for a senior thesis. Describing her data to her professor Marissa gave the professor the standard deviation of the scores. Marissa used this statistic to describe the VARIATION in the data.

a) I understand this statement.
b) I do not understand this statement.

(exam)
The standard deviation is a measure of:

a) Central tendency
b) Variation
c) Statistical significance
d) Correlation

Target item: MEAN

(hard)
Dr. Honeywell added all the scores in his data set and then divided the sum of scores with the number of scores. Dr. Honeywell was interested in knowing the ___________ of the data set.

a) Standard deviation
b) Mode
c) Mean
d) Median

(easy)
Dr. Honeywell added all the scores in his data set and then divided the sum of scores with the number of scores. Dr. Honeywell was interested in knowing the ___________ of the data set.

a) Genus
b) Stage
c) Mean
d) Illumination

(study)
Dr. Honeywell added all the scores in his data set and then divided the sum of scores with the number of scores. Dr. Honeywell was interested in knowing the MEAN of the data set.

a) I understand this statement.
b) I do not understand this statement.

(exam)
The most commonly reported measure of central tendency is the:

a) Mode
b) Mean
c) Normal distribution
d) Median
e) Standard deviation

Target item: PREDICTIONS
(hard)
Hypotheses guide research because they help researchers make ________________.

a) Assumptions
b) Predictions
c) Inferences
d) Suppositions

(easy)
Hypotheses guide research because they help researchers make ________________.

a) Money
b) Predictions
c) Deliberation
d) Stagnation

(study)
Hypotheses guide research because they help researchers make PREDICTIONS.

a) I understand this statement.
b) I do not understand this statement.

(exam)
Hypotheses are best described as _________________.

a) Assumptions
b) Replication
c) Explanation
d) Confirmation
e) Prediction