Summer 5-7-2010

Interteaching and the testing effect: How quizzes alter the efficacy of interteaching

Tonya Lambert
James Madison University

Interteaching and the Testing Effect: How Quizzes Alter the Efficacy of Interteaching

Tonya Lambert

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Psychological Sciences

August, 2010

Acknowledgements

First, I would like to acknowledge my wonderful thesis committee, especially my advisor, Dr. Bryan Saville. Without his guidance, love of research, and subtle sarcasm, this project would not have been possible. As for the rest of my committee, Drs. Tracy Zinn and Jessica Irons, I would like to thank them for not only being so supportive and challenging at the same time, but also for being flexible during times of scheduling and changes made to my thesis timeline.

I would also like to thank my family and friends for supporting me and never letting me give up. Without them, I would not be where I am today and I am eternally grateful for such amazing people in my life.

Table of Contents

List of Figures

Abstract

In recent years, educational systems have come under great scrutiny. In response to this scrutiny, researchers have developed numerous behavior-analytic teaching methods, the most recent being interteaching. In a growing number of studies, interteaching has proven to be more effective than traditional methods of instruction, but little research has examined ways to make interteaching even more effective. Research on the testing effect suggests that frequent testing may improve student-learning outcomes. Thus, including post-discussion quizzes as a part of interteaching might be one way to make it more effective. In the present study, participants completed interteaching in a simulated classroom setting. Some participants completed a post-discussion quiz, whereas others completed anagrams. All participants then returned 1 week later to take a brief quiz. Participants who completed anagrams had higher quiz scores than those who completed post-discussion quizzes. Although further research is needed, these results suggest that the testing effect may not enhance interteaching and, in fact, may diminish its efficacy.

Chapter 1: Introduction

In recent years, educational systems, especially those in higher education, have come under great scrutiny. For instance, Hersh and Merrow (2005) introduced their edited book with a call for people to think about and discuss the "mediocrity" and "underperformance" of college institutions in training students, not only for careers, but also for general knowledge about life. Throughout the book, contributors suggested reasons why colleges are failing the students, ranging from privileges awarded to athletes to the "big business mentality" of administrators. Another possible reason for the purported downfall of higher education is the continued use of ineffective teaching methods.

Criticisms of traditional educational practices are by no means new. In 1954, B. F. Skinner, for instance, identified several problems with traditional teaching methods. First, students in these environments are constantly trying to avoid aversive consequences (e.g., bad grades). Along with avoiding aversive consequences, students rarely receive positive reinforcement. Finally, the typical situation lacks an efficient shaping program. Without a shaping program, educators cannot effectively develop the behaviors that are necessary for current and future skills.

Lecture is the most commonly used teaching method (Benjamin, 2002). Although Skinner criticized traditional teaching methods, they can be effective pedagogical tools. As McKeachie and Svinicki (2006) pointed out, lectures are good for presenting up-to-date information (such as when the textbook in use is older), integrating material from a variety of sources, linking course material to the interests of students, providing orientation when introducing a difficult concept, and helping students focus on key ideas.

Most importantly, lectures should be a bridge between students' interests and things they already know about the subject matter. To make lectures more effective, McKeachie and Svinicki suggested breaking up the lecture to maintain students' attention, training students to be better listeners through writing activities, encouraging deep processing by pointing out relations between concepts, and training students to write better notes by collecting them. Saville (2008) also suggested that instructors should be enthusiastic about the material and that instructors get the students involved. In spite of these recommendations, numerous studies have shown that traditional lectures tend to be less effective at improving student learning than other alternative teaching methods (e.g., Alba & Pennypacker, 1972; Benedict & Anderton, 2004; Binder, 1988, Dochy, Segers, Van de Bossche, & Gijbels, 2002; Kulik, Kulik, & Cohen, 1979; Saville, Zinn, & Elliott, 2005).

In the past few decades, educators have attempted to address the concerns of Skinner and others by developing alternative teaching methods that seem to affect learning more positively than traditional lectures. Three recent alternatives are Just-in-Time Teaching (JiTT), Problem-Based Learning (PBL), and Peer Instruction (PI). JiTT integrates classroom activities that promote active learning and internet resources that enhance the classroom component (Novak, Patterson, Gavrin, & Christian, 1999). Instructors identify the course objectives for discussion on a given day and write questions to address those objectives. There is typically one question for each objective, and the questions consist of at least one open-ended question and several multiple-choice questions. The instructor then posts these questions on a course Web page, and students have several days to study the material and submit their answers electronically. Once the

instructor receives the submissions, he identifies the questions that need further discussion and prepares a lecture "just in time" for class. During class, the instructor reviews each question and objective, elaborates on the course material, and stimulates discussion by providing examples of good answers and poor answers to the questions (Novak et al., 1999). In sum, JiTT creates a student-centered course, increases the likelihood that students will study the material before class, provides both students and instructors with immediate feedback, promotes increased interaction between students and between students and the instructor, and creates a dynamic and interactive classroom (Barron, Benedict, Saville, Serdikoff, & Zinn, 2007).

PBL requires students to work together to solve complex, "real world" problems without any real correct answer. Although there is some variability in how PBL is implemented (Rhem, 1998), instructors typically create problems that are difficult and require students to research, learn, and apply course concepts while developing their solution. The distribution of problems precedes learning of course material, so the instructors must write good problems. The instructor distributes one problem to each group of 5 to 10 students. The group then discusses the problem and what they do and do not know about the problem. They decide what they need to learn more about, divide the responsibilities, and begin to gather information. The groups must continue to research and share information until they feel they have learned enough about each piece of the problem and they can come to a consensus on the solution. Because the problems do not have a definite solution, the group must be able to defend how they gathered their information and the solution. Throughout this time, the instructor serves as a facilitator of discussion and meets frequently with each group to discuss their progress (Antepohl &

Herzig, 1999; see also Connor-Greene, 2005).

In a PI classroom, students must apply the concepts discussed each day as well as explain and discuss those concepts with fellow classmates. The instructor divides the lecture into sections with each section focused on a concept. At the end of each section, there is a question (ConcepTest) based on the concept just discussed. Students have a couple of minutes to answer the question individually and in some fashion report their answer to the instructor. The students then get into groups, discuss their answer with others, and are urged to convince each other of the correctness of their own answer. Following the discussion period, the instructor has the students again report their answers to the question since these may have changed during the discussion. The instructor then explains the correct answer and moves on to the next section of the lecture. Students receive minimal credit for participating in the ConcepTests throughout the semester and many of the questions appear on exams. In order to save time in class, students are required to read prior to each class.

Review of Behavioral Teaching Methods

Behavioral researchers have also developed educational methods in response to the issues discussed by Skinner (1954). Among these methods, there are certain common features. First, each program uses reinforcement frequently and sets a goal for each student to master the material. Next, teachers assess some measure of student learning regularly and directly. Finally, the tasks and concepts are broken down into small components, and students are taught necessary pre-requisite skills (Kinder & Carnine, 1991). The four most commonly used behavioral instruction methods are precision teaching (Lindsley, 1964), programmed instruction (Skinner, 1968), direct instruction

(Engelmann & Carnine, 1982), and the personalized system of instruction (PSI; Keller, 1968).

Precision Teaching

Ogden Lindsley (1964) developed precision teaching based on the idea that students, like any other organism a researcher may study, should have their behavior analyzed frequently. To facilitate data collection on learning, Lindsley created the standard celeration chart (Binder & Watkins, 1990). Celeration is the increase or decrease in rate of responding over time. A standard celeration chart allows for the standardized charting of celeration and analyzing how frequency of behavior changes over time. The standard celeration chart is a semilogarithmic chart that can accommodate response rates as low as 1 every 24 hrs or as high as 1,000 every min. The four types of scaling are daily, weekly, monthly, and yearly. The chart is standardized because of its consistent display of celeration. A 34° trend line drawn from the bottom left corner to the top right corner through graphed data points represents the direction and degree of the actual trend in the data (Cooper, Heron, & Heward, 2007). The key components of precision teaching are that students are free to respond at their own pace, frequency of responses are monitored daily, students self-record, and the standard celeration charts are used to display changes in learning (Lindsley, 1990). Students are required to practice the responses frequently and report the appropriate response associated with each concept. In most cases, the number of students greatly outweighs the number of teachers. Therefore, it is typical for students to collect data on their own behavior (Maloney, 1982). In fact, many teachers have found that changes in behavior are greater when students take an active role in monitoring (Lindsley, 1990). Teachers then use the data to make

decisions about whether students have mastered the material and are ready to move on to new material.

A major benefit of precision teaching is that it combines well with any curricular approach that allows for counting and charting. The Morningside Academy (established 1980 in Seattle, Washington), for instance, combines precision teaching with direct instruction (see below) and offers a money-back guarantee if a student does not gain at least two grade levels in one year in their skill of greatest deficit (Lindsley, 1992; http://www.morningsideacademy.org/about/indepth.php). Elementary school students in Great Falls, Montana engaged in 20 to 30 min per day of timed practice, charting, and decision-making over a period of four days. Students improved between 19 and 44 percentile points on subtests of the Iowa Test of Basic Skills compared with children in control group classrooms in the same school district (Binder, 1988). Precision teaching is more effective at aiding students in the learning process, but there are very few who use this method (Binder & Watkins, 1990).

Programmed Instruction

Shortly following the introduction of precision teaching, B.F. Skinner (1968) developed programmed instruction. Programmed instruction entails a system of small steps, originally displayed through a "teaching machine," in order to increase reinforcement and decrease students' contact with aversive consequences. Because the material is taught through the machine, students move at their own pace, and the teacher is available to help individuals as necessary. Concepts from the material are divided into frames, typically in sets of three, and questions regarding the material are presented to each individual student. The teaching machines provide immediate reinforcement for the

correct response, or they provide feedback when students respond incorrectly. Specifically, when a student responds incorrectly, the program may take one of three courses: The program may break out of sequence to remediate the error and then return to the original sequence, the set may start over, or the program may go back several steps and repeat the instruction (Skinner, 1968). Although early studies showed positive outcomes, programmed instruction never caught on due to the cost of the machines and inconclusive research about its effectiveness (Kulik, Schwalb, & Kulik, 1982; McDonald, Yanchar, & Osguthorpe, 2005).

Direct Instruction

Siegfried Engelmann first developed direct instruction based on the idea that disadvantaged children learn more if taught more (Engelmann & Carnine, 1982). Teachers must be trained on the proper use of direct instruction, and students are required to complete a placement test prior to beginning the program and then again frequently throughout (Adams & Carnine, 2003). Instructors teach each component and then provide practice and feedback before testing the students to determine mastery of the material (Engelmann & Osborn, 1999). When prompted to give an answer, all students in the group must respond correctly and in chorus. If even one student does not do so, the teacher will provide feedback (e.g., the correct answer) and they repeat the answer until all respond in unison. The goal is that students will quickly encounter the natural reinforcers in the classroom setting (Engelmann & Carnine, 1982).

Direct instruction was part of a mass research project during the 1960s known as Project Follow Through. The project used 22 educational programs across 51 school districts. The programs were divided into three categories: Basic skills (DI was in this

category), cognitive-conceptual, and affective-cognitive. Direct instruction had a "considerable impact" on reading comprehension and problem-solving skills and was the most effective program in teaching basic literacy and mathematical competence (Watkins, 1988). All direct instruction students performed close to or at the norm for all measures (Engelmann, Becker, Carnine, & Gersten, 1988). Results from follow-up studies of PFT have shown that high school students who received direct instruction scored higher on standardized tests, had lower drop out rates, and a higher proportion of students applying to college compared to control groups in similar communities (Gersten & Keating, 1987).

As with most non-traditional teaching techniques, there are disadvantages to direct instruction. Because of the requirement for all students to answer in unison, instructors may spend excessive amounts of time correcting an individual within the group, thus failing to hold the attention of the other students. To overcome this obstacle, teachers must administer placement tests frequently so students can progress to more advanced levels as soon as possible (Catania & Brigham, 1978).

Personalized System of Instruction

Arguably the most well-known behavioral teaching method is Keller's (1968) PSI, which has five defining features: self-pacing, unit-perfection requirement, use of lectures to provide motivation and/or reinforcement, stress on the written word, and use of proctors (Johnson & Ruskin, 1977; Keller, 1968). On the first day of class, the instructor gives students a description of the class, instructions, and words of advice about how best to complete the assignments and the course. Essentially, a student may complete all requirements in less than one semester, or it could take several semesters

depending on the restrictions given by the instructor. Following the introduction to the class, students receive their first assignment along with study questions. The classroom serves as a study hall, and students are encouraged to use it at least part of the time. The instructor also offers motivational lectures to students who have "earned" them, meaning that they are on track with the course. These lectures are typically either demonstrations or discussions and are not part of the exams. Once students are ready to take the first test, they come to class on the assigned days, and a student proctor administers the test. The instructor sets a minimum criterion that students must meet before they have a chance to defend their answers. If the student cannot defend the answers given, the proctor sends him or her to study for a minimum of 30 min with advice about the material to study. In the event that the student gets all of the questions correct, the proctor may ask him to explain one or two of the answers to be sure that it was, in fact, the student who completed the test and to get the student further acquainted with the material. Once the student successfully completes the test, the proctor records the success, and the student may move on to the next assignment. In a number of studies, researchers found PSI superior to conventional lecture, and most students actually prefer the PSI method (e.g., Alba & Pennypacker, 1972, Corey & McMichael, 1970, Morris & Kimbrell, 1972).

Although behavioral teaching methods tend to produce superior learning outcomes relative to more traditional lecture-based methods, there are some perceived problems with these methods (see Buskist, Cush, & DeGrandpre, 1991; Saville et al., 2005). First, many administrators do not like the perceived change in the role of the teacher. In each of these methods, the teacher is no longer the "star" of the classroom but serves as a facilitator or tutor for students. Second, because of their focus on mastery,

behavioral teaching methods typically produce a negatively skewed grade distribution (i.e., mostly As), an outcome that some administrators see as grade inflation. Third, some behavioral teaching methods—for example, PSI with its focus on self-pacing—do not fit nicely into the traditional semester system. Finally, many educators do not fully understand behavioral principles and, thus, misapply them (e.g., Ainsworth, 1979).

Interteaching

In 2002, Boyce and Hineline introduced interteaching as a user-friendly teaching alternative that attempts to rectify some of the problems with other behavioral teaching methods. Interteaching is, like the methods mentioned above, based on behavior-analytic principles, with the belief that learning is something individuals do, not something that happens to them. This method is similar to PSI in that there is less emphasis on lectures and more emphasis on active learning. In addition, interteaching allows for self-pacing but within each class period instead of across the semester. To account for this slight difference, interteaching allows students to decide what material the instructor covers during class, making the lectures, theoretically, reinforcing (Boyce & Hineline, 2002).

Interteaching works as follows (see Saville, Lambert & Robertson, in press, for a complete description). The instructor prepares a number of questions (a "prep guide") for the students to answer as they complete their reading before class. Prep guides are meant to guide students through the course material and thus provide focus for students as they read the material. During class, students break off into dyads or triads to discuss the questions from the prep guide. While students are discussing the material, the instructor, and any teaching assistant(s) who may be available, attend to each discussion group and clarify any questions that students may have about the prep guide. At the end of the

discussion, students complete a record sheet on which they list any questions that gave them problems or for which they would like further clarification. Students also report how well their discussions went and which topics they found interesting. The instructor then uses the record sheets to prepare a brief lecture for the next day's class, addressing the questions that students had difficulty answering or understanding.

Other components of interteaching include frequent exams (i.e., at least five per semester), which include questions based on the prep-guide items; participation points for completing the discussions; and quality points, the purpose of which is to improve the pair discussions. As an example of quality points, if two students pair off during discussions and each later earns the equivalent of an "A" or "B" on an essay question, they both receive a small number of quality points towards their class grade. If either of them earns less than the equivalent of a "B", however, neither receives quality points. Quality points do not serve as extra credit, however, but count as 10% of the total course grade.

Research on Interteaching

Few empirical articles have examined the effectiveness of interteaching. Of those that have been published (Saville et al., 2005, Saville, Zinn, Neef, Van Norman, & Ferreri, 2006), interteaching has proven to be more effective than typical lecture at improving student-learning outcomes. Saville et al. (2005) conducted a study in which students came to a laboratory and were randomly assigned to one of four conditions: interteaching, lecture, reading, or control. Participants in the interteaching condition spent 15 min reading a short journal article and completing a prep guide. They then formed pairs and discussed the questions for another 15 min. After the discussion period,

participants informed the instructor of questions they would like reviewed, and the instructor spent the remaining 15 min discussing those questions. Participants in the lecture condition simply heard a 45-min lecture on the same article. In the reading condition, the researchers instructed participants to study the article for 45 min. Participants in each of these three conditions (interteaching, lecture, and reading) returned 1 week later to take a 10-question, multiple-choice quiz based on the article. Participants in the control condition only reported to the laboratory once to take the quiz. Saville et al. found that participants in the interteaching group did significantly better on the quiz than participants in the lecture group, the reading group, or the control group. Interestingly, there were no significant differences among any of the latter groups.

In another study, Saville et al. (2006) conducted two studies in which they compared interteaching to lecture in college courses. In the first study, graduate students in a master's-level class served as participants. At the start of the semester, the instructor administered pre-tests on major course concepts. The instructor then alternated between interteaching and lecture classes throughout the semester so that neither interteaching nor lecture occurred for more than two class days in a row and each occurred an equal number of times. They found that the mean difference between pre- and post-test scores was consistently higher in the interteaching condition than in the lecture condition. In Saville et al.'s second study, undergraduate students in two sections of a research methods course served as participants. A similar alternating design was used, with the teaching method (interteaching or lecture) counterbalanced across sections. After each unit, the two sections took the same unit test. The authors then compared the mean score for students in the lecture condition to the mean score for students in the interteaching

condition. Mean unit test scores were consistently higher when students were in the interteaching condition. For each of the two studies, Saville et al. also collected social validity data. The majority of students from the first study reported that they preferred interteaching to lecture and that they learned equally either well with interteaching and lecture or learned more with interteaching. Most students in the second study also reported a preference for interteaching and that they learned more with interteaching.

Goto and Schneider (2009) combined interteaching with lectures in two upper-level nutrition courses with very minor modifications to better suit the discipline. At the end of the course, students completed an evaluation of interteaching. Overall, students reported that preparation guides helped them prepare for class and that the in-class discussions encouraged critical thinking. Students also reported that they enjoyed the interteaching method and found it beneficial to their learning.

In a move towards component analysis, Cannella-Malone, Axe, and Parker (2009) studied the effects of different prep-guide types on quiz scores. Seven seniors in an undergraduate special education course participated in the study and provided social validity data at the end of the semester. As part of an alternating treatments design, students either answered 10 questions on a prep guide or wrote at least four questions on topics they thought were important or that they did not understand. During each class period, students spent the first 25 min discussing the prep guides or written questions. Then they completed an interteaching record. Following a 30-min lecture, students took a 20-min quiz. Quizzes consisted of multiple-choice, fill-in-the-blank, factual short answer, and problem-solving short answer questions. When Canella-Malone and her colleagues compared total quiz scores in each of the conditions, the average differences

were very small. The average score when students answered prep guide questions was 81%, and the average score when students wrote questions was 82%. When the researchers compared scores on the different question types across the two conditions, however, there were larger differences. Average scores on multiple-choice questions were 86% and 79% for writing questions and answering questions, respectively. Fill-in-the-blank average scores were 81% when students wrote questions and 88% when students answered questions. Factual short answer average scores were 80% when students wrote questions and 88% when students answered questions. Ranges of scores for the different types of questions were broad except for problem-solving short-answer questions. These were the most consistently different scores, although the difference was not that large: Students who wrote questions had an average score of 81% and those who answered questions had an average score of 75%. Overall, different preparation types (answering instructor-written prep-guide questions or writing one's own questions) did not significantly influence quiz performance.

Most recently, Scoboria and Pascual-Leone (2009) examined the associations between academic performance, student engagement, and participation in interteaching in two large abnormal psychology courses. The instructors followed the interteaching model except (a) they divided the class evenly between discussion and lecture instead of spending two-thirds of class on discussion, and (b) they did not emphasize an explicit connection between the prep guides and exams. Measures included the number of discussions attended, grades on exams, and writing assignments recorded, in addition to a composite course average and preference and motivation assessments. Scoboria and Pascual-Leone found that, compared to performance in another lecture-based class from a

previous semester, the performance on the writing assignment was significantly higher. They also found that attendance correlated significantly with course average, writing assignment grades, and exam scores. After controlling for academic ability in the first course, the authors also found significant relations between performance and attendance for the composite average, writing assignments, and the second exam. The authors completed a similar regression with the second course, but controlled for motivation and found a significant relation for the composite score and the writing assignments. Students also preferred interteaching in both courses.

Recently, researchers have conducted component analyses of interteaching. Although there is still a need for more empirical research on the effectiveness of interteaching (Saville et al., in press), the component analysis serves to determine what components make interteaching effective. Saville and Zinn (2009) recently completed the first of these component analyses on the quality-points component of interteaching. For this study, participants were students in two sections of an introductory psychology course, and the method followed that of Saville et al. (2006, Study 2). More specifically, Saville and Zinn used an alternating treatments design, switching between interteaching with quality points and interteaching without quality points several times, counterbalancing across sections. Following each unit, both sections of the class completed the same exam consisting of two five-point essay questions and several fill-in-the-blank and short answer questions. There was not a significant difference between sections on five of the six exams. Moreover, on the exam where there was a difference, the difference was in the opposite direction of that expected: Students in the section with no quality points performed better on the exam than those in the section with quality

points. Although there are several possible explanations for these results, the most likely one is that the delivery of quality points was too delayed, thus decreasing their effectiveness.

The Testing Effect

Although interteaching seems to be a promising new teaching method, there are likely ways to make it even more effective (see Saville et al., in press). Boyce and Hineline (2002) mentioned that daily quizzing might be aversive to students, but there has been some empirical evidence that frequent testing may help students retain information (see Roediger & Karpicke, 2006a, for full review). Therefore, the positive effects of frequent testing may outweigh the potentially aversive affects reported by students and make interteaching a more effective teaching method.

Since the early 1900s, psychologists have studied a phenomenon known as the testing effect. In the first study on the effects of testing, Gates (1917) tested groups of children from grades 1, 3, 4, 5, 6, and 8. Participants read nonsense syllables or biographical facts and were then instructed to recall the material. Participants were allowed to glance at the materials during recall, however. Gates varied the amount of time spent reciting so that participants spent 0, 20, 40, 60, 80, or 90% of their time recalling material. In almost all of the conditions, Gates observed a positive effect of recall. The only group that showed no change in performance was the first-grade participants who read nonsense syllables.

Several years later, Spitzer (1939) had 91 sixth-grade students study articles similar to what they studied in school. The students then took a 25-question, multiple-choice test at various times over the following 63 days. When students were given a test

and then retested later, performance on the test did not decrease and even increased in some cases. Moreover, the sooner a test was administered following the study time, the better students did on later tests.

Although both of these studies were methodologically flawed, Forlano (1936) replicated Gates' (1917) study, and Sones and Stroud (1940) replicated Spitzer's study with similar results. For multiple reasons, research on the testing effect was abandoned until the late 1960s when Tulving (1967) began a new line of research. In his study, participants read a list of 36 words presented in a random order and then recalled items aloud. There were three conditions: One group alternated between studying and testing, another group studied three times and then recalled lists once, and the third group studied the list once and then recalled the lists in three consecutive trials. Interestingly, the learning curves were almost identical for each condition, showing no positive effect of testing.

Roediger and Karpicke (2006b) recently replicated Tulving's study using prose passages instead of word lists. In the first study, 120 undergraduate students read two prose passages that the researchers divided into 30 "idea units." Participants read the passages within the first phase of the experiment, which consisted of four 7-min periods. During any given period, the researcher could prompt the participant to study a passage for the first time, restudy the passage, or take a recall test. The order of the passage read was counterbalanced, with students restudying or taking a recall test. The recall test consisted of writing as much information as students could remember on a sheet of paper, simulating an essay test. Between each period, the students solved multiplication problems for 2 min and did the same for 5 min following the final period. Phase 2 then

occurred 5 min, 2 days, or 1 week later. In Phase 2, students were asked to recall passages learned in Phase 1 and were given a point for each recalled "idea unit." Roediger and Karpicke observed a significant interaction between the learning condition and the interval. After 5 min, those who restudied the passage did better than those who completed a test on the passage. The opposite was true of the students who were in the 2-day or 1-week condition. Specifically, those who took a test on the passage did better than those who re-studied the material following a 2-day or 1-week interval.

Roediger and Karpicke found similar results in their second experiment. Students read the same two passages from the first experiment, but the conditions were slightly different. In Phase 1, students had either four periods of reading the passages, three periods of studying and one period of testing, or one period of studying and three periods of testing. The intervals remained the same (5 min, 2 days, or 1 week). Roediger and Karpicke observed the same interaction: Students in the 5-min interval group performed better on the test if they had studied the passages four times than if they studied the passage three times and took a test or if they studied once and took the test three times. Those in the 2-day or 1-week group performed better if they studied the passage once and took the test four times.

Other researchers who have studied the testing effect have found that it can also enhance the recall of nontested but related material. Chan, McDermott, and Roediger (2006) had participants read an article for 25 min and then instructed them to complete a brief test (22 questions) twice through, read 22 statements twice, or dismissed them after reading the article (control condition). There were two versions of the test, each with 18 related questions and 5 "filler" questions, and the order in which students took the tests

was counterbalanced. Participants then returned the following day to complete a 40-item test. Participants in the testing condition performed better on both the tested and nontested but related items. In addition, participants in the testing condition performed better overall than participants in the extra study or control groups.

For a second part of the study, which served as a replication, participants read two of four articles. After reading an article, they either took a 12-item quiz or waited to begin reading the second article. As with Study 1, the students returned the following day to complete a 48-item test with items tested the previous day, items not tested but related to the previous test, and 24 questions from the control article. The authors observed that participants in the testing condition performed better on the tested and nontested but related items than those in the read-only condition.

Butler and Roediger (2007) recently studied the testing effect in a simulated classroom. Twenty-seven undergraduates served as participants, and the researchers assigned them to a multiple-choice, short-answer, or study control condition. Students came to the "classroom" and watched a video lecture and then either studied the material, took a multiple-choice test, or took a short-answer test. They then returned to the classroom for the next 2 days and followed the same routine, watching two other video lectures and completing two other postlecture activities. Videos and postlecture activities were counterbalanced. Because Butler and Roediger were also interested in the effects of feedback on the testing effect, some students received feedback in the form of the correct answer.

One month following the final "learning" session, the participants returned to take a comprehensive short-answer test. Participants who took a test following the lecture did

better on the final test than those who studied the material. Moreover, those who completed the short-answer test performed better than those who completed the multiple-choice test. The researchers also found that feedback had no significant effect on how the participants performed on the final test.

Although the majority of research on the testing effect has been conducted in laboratory or simulated classroom settings, McDaniel, Anderson, Derbish, and Morrisette (2007) examined the testing effect in a web-based course. Participants were assigned to read pages in a textbook and then completed a 10-item quiz in either multiple-choice, short answer, or read-only format. Immediately after completing the quiz, students were given access to feedback. After 3 weeks and three quizzes (one of each type), students completed a unit test that consisted of questions from the quizzes as well as related questions. McDaniel et al. found that participants were more likely, on the weekly quizzes, to answer multiple-choice questions correctly than short-answer questions. They also found that participants performed better on previously quizzed items on the unit test than on non-quizzed items, regardless of whether they took the multiple-choice or short-answer quiz. They did find, however, that students who completed the short-answer quiz more accurately learned and retained facts (quizzed or not quizzed) than those who completed the multiple-choice or read-only quiz. Similar results were found on a final comprehensive exam.

Integrating Interteaching and the Testing Effect

One method of capitalizing on the testing effect with interteaching would be to administer quizzes post-discussion, before the instructor has given the clarifying lecture. Research by Roediger, McDaniel, and colleagues (e.g., Chan et al., 2006, Roediger & Karpicke, 2006b) suggests that including a post-discussion quiz following the

interteaching pair discussions should improve quiz performance. Therefore, the purpose of the present study is to evaluate the effectiveness of adding post-discussion quizzes to the interteaching method.

Chapter 2: Method

Participants

One hundred seventeen undergraduate students (84 women, 33 men) from James Madison University participated. Their average age was 18.81 years, and there were 82 freshman, 21 sophomores, 9 juniors, and 4 seniors; 1 participant did not report her year in school. The students were enrolled in various undergraduate psychology courses and received partial course credit for participating; students also had the opportunity to win a $25 gift card.

Materials and Procedure

The materials and general procedure in the present study are based largely on those reported by Saville et al. (2005) in their lab-based interteaching study. Groups of 11 or 12 students reported to a classroom. Participants were assigned to one of two conditions: interteaching with a post-discussion quiz (quiz condition, $n = 58$) or interteaching with no post-discussion quiz (no-quiz condition, $n = 59$). Specifically, I quasi-randomly chose the conditions before each session so that there were five or six participants in each condition within each group. As soon as participants arrived in the room, but before the actual session began, they read an informed consent with information regarding what they would be doing as part of the study. After providing consent, participants spent the next 55 min completing an interteaching session.

During the first 15 min of the session, participants read a brief article by Allen (2003), which discussed whether pets have a positive effect on blood pressure, and completed a 10-item preparation guide (see Appendix A) that contained questions about the article. I then instructed the participants to form pairs and discuss the prep-guide

items for another 15 min. The only formal instruction given to participants on how to choose their partner was that they could not pair up with someone they knew prior to the session. Once the participants finished their discussions, they took approximately 5 min to complete a record sheet, on which they reported the quality of their discussion (on a 7-point scale where 1 = *poor* and 7 = *superb*), questions on the prep guide with which they had difficulty, questions they wanted reviewed, and demographic information. I then collected all materials (informed consent, article, prep guide, and record sheet).

Next, participants in the quiz condition had 5 min to complete an 8-item, short-answer quiz (Quiz 1; see Appendix B); participants in the no-quiz condition completed anagrams for 5 min. I handed out quizzes and anagrams randomly so it was possible that participants who paired off for the discussion may have been in different groups. During this time, I reviewed the record sheets and decided which questions to cover during the lecture. After the quizzes and anagram sheets were collected, participants heard a 15-min lecture over prep-guide items they asked to be reviewed. After the lecture, participants were dismissed. Participants returned 1 week later to complete a 16-item, multiple-choice quiz (Quiz 2; see Appendix C). Eight of the questions on Quiz 2 were similar to the short-answer questions from Quiz 1 but presented in multiple-choice format; the eight remaining questions were related to the original questions but previously untested. All questions were based on material presented in the article.

Interobserver Agreement (IOR)

An undergraduate assistant and I independently scored 31% of the short-answer quizzes (Quiz 1). We scored each quiz on a separate sheet to ensure independent grading. I calculated agreement scores by dividing the total number of agreements by the

total number of items (8 on each test, 136 total items) and multiplying by 100%. The mean agreement score was 93%, with a range of 75% to 100%. Because the correct answers on Quiz 2 were unambiguous (i.e., because the items were in multiple-choice format), I did not collect any IOR data.

Chapter 3: Description of Data Analysis

Of the original 117 participants who completed the first phase of the study, 8 did not return the following week to take Quiz 2. As such, only the data from the remaining 109 participants (54 in the quiz condition and 55 in the no-quiz condition) are included in the following analyses. To examine whether there was a significant difference between groups on Quiz 2, I conducted a one-way ANOVA for independent groups, which produced a statistically significant outcome, $F(1, 107) = 5.42$, $p = .01$, $d = .45$. As shown in Figure 1, participants in the no-quiz condition answered a greater percentage of questions correctly ($M = 66.70$, $SD = 11.67$) than participants in the quiz condition ($M = 61.23$, $SD = 12.87$).

To investigate the differences between the two groups further, I divided the questions that participants answered into two categories: the eight items that appeared on Quiz 1 and Quiz 2 (the tested items), and the eight items that appeared only on Quiz 2 (the non-tested items). I then conducted a 2 (condition) x 2 (question type) mixed factorial ANOVA. There was no interaction of condition and question type, $F(1, 107) = 0.92$, $p = .34$, but there was a main effect of condition, $F(1, 107) = 5.42$, $p = .01$, $d = .45$, and a main effect of question type, $F(1, 107) = 54.39$, $p < .001$, $d = 0.94$. Specifically, participants in the no-quiz condition answered more questions correctly, on average, than participants in the quiz condition; and participants in both conditions answered correctly more tested items than non-tested items (see Figure 2).

To provide further information on how participants in both conditions answered the different types of questions, I conducted two independent-samples *t* tests, one on the tested items and one on the non-tested items. There was no significant difference

between conditions on the non-tested items, $t(107) = 0.93$, $p = .18$ (quiz condition = 54% correct; no-quiz condition = 58% correct). There was, however, a significant difference on the tested items $t(107) = 2.75$, $p = .003$, $d = 0.53$ (quiz condition = 68% correct; no-quiz condition = 75% correct). Because participants in the no-quiz condition did not actually see the tested items prior to Quiz 2 (they only completed anagrams), these findings suggest that the tested items (i.e., those that appeared on both Quiz 1 and Quiz 2) were unintentionally easier to answer than the non-tested items.

One possible reason for why participants in the quiz condition performed worse, on average, than participants in the no-quiz condition on Quiz 2 is that their answers on Quiz 1 may have affected the answers they gave on Quiz 2. Specifically, if participants in the quiz condition incorrectly answered questions on Quiz 1, they may have incorrectly remembered the same information when it came time to answer questions on Quiz 2, a phenomenon known in cognitive psychology as the "negative suggestion effect" (e.g., Brown, Schilling, & Hockensmith, 1999; Remmers & Remmers, 1926; Toppino & Luipersbeck, 1993). To examine this possibility, I first analyzed whether the performance of participants in the quiz condition differed on Quiz 1 and Quiz 2. On Quiz 1 and Quiz 2, participants in the quiz condition had an average score of 63.89% and 68.05%, respectively, an increase that approached, but did not quite reach, traditional levels of significance, $t(53) = 1.47$, $p = .07$, $d = 0.23$. It is important to remember, though, that the formats of Quiz 1 (short-answer) and Quiz 2 (multiple-choice) were different. Thus, any differences may have been a function of quiz format and not necessarily changes in how much they remembered.

Nevertheless, although the questions on Quizzes 1 and 2 differed in format, I still wanted to examine more closely whether participants may have been remembering incorrect information on Quiz 2 as a function of their answers on Quiz 1. In order to analyze this, I first found the proportion of questions that participants answered incorrectly on Quiz 1. I then checked for correctness of the related questions on Quiz 2 (i.e., those that covered the same material). The average proportion of questions answered incorrectly on both quizzes was 0.31, which suggests that most participants were not retaining incorrect information from Quiz 1 to Quiz 2.

Chapter 4: Discussion

The purpose of the present study was to examine whether the inclusion of post-discussion quizzes increased the efficacy of interteaching. College students reported to a classroom, took part in interteaching with or without a post-discussion quiz, and then returned 1 week later for a follow-up quiz. In summary, the participants who took a post-discussion quiz over prep-guide material performed significantly *worse* on a follow-up quiz than participants who instead completed anagrams. Moreover, participants in both groups performed significantly better on the tested questions (i.e., those questions that appeared on the post-discussion quiz and the follow-up quiz) than on the non-tested questions (i.e., questions that only appeared on the follow-up quiz). These results do not support my primary hypothesis that adding post-discussion quizzes would increase the efficacy of interteaching.

As mentioned previously, there have been several studies comparing interteaching to more traditional methods of college instruction (e.g., Saville et al., 2005; Saville et al., 2006). In the first experimental study of interteaching, Saville et al. (2005) found that participants in an interteaching condition performed significantly better on a quiz than participants in a lecture, reading, or control condition. Saville et al. (2006) subsequently compared interteaching to lecture in a graduate-level course and in an undergraduate-level course. In both courses, they found that students performed better on exams following interteaching than they did following lectures. Since Saville et al.'s (2005, 2006) initial studies, several researchers have reported similar findings (e.g., Cannella-Malone et al., 2009; Goto & Schneider, 2009; Scoboria & Pascual-Leone, 2009). Thus, interteaching seems to be an effective alternative to more traditional methods of

instruction. Nevertheless, as Saville et al. (in press) suggested, there are likely ways to improve interteaching further.

At first glance, the testing effect provides a promising means of increasing the efficacy of interteaching. To reiterate, the testing effect is a phenomenon in which taking a test over studied material enhances remembering more so than additional studying. Several researchers have observed this effect not only with tested material (Butler & Roediger, 2007; Roediger & Karpicke, 2006b), but also with previously non-tested, related material (Chan et al., 2006; McDaniel et al., 2007). Thus, in the context of interteaching, including post-discussion quizzes should enhance learning. This, however, was not the case in this study. Rather, participants who received a brief quiz after their pair discussions performed significantly worse on a follow-up quiz taken 1 week later than participants who completed anagrams after the discussions. There are at least three possible explanations for why adding post-discussion quizzes did not enhance learning.

First, participants in the quiz condition may have performed worse because of the negative suggestion effect (e.g., Butler, Marsh, Goode, & Roediger, 2006; Remmers & Remmers, 1926; for a full review see Roediger & Karpicke, 2006a). The negative suggestion effect is said to occur when respondents answer questions incorrectly and are not given correct information prior to the next testing period. Consequently, respondents may believe that answers given on the first test were correct and thus give similar responses on the second test. A number of researchers have observed this effect. For instance, Roediger and Marsh (2005) had participants read 18 of 36 prose passages and then gave them a multiple-choice test on all 36 of the passages. Tests consisted of questions with two, four, or six alternatives. All participants then completed a filler task

(brainteasers) for 5 min, followed by a cued-recall test. Participants performed better on questions when there were fewer alternatives. Interestingly, though, when participants gave wrong answers on the follow-up test, the incorrect answers often were those that had been incorrectly marked as the answer on the multiple-choice tests. Similar negative effects have been observed in studies of word recall (e.g., Meade & Roediger, 2006; Neely, Schmidt, & Roediger, 1983) and on true-false tests (e.g., Toppino & Brochin, 1989).

In the present study, participants in the quiz group may have answered questions incorrectly on Quiz 1, which then strengthened the likelihood that they would choose the same incorrect answers on Quiz 2. Given past research on the negative testing effect, especially with multiple-choice tests, it seems possible that the presence of "lures" (i.e., incorrect alternatives) on Quiz 2 may have prompted participants to make incorrect responses. That is, because participants in the quiz condition had the opportunity to answer incorrectly on Quiz 1, they may have remembered an incorrect alternative as correct on Quiz 2, even though that information may have been corrected in the clarifying lectures. In contrast, because participants in the no-quiz condition did not complete post-discussion quizzes, there was no opportunity for their answers on Quiz 1 to affect their answers on Quiz 2. Rather, their misunderstandings may have been corrected during the lectures, which then carried over to Quiz 2.

Although the negative suggestion effect is a plausible explanation, it, however, seems unlikely in the present study. There were few instances when participants in the quiz condition answered a question incorrectly on Quiz 1 and then again on Quiz 2. In fact, this occurred only about one third of the time, on average, which suggests that the

clarifying lectures served to correct any misunderstandings that participants had. Some studies from the testing effect literature show that feedback leads to better performance on follow-up quizzes (e.g., Cull, 2000; Pashler, Cepeda, Wixted, & Rohrer, 2005). Assuming participants were listening to the clarifying lecture, participants had the opportunity to receive feedback from the lecture, which presumably decreased the likelihood that they would get the same answers wrong on Quiz 2.

Rather, a better explanation for the present findings may be found in the preexisting contingences that are part of the interteaching format. According to Boyce and Hineline (2002), the clarifying lectures should function as reinforcers because students request, or mand, the information that is presented. Therefore, the lectures should reinforce some, or many, of the behaviors that preceded them: reading, discussing the material, asking questions, and so on—behaviors that presumably impact learning. Thus, any aspect of the lectures that makes them more reinforcing is likely to have a positive effect on learning. Similarly, any aspect of the lectures that makes them less reinforcing is likely to have a negative effect on learning.

One factor that affects the reinforcing nature of a stimulus is the amount of the stimulus that an organism has already "consumed" (e.g., Laraway, Snycerski, Michael, Poling, 2003; Michael, 1982). For instance, when an organism has eaten a good amount of food, the food loses some of its reinforcing value. In the same way, being exposed to large amounts of information in a classroom setting may have the same effect: The information may become less interesting, or reinforcing, to students. In this study, participants requested the information they wanted to hear in the lecture, but in the quiz condition, there was no break from the material. Therefore, it may have been that the

participants in the quiz condition were "satiated" with the material, and the lecture no longer served as a reinforcer (or, at least, was less reinforcing). Benjamin (2002) and McKeachie and Svinicki (2006) have noted that students' attention spans in lecture-based courses tend to be very limited (typically around 10 min). When breaks are introduced, however, lectures tend to maintain attending better. Because students in the no-quiz condition had a brief break from the material when they completed anagrams, the lectures may have been more reinforcing and thus might have functioned to improve their performance on Quiz 2.

Because this was a lab-based study, it is possible that in a typical interteaching classroom, where the discussions and lectures are usually separated by a day or more, students might still benefit from post-discussion quizzes. Having a separation between the discussions and lectures might preclude any "satiation" effects that occurred in the present study and thus further enhance the efficacy of interteaching. It is also possible that moving the quizzes to a different location within the method would make them more effective. For example, the quizzes could take place after the lectures. By having the quizzes after the lectures, students would have the opportunity to receive clarification before answering quiz questions. In this way, having correct information before taking the quizzes might ultimately produce a "positive suggestion effect."

A final possible explanation for the present findings is that interteaching already contains components that enhance learning such as active learning (Mathie et al., 1993; Yoder & Hochevar, 2005) and frequent feedback (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2007; Leeming, 2002). In interteaching, active learning occurs in a number of ways, for instance, when students complete the prep guides and when they discuss

material with the instructor. Mostly, though, active learning occurs during the pair discussions. These discussions allow students to practice skills that most teachers wish to promote. Mathie et al. (1993) and Yoder and Hochevar (2005), among others, have found that active involvement tends to promote learning. Similarly, interteaching includes feedback in a variety of forms. Students receive immediate feedback from their peers and the instructor during the discussion as well as delayed feedback from the instructor during the lecture, both of which seem to have a positive effect on learning and remembering (e.g., Agarwal et al., 2006; Leeming, 2002). Thus, interteaching may be an effective teaching method partially because it already includes both immediate and delayed feedback. Moreover, the general format of interteaching is such that students are already likely to engage in some (or many) of the behaviors that impact learning. As such, adding post-discussion quizzes to enhance remembering may do little to increase efficacy of interteaching (H. Roediger, personal communication, September 27, 2009).

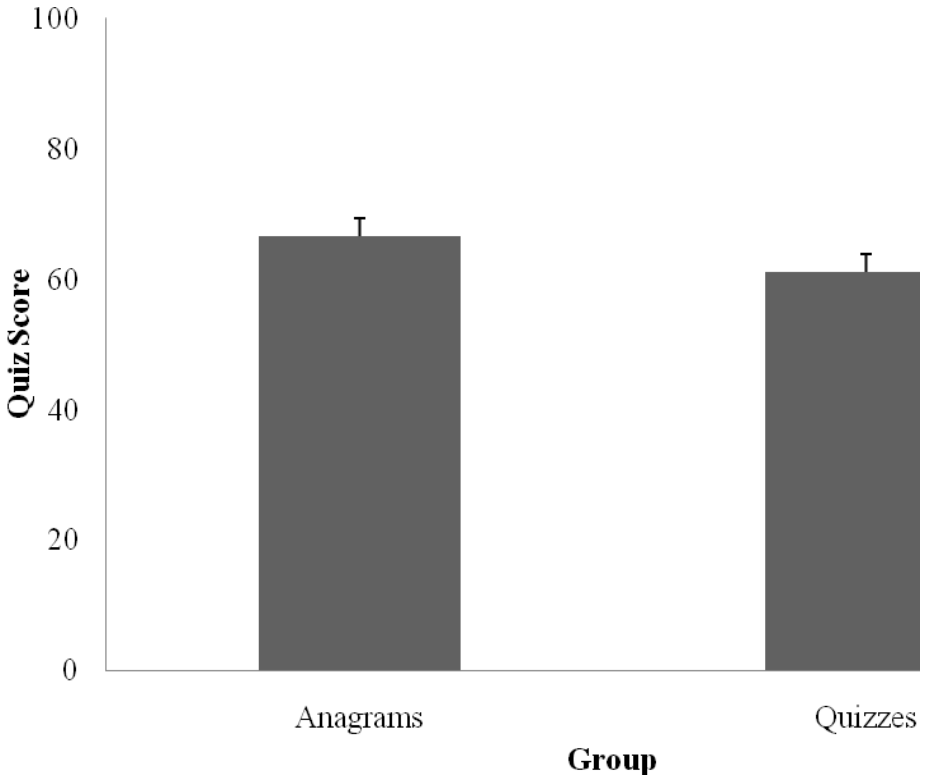

Figure 1. Mean Quiz 2 scores for each interteaching condition. Bars show standard error.

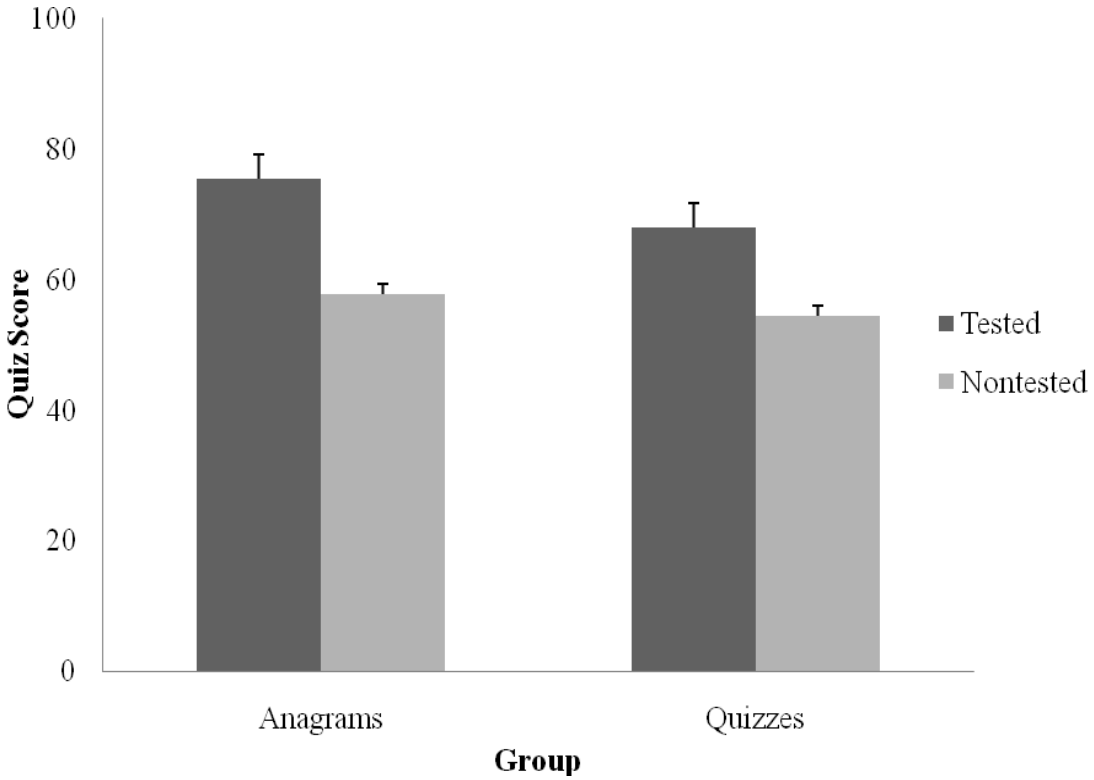

Figure 2. Mean Quiz 2 scores on tested and nontested items for each interteaching condition. Bars show standard error.

Appendix A

1. What is the "commonly known fact [that] appears in television commercials and even in publicity for nursing homes and hospitals" that this article is examining? Have you heard this fact before? Do you believe it to be true?

2. Approximately how many pet dogs and pet cats are there in the United States? How much do pet owners pay each year to keep these animals as pets? Do most pet owners view their pets as important parts of their families? If it costs so much to keep these pets as animals, what are some reasons why pet owners have pets?

3. Discuss the background research on pets and health (see bottom of p. 236 and top of p. 237). Do pets seem to have a positive effect on victims who suffered heart attacks? Explain. What about elderly people? AIDS victims? Children? Compared to talking to another person, how does talking to a pet affect cardiovascular responses?

4. Are there other benefits to pet ownership? If so, discuss what they are. If you are a pet owner, have you experienced any of these benefits? If you are not a pet owner, do you know others who may have experienced some of these benefits?

5. How do researchers typically assess whether pets have a positive effect on various cardiovascular measures (see section on "Theoretical Framework")?

6. What is the "main research question addressed in studies about pets and social support" (p. 237)? In these studies, why have researchers typically used animals that belong to the research participants? Do you think it makes sense to do this? Explain your answer.

7. Discuss the two studies conducted by Allen and her colleagues (Allen, Blascovich, Tomaka & Kelsey, 1991; Allen Blascovich, & Mendes, 2002) on p. 237 under "Blood Pressure and Pets: Some Research Findings." In general, how did dogs and cats affect blood pressure compared to friends and spouses? How did the author explain this "pet effect"?

8. What other explanations did the author list as possible reasons for the "pet effect?"

9. Explain why the author and her colleagues (Allen, Shykoff, & Izzo, 2001; see p. 238) decided to use "clinical randomized pet adoption" in their study of stress in stockbrokers? Describe how they carried out the study. Also, describe the general results of the study. How did taking blood pressure medication, Lisinopril, and having a pet affect blood pressure compared to simply taking Lisinopril?

10. In conclusion, is it accurate to say that pets lower blood pressure? Explain.

Appendix B

Please answer the following questions to the best of your ability…

1. What common phrase seems to be appearing on television shows and even in publicity for nursing homes and hospitals?

2. What do Americans spend $30 billion on each year?

3. Approximately what percentage of Americans describes their pets as important family members?

4. What is one of the health benefits of pet ownership?

5. In past studies examining the effects of pets on blood pressure, how are the pets usually obtained for use in the studies?

6. What is one explanation for the “pet effect”?

7. Briefly describe the results of these studies.

8. What is one shortcoming of studies that have examined if pets have an effect on health?

Appendix C

Please answer the following questions to the best of your ability…

1. "Pets lower your blood pressure" is a common phrase that seems to be appearing
    a. On television shows.
    b. In publicity for nursing homes and hospitals.
    c. On billboards.
    d. Both A and B.
    e. All of the above.

2. Today in the United States, there are approximately ____ pet dogs.
    a. 68,000
    b. 6.8 million
    c. 680,000
    d. 68 million
    e. 6.8 billion

3. Today in the United States, there are approximately ____ pet cats.
    a. 75,000
    b. 7.5 million
    c. 750,000
    d. 75 million
    e. 7.5 billion

4. Americans spend approximately _____ on their pets each year.
    a. $30 million
    b. $750 million
    c. $1 billion
    d. $10 billion
    e. $30 billion

5. Approximately 90% of Americans describe their pets as
    a. Important family members
    b. Annoying
    c. Their best friend
    d. As important to them as their own children
    e. None of these

6. Which of the following is NOT true regarding the health benefits of pet ownership?
    a. People who suffer heart attacks are more likely to live at least 1 year after the heart attack if they own a pet.
    b. People who own pets tend to have fewer stressful events than people who do not own pets.
    c. Elderly people who own pets typically make fewer visits to the doctor.

d. People with AIDS who own pets suffer less from depression than those who do not own pets.
e. Children who read in the presence of pets have lower blood pressure than children who do not read in the presence of pets.

7. In studies examining the effects of pets on blood pressure, how are the pets obtained for use in the studies?
    a. They typically belong to the researchers, who purchase them from breeders for the purpose of using them in studies.
    b. They belong to the participants because it is best to study how participants respond to their own pets.
    c. They are adopted by the researchers from humane societies so the pets can be saved and used in psychological research.
    d. They are owned by the universities where the researchers work and can be used by anyone wanting to study pet-human bonding.
    e. They are owned by the researchers, so they can know for sure how the pets will behave during the studies.

8. In one study, female participants who performed mental arithmetic in the presence of their pets had blood pressure that was _____ compared to when they completed the same task in the presence of their friends.
    a. lower
    b. higher
    c. the same
    d. lower at times, but higher at other times
    e. The results were not mentioned in the article.

9. In another study, female participants who performed mental arithmetic in the presence of their pets had blood pressure that was _____ compared to when they completed the same task in the presence of their spouses.
    a. lower
    b. higher
    c. the same
    d. lower at times, but higher at other times
    e. The results were not mentioned in the article.

10. One explanation for the "pet effect" is that:
    a. Pets serve as an entertaining distraction, which interferes with participants' ability to complete the mental arithmetic task, but also reduces stress.
    b. The presence of pets allows participants to relax during the mental arithmetic task, which often brings out the best in pet owners.
    c. Pet owners often get more exercise than non-pet owners, which allows them to concentrate better during the mental arithmetic task.
    d. People who own pets tend to be better at science and math than people who do not own pets.

e. People who own pets are more likely to have been raised in households where there was little stress, which has made them more focused in general.

11. What is one of the problems with previous pet studies that may make it difficult to determine exactly whether pets have positive effects on health?
    a. The people in these studies often didn't have many close friends
    b. The people in these studies typically don't own pets prior to the study
    c. The people in these studies may have been healthier to begin with.
    d. The people in these studies often didn't like pets.
    e. The people in these studies still experienced high levels of stress but were just more likely to report that they felt fine.

12. Which of the following most accurately describes the results of a study in which stockbrokers either (a) took blood pressure medication, or (b) took blood pressure medication and owned a pet, and then experienced a stressful situation?
    a. Stockbrokers who took the medication experienced a greater decrease in blood pressure than stockbrokers who took the medication and owned a pet.
    b. Both stockbrokers who took the medication and stockbrokers who took the medication and owned a pet experienced a significant decrease in blood pressure.
    c. Stockbrokers who took the medication and owned a pet experienced a greater decrease in blood pressure than stockbrokers who only took the medication.
    d. Both groups of stockbrokers experienced a significant increase in blood pressure while under stress.
    e. Although both groups experienced an increase in blood pressure under stress, stockbrokers who took the medication and owned a pet experienced significantly smaller increases in blood pressure than stockbrokers who simply took the medication.

13. Why did the researchers who studied the stockbrokers (from Question 12) use randomized pet adoption in their study?
    a. So they could be sure that everyone got a good pet.
    b. So they could be sure that the participants were completely health before pet adoption.
    c. So they could be sure that the pet owners and non-pet owners were similar before the study started.
    d. So they could be sure that the pets were similar before the study started.
    e. So they could be confident that the researchers were completely unbiased.

14. What was one other important finding from the study mentioned in Question 12?
    a. People who had the most friends benefitted the most from having a pet.
    b. The pets also showed fewer signs of stress at the end of the study.
    c. People with the fewest friends benefitted the most from having a pet.

d. The pets actually showed slightly elevated signs of stress from interacting with the stressed out stockbrokers.
e. There were no other important findings mentioned.

15. Which of the following most accurately describes the results of these studies?
    a. Pets lower blood pressure
    b. Pets increase blood pressure
    c. Pets help control blood pressure
    d. Pets have little long-lasting effect on blood pressure.
    e. Pets have no effect at all on blood pressure.

16. What is one shortcoming of studies that have examined if pets have an effect on health?
    a. Most researchers are animal lovers and are biased in their interpretations.
    b. The pets used on most of these studies have been relatively calm to begin with.
    c. Researchers have only examined how dogs affect health.
    d. The research participants are usually in poor health to begin with.
    e. Little research has examined negative effects that pets might have.

References

Adams, G., & Carnine, D. (2003). Direct Instruction. In H. L. Swanson, K. R. Harris, S. Graham (Eds.), *Handbook of learning disabilities* (pp. 403-416). New York, NY: Guilford Press.

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III, & McDermott, K. B. (2007). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861-876. doi: 10.1002/acp.1391

Ainsworth, L. L. (1979). Self-paced instruction: An innovation that failed. *Teaching of Psychology, 6,* 42-46. doi: 10.1207/s15328023top0601_13

Alba, E., & Pennypacker, H. S. (1972). Multiple change score comparison of traditional and behavioral college teaching procedures. *Journal of Applied Behavior Analysis, 5,* 121-124. doi: 10.1901/jaba.1972.5-121.

Allen, K. (2003). Are pets a healthy pleasure? The influence of pets on blood pressure. *Current Directions in Psychological Science, 12,* 236–239. doi: 10.1046/j.0963-7214.2003.01269.x

Antepohl, W., & Herzig, S. (1999). Problem-based learning versus lecture-based learning in a course of basic pharmacology: A controlled, randomized study. *Medical Education, 33,* 106-113. doi: 10.1046/j.1365-2923.1999.00289.x

Barron, K. E., Benedict, J. O., Saville, B. K., Serdikoff, S. L., & Zinn, T. E. (2007). Innovative approaches to teaching statistics and research methods: Just-in-Time Teaching, Interteaching, and Learning Communities. In D. S. Dunn, R. A. Smith, & B. C. Beins (Eds.), *Best practices for teaching statistics and research methods in the behavioral sciences* (pp. 143-158). Mahwah, NJ: Erlbaum.

Benedict, J. O., & Anderton, J. B. (2004). Applying the Just-in-Time Teaching approach to teaching statistics. *Teaching of Psychology, 31,* 197-199.

Benjamin, L. T. Jr. (2002). Lecturing. In S. F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 57-67). Mahwah, NJ: Erlbaum.

Binder, C. (1988). Precision Teaching: Measuring and attaining exemplary academic achievement. *Youth Policy, 10,* 12-15.

Binder, C., & Watkins, C. L. (1990). Precision Teaching and Direct Instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3,* 74-96. doi: 10.1111/j/1937-8327.1990.tn00478.x

Buskist, W., Cush, D., & DeGrandpre, R. J. (1991). The life and times of PSI. *Journal of Behavioral Education, 1,* 215-234. doi: 10.1007/BF00957005

Boyce, T. E., & Hineline, P. N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst, 25,* 215 - 226.

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L., III. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology, 20,* 941-956. doi: 10.1002/acp.1239

Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19,* 541-527. doi: 10.1080/09541440701326097

Cannella-Malone, H. I., Axe, J. B., & Parker, E. D. (2009). Interteach preparation: A comparison of the effects of answering versus generating study guide questions on quiz scores. *Journal of the Scholarship of Teaching and Learning, 9,* 22-35.

Catania, A. C., & Brigham, T. A. (1978). *Handbook of applied behavior analysis.* New York, NY: Irvington.

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

Connor-Greene, P. A. (2005). Problem-based learning. In W. Buskist & S. F. Davis (Eds.), *The handbook of the teaching of psychology* (pp. 70-77). Madlen, MA: Blackwell.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis.* (2nd ed.). Upper Saddle River, NJ: Pearson.

Corey, J. R., & McMichael, J. S. (1970). *Using personalized instruction in college course.* New York, NY: Meredith Corporation.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215-235. doi: 10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Dochy, F., Segers, M., Van de Bossche, P., & Gijbels, D. (2002). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13,* 533-568.

Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The Direct Instruction Follow Through model: Design and outcomes. *Education and Treatment of Children, 11,* 303-317.

Engelmann, S., & Carnine, D. W. (1982). *Theory of instruction: Principles and application.* New York: Irvington.

Engelmann, S., & Osborn, J. (1999). *Language for Learning: Teacher's guide.* Columbus, OH: SRA/McGraw-Hill.

Forlano, G. (1936). *School learning with various methods of practice and rewards* (Teachers College Contributions to Education No. 688). New York: Teachers College, Columbia University, Bureau of Publications.

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6(40).*

Gersten, R., & Keating, T. (1987). Long-term benefits from Direct Instruction. *Educational Leadership, 44,* 28-31.

Goto, K., & Schneider, J. (2009). Interteaching: An innovative approach to facilitate university student learning in the field of nutrition. *Journal of Nutrition Education and Behavior, 41,* 303-304. doi: 10.1016/j.jneb.2009.02.003

Hersh, R. H., & Merrow, J. (Eds.). (2005). *Declining by degrees.* New York: Palgrave McMillan.

Johnson, K. R., & Ruskin, R. S. (1977). *Behavioral instruction: An evaluative review*. Washington, D. C.: American Psychological Association.

Keller, F. S. (1968). Good-bye teacher . . . *Journal of Applied Behavior Analysis, 1,* 79-89. doi: 10.1901/jaba.1968.1-79.

Kinder, D., & Carnine, D. (1991). Direct instruction: What it is and what it is becoming. *Journal of Behavioral Education, 1,* 193-213. doi: 10.1007/BF00957004

Kulik, J.A., Kulik, C. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist, 34,* 307-318. doi: 10.1037/0003.066X.34.4.307

Kulik, C. C., Schwalb, B. J., & Kulik, J. A. (1982). Programmed Instruction in secondary education: A meta-analysis of evaluation findings. *Journal of Educational Research, 75,* 133-138.

Laraway, S., Snycerski, S., Michael, J., & Poling, A. (2003). Motivating operations and terms to describe them: Some further refinements. *Journal of Applied Behavior Analysis, 36,* 407-414. doi: 10.1901/jaba.2003.36-407

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29,* 210-212. doi: 10.1207/S15328023TOP2903_06

Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education, 147,* 62-81.

Lindsley, O. R. (1990). Precision Teaching: By teachers for children. *Teaching Exception Children, 22,* 10-15.

Lindsley, O. R. (1992). Precision Teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25,* 51-57. doi: 10.1901/jaba.1992.25-51

Maloney, M. (1982). Teaching the standard behavior chart: A Direct Instruction approximation. *Journal of Precision Teaching, 2,* 11-30.

Mathie, V. A., Beins, B., Benjamin, L. T., Jr., Ewing, M. M., Hall, C. C., Henderson, B., et al. (1993). Promoting active learning in psychology courses. In T. V. McGovern (Ed.), *Handbook for enhancing undergraduate education in psychology* (pp. 183-214). Washington, DC: American Psychological Association.

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. doi: 10.1080/09541440701326154

McDonald, J. K., Yanchar, S. C., & Osguthorpe, R. T. (2005). Learning from Programmed Instruction: Examining implications for modern instructional technology. *Educational Technology Research and Development, 53*, 84-98. doi: 10.1007/BF02504867

McKeachie, W. J., & Svinicki, M. (2006). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers* (12th ed.). Boston: Houghton Mifflin.

Meade, M. L., & Roediger, H. L., III. (2006). The effect of forced recall on illusory recollection in younger and older adults. *American Journal of Psychology, 119,* 433-462.

Michael, J. (1982). Distinguishing between discriminative and motivational functions of stimuli. *Journal of the Experimental Analysis of Behavior, 37*, 149-155. doi: 10.1901/jeab.1982.37-149

Morningside Academy in depth. (n.d.) Retrieved March 24, 2010, from http://www.morningsideacademy.org/about/indepth.php

Morris, C., & Kimbrell, G. (1972). Performance and attitude effects of the Keller Method in an introductory psychology course. *Psychological Record, 22,* 523-530.

Neely, J. H., Schmidt, S. R., & Roediger, H. L., III. (1983). Inhibitory priming effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 196-211. Novak, G. M., Patterson, E. T., Gavrin, A. D., & Christian, W. (1999). *Just-in-Time Teaching: Blending active learning with Web technology.* Upper Saddle River, NJ: Prentice Hall.

Pashler, H., Cepeda, N. J., Wixted, J., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3-8.

Remmers, H. H., Remmers, E. M. (1926). The negative suggestion effect on true-false examination questions. *Journal of Educational Psychology, 17,* 52-56. doi: 10.1037/h0070067

Rhem, J. (1998). Problem-based learning: An introduction. *The National Teaching and Learning Forum, 8*(1), 1-4.

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. doi: 10.1111/j.1745-6916.2006-00012.x

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

Roediger, H. L., III & Marsh, E. J. (2005). The positive and negative consequence of multiple-choice testing. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 31*, 1155-1159. doi : 10.1037/0278-7393.31.5.1155

Saville, B. K. (2008). *A guide to teaching research methods in psychology*. Malden, MA: Blackwell.

Saville, B. K., Lambert, T. L., & Robertson, S. (in press). Interteaching: Bringing behavioral education into the 21st century. *The Psychological Record*.

Saville, B. K., & Zinn, T. E. (2009). Interteaching: The effects of quality points on exam scores. *Journal of Applied Behavior Analysis, 42,* 369-374. doi: 10.1901/jaba.2009.42-369

Saville, B. K., Zinn, T. E., & Elliott, M. P. (2005). Interteaching versus traditional methods of instruction: A preliminary analysis. *Teaching of Psychology, 32,* 161-163. doi: 10.1207/s15328023top320B_6

Saville, B. K., Zinn, T. E., Neef, N. A., Norman, R. V., & Ferreri, S. J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39,* 49-61. doi: 10.1901/jaba.2006.42-05

Scoboria, A., & Pascual-Leone, A. (2009). An 'Interteaching' informed approach to instructing large undergraduate classes. *Journal of the Scholarship of Teaching and Learning.*

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24,* 86-97. (Reprinted in *Cumulative record, definitive edition,* pp. 179-191, by B. F. Skinner, 1999, Cambridge, MA: B. F. Skinner Foundation).

Skinner, B. F. (1968). *The technology of teaching*. New York: Meredith Corporation.

Sones, A. M., & Stroud, J. B. (1940). Review, with special reference to temporal position. *Journal of Educational Psychology, 31,* 665-676.

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

Toppino, T. C., & Brochin, H. A. (1989). Learning from tests: The case of true-false examinations. *Journal of Educational Research, 83,* 119-124.

Toppino, T. C., & Luipersbeck, S. M. (1993). Generality of the negative suggestion effect in objective tests. *Journal of Educational Psychology, 86,* 357-362.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175-184. doi: 10.1016/s0022-5371(67)80092-6

Watkins, C. L. (1988). Project Follow Through: A story of the identification and neglect of effective instruction. *Youth Policy, 10,* 7-11.

Yoder, J. D., & Hochevar, C. M. (2005). Encouraging active learning can improve students' performance on examinations. *Teaching of Psychology, 32,* 91-95. doi: 10.1207/s15328023top3202_2