Retrieval practice as a way to enhance learning and transfer in a high school biology classroom

by

Shannon M. Ralph

B.S., Washburn University, 1984
M.S., Fort Hays State University, 2003

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Curriculum and Instruction
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Abstract

Educators often seek methodologies that enhance enduring knowledge for their students if their goal is to teach effectively. Students who strive to improve their own learning often use less effective study strategies as a way to improve long-term memory or content transfer. If the goal of education is to teach students how to retrieve information that they have learned in the past to solve a novel problem, then teachers and students could benefit from evidence as to which strategies support this process. The goal of this study was to investigate retrieval practice in the form of quizzing as a way to improve teaching and learning in a high school biology classroom. A pre-test was given to all of the student participants prior to instruction in the targeted unit. Two retrieval quizzes were administered to the treatment group while the control group was given a review sheet covering the same information over which to study. The instructor provided the same content and activities to all students throughout the three-week time frame in which the study took place. A post-test was given to all students at the conclusion of the learning unit, and the scores were used to determine the differences in the learning gains that occurred between the treatment and control groups. Statistical analysis revealed that there were no significant differences in the post-test scores between the two groups. The non-significant results could have been influenced by one or more of the limitations of the study such as the low number of participants, the short time frame of the study, or using a testing instrument without demonstrated reliability. It would be beneficial to conduct further studies in authentic classroom settings as a way to identify effective teaching and learning strategies.

Retrieval practice as a way to enhance learning and transfer in a high school biology classroom

by

Shannon M. Ralph

B.S., Washburn University, 1984
M.S., Fort Hays State University, 2003

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Curriculum and Instruction
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Dr. Thomas Vontz

Copyright

Abstract

Educators often seek methodologies that enhance enduring knowledge for their students if their goal is to teach effectively. Students who strive to improve their own learning often use less effective study strategies as a way to improve long-term memory or content transfer. If the goal of education is to teach students how to retrieve information that they have learned in the past to solve a novel problem, then teachers and students could benefit from evidence as to which strategies support this process. The goal of this study was to investigate retrieval practice in the form of quizzing as a way to improve teaching and learning in a high school biology classroom. A pre-test was given to all of the student participants prior to instruction in the targeted unit. Two retrieval quizzes were administered to the treatment group while the control group was given a review sheet covering the same information over which to study. The instructor provided the same content and activities to all students throughout the three-week time frame in which the study took place. A post-test was given to all students at the conclusion of the learning unit, and the scores were used to determine the differences in the learning gains that occurred between the treatment and control groups. Statistical analysis revealed that there were no significant differences in the post-test scores between the two groups. The non-significant results could have been influenced by one or more of the limitations of the study such as the low number of participants, the short time frame of the study, or using a testing instrument without demonstrated reliability. It would be beneficial to conduct further studies in authentic classroom settings as a way to identify effective teaching and learning strategies.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This project would not have been possible without the guidance and support of my advisor, Dr. Thomas Vontz. I am thankful for his dedication and patience as I traveled this journey. I am also deeply grateful to my committee members: Dr. Lydia Yang, Dr. Todd Goodson, and Dr. J. Spencer Clark. Their selfless mentorship provided me valuable insights.

The support of my husband has been relentless, and without him I would not have finished this work. Thank you Brad for always reminding me of my purpose.

I deeply appreciate the support I received from my children. Michael, whose help and expertise guided my growth as I navigated this process. Jen was my sounding board and provided a listening ear when I just needed to vent. Kelli constantly cheered me on.

My parents, V.L. and Judy, provided unending and unwavering support from the very beginning. I learned the power of perseverance from them, and I am who I am because of the foundation they provided.

To my former and current students, this work is actually for you, because learning never stops. I strive to demonstrate that. That's how teaching works.

Dedication

I dedicate this work to my grandchildren:
Addison, Cameron, Clare, and Luke.

Learning is indeed a life-long process.

Preface

Educators often seek methodologies that enhance enduring knowledge for their students if their goal is to teach effectively. Pupils are exposed to vast amounts of information at a rapid-fire pace as they navigate their educational world in today's society. The current learning environment looks very different from the school of their parents and grandparents, and begs the question, "What does effective teaching look like that will meet our students' learning needs?" Some teaching strategies will provide higher learning gains (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Einstein, Mullet, & Harrison, 2012; Karpicke &Roediger, 2008; Koh, Lee, & Lim, 2018), and therefore, more adequately support the desired educational goals for teachers and students. One such strategy is retrieval practice. The researcher designed this study to investigate the effect of retrieval practice, in the form of short, intermittent quizzes, on long-term memory of students in a high school biology class. In addition, the study sought to determine if the quizzes would influence the ability of high school students to transfer their biological content knowledge to solve a novel, real-world problem. The researcher did not find significant learning differences between the treatment group, those administered quizzes, and the control group, those who studied a review sheet and did not quiz. The possible reasons for the lack of statistically significant gains are explored, and the implication for future research is proposed.

Chapter 1 - Overview of the Study

Educational practices from decades ago might inadequately prepare students to be productive and successful in the world today. A new learning paradigm has emerged to more efficiently meet the constantly changing needs of students. The Partnership for 21st Century Skills (P21) is an advocacy organization whose mission is to encourage the infusion of technology into education. National Education Association (NEA) was a founding member of P21, along with a wide range of business, media, and education-related organizations, and the Framework for 21st Century Learning was developed through this portal. Kivunja (2014, p. 85) explains the importance of developing 21st century fluencies when he said, "that whereas every educated person should have an appreciation of core skills in literacy and numeracy (e.g., the 3Rs of reading, -riting and –rithmetic), to succeed in the 21st century, an educated person must also have skills that enable him or her to think logically and to solve problems effectively and independently." Successful education in the twenty-first century will require methodologies leading to authentic learning that will deepen students' ability for critical thinking, problem-solving and collaboration (Lombardi, 2007).

Trilling and Fadel (2009), in *The 21st Century: Learning for Life in Our Times,* describe a learning and innovation skills domain which includes teaching students two skills: thinking critically and problem-solving. This domain states that students should be able to reason effectively, use systems thinking, make judgments and decisions, and solve problems. These types of abilities only work when students remember and internalize information; they must move it from short-term memory to long-term memory (Crockett, Jukes, & Churches, 2011). In addition, effective learning might allow students to make connections between their existing knowledge and new information (Crockett et al., 2011).

The researcher in this study was a veteran high school biology instructor with over 20 years of experience in the field of education. The purpose of the study evolved from her desire to investigate the effectiveness of an easily implemented teaching and learning technique (retrieval) in an authentic classroom environment. It is critical for teaching and learning strategies to respond to students whose experiences may be grounded in an environment quite different from that found in previous generations. To that end, the results of the study could provide value to students, as they seek to find ways to proactively learn information, and use that information to solve the inevitable problems encountered in life. The results could provide value to teachers who desire to improve their craft and increase the learning gains in their students. Finally, the results could inform educational preparation programs, providing future teachers with an additional tool to use as they enter the field of education.

The Kansas State Department of Education (KSDE) provides the learning criteria Pre-K-12 Kansas educators must consider in their instruction. KSDE, seeking to improve student education, defines a successful high school graduate as one who has "the academic preparation, cognitive preparation, technical skills, employability skills and civic engagement to be successful in postsecondary education, in the attainment of an industry recognized certification or in the workforce, without the need for remediation" (www.ksde.org/Agency/Fiscal-and-Administrative-Services/Communications-and-Recognition-Programs/Vision-Kansans-Can, retrieved 6/24/19). To summarize, effective teachers consider methodologies that build 21st century skills to better prepare graduating students to be successful.

Applied research has shown that retrieval practice improves long-term retention of targeted learning material (Agarwal, Bain, & Chamberlain, 2012). Evidence also supports the concept that retrieval practice will increase students' ability to have some degree of flexible use for the target content (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013). Butler (2010a) proposed that retrieval practice (test-enhanced learning) not only improves retention of content, but also improves the ability to transfer the information to different contexts. Lombardi (2007) suggests that authentic learning focuses on real-world problems and their solutions, which require transferability of information from school content courses to problems encountered in the work place, and is therefore an important component of education.

It is likely that educators teach with the goal for students to remember their instruction, and that the information learned provides knowledge from which to draw to solve problems when they are encountered. It was this presumption that provided the basis for the research questions in this study: 1) Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course? 2) Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

Considering memory systems is important in the context of teaching techniques that will improve the likelihood that students move content from short-term memory into long-term memory. In addition, a baseline of knowledge found in long-term memory provides a framework from which to draw when needed to solve real-world problems. Schacter & Tulving (1994) define a memory system in three ways: by its brain mechanisms, by the kind of information it processes, and by the principles of its operation. Long-term explicit memory, which is the focus of this study, is comprised of two mechanisms: episodic memory and semantic memory; together, these two systems result in declarative memory. Tulving (Bower, Donaldson, & Tulving, 1972) first distinguished these memory systems; semantic memory can be considered memory of facts which do not have to be personally experienced, while episodic memory is memory recollection from personal experience. Liu, Rosburg, Gao, Weber, Guo (2017) demonstrated that the memory system accessed during retrieval practice influences the power of the retrieval effect. They suggest that content reinstatement, versus semantic elaborative processing, provides a stronger long-term learning effect (Liu et al. 2017). However, the ability of students to recollect an episodic event that is specific to the content being studied is highly variable. Determining which memory system was employed during the retrieval events was beyond the scope of this study to define, but even so, the interactivity of both systems when processing memories has been shown (Greenberg, Keane, Ryan, & Verfaellie, 2009). The interdependence of both forms of memory can affect each other both at encoding and retrieval (Greenberg & Verfaellie, 2010); therefore, it is the presumption of the study that retrieval could provide learning gains regardless of the memory system employed. If learning gains are realized, then students have increased their base of knowledge with which to solve real-world problems.

The benefits from retrieval practice can be accomplished in many different ways (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013); however, in this study the researcher chose quizzes that were administered each week over the course of instruction. Researchers have suggested that the quiz format can vary (such as multiple-choice questions or short answer questions) without affecting the benefit of the recall practice (McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014). However, Larsen (2018) noted that writing multiple-choice questions which will provide maximum information retrieval and schema building is difficult. He suggested that other test formats might better accomplish the transfer of information to long-term memory. Additional evidence illustrates that multiple-choice questions are not as effective in promoting long-term retention because they require little to no effortful retrieval (Kang, McDermott, & Roediger, 2007). The Revised Bloom's Taxonomy (Forehand, 2011) provided insight when selecting questions that would be used on the quizzes given to the treatment group as a way to improve their effectiveness. As defined in the revised taxonomy, the remembering level asks students to retrieve, recognize, and recall relevant knowledge from long-term memory. The Bloom's level of analyzing requires students to determine how different parts relate to one another, using the skills of differentiating, organizing, and attributing. The researcher's retrieval quizzes used a mixed format, including both multiple-choice and short answer questions. The reasoning for using a mixed format of questions was to create an experience for students to do more than personal mental recall. For example, one multiple-choice question that was related to energy transfer in a food chain stated, "Suppose 10,000 units of energy are available at the level of the grasses. What is the total number of energy units that are available to the coyote?" The researcher added this question as a way to determine whether students could perform the calculation necessary to arrive at the right answer, as opposed to a more general question over energy transfer in a food chain. The inclusion of short answer questions on the quizzes allowed the researcher to provide questions that would require students to move to an analysis level of cognitive processing, possibly strengthening the retrieval process. For example, the Calvin Cycle was taught during the course of instruction, and one quiz question asked students to identify the main function of the cycle. Students were required to synthesize and analyze what they knew and then formulate an answer using their general knowledge of the Calvin Cycle, even though they were not explicitly taught the main function.

Another component of the research design included the decision to provide feedback following the administration of the retrieval quizzes. Support for the value of providing feedback varies in the research. For example, Pashler, Cepeda, Wixted, and Rohrer's study

(2005) included feedback for their participants for some items but not others, resulting in no benefit to long-term retention in successful retrieval events. However, feedback after an incorrect response improved final recall by 494%. Feedback is not always beneficial, and can be costly if the time used to review is spent on material that has already been mastered. The opportunity for a participant to skip feedback on already learned material in a computer-based exam could be ideal. However, it is more difficult to make this determination in a classroom setting using paper/pencil quizzes. The potential learning gains seemed to outweigh the cost in the current study's conditions. In addition, Roediger & Karpicke (2006b) suggested that feedback increases the power of the testing effect, supporting its use in the study's protocol. Feedback could allow students the opportunity to process incorrect answers and guide their study for the final cumulative exam. The researcher's hypothesis stated that this sequence of quizzing, reflection, and studying could improve students' skills for recalling information and then using that information to solve a novel problem.

Statement of the Problem

The assumption that teaching results in learning is a common misconception found in education (Larsen, 2018). K-12 schools are given the task of preparing students to be productive members of society. The 21st century skills necessary for student success requires the ability to think critically, problem solve, and apply innovations to novel situations (Trilling & Fadel, 2009). However, one must have access to information beyond its testing window in order to successfully navigate and solve problems. While educators understand that forgetting information is an inherent part of learning, the teaching process rarely accounts for it. The inability of students to master the skills of recall and transfer is a result of the lack of planning for long-term retention in many school settings (Larsen, 2018). While retrieval practice has been demonstrated to improve long-term retention of material (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Einstein, Mullet, & Harrison, 2012; Karpicke & Roediger, 2008; Roediger, Agarwal, McDaniel, & McDermott, 2011; Roediger & Karpicke, 2006), recall of information, even beyond the testing window, is not enough. In addition to access, students must also be able to solve a problem when one is encountered. Hostetter, Penix, Norman, Batsell, & Carr (2019) found that retrieval practice not only improved long-term retention of story structures, but also had the extra effect of making the information more accessible for use in identifying schematic similarities and applying to the problems when a hint was given. Hostetter et al.'s study (2019) did not support the idea that retrieval practice assisted students in realizing the relationship of information studied to the novel problem to be solved without a hint given. However, demonstrating that retrieval practice increased long-term retention of course content, and provided some level of transferability to questions in a different domain when provided a hint, has been demonstrated (Butler, 2010; Hostetter et al., 2019). The purpose of this study was to investigate whether students can 1) improve their long-term memory of content when intermittent quizzes are given over the course of a learning unit, and 2) use that information to solve a problem, even without a hint. If the goal of education is to teach students how to retrieve information that they have learned in the past to solve a novel problem, then teachers and students could benefit from evidence as to which strategies support this process. Studies demonstrating effective teaching approaches that will improve the opportunity to meet the educational goals of students could inform the pedagogy of teacher preparation programs, practicing educators, and students who strive to improve their ability to learn.

Purpose of the Study

A need exists in education to teach students skills providing them opportunities for fulfilling careers, community involvement, and societal contributions. Some key 21st century skills required to meet this need include critical thinking and problem-solving (Lombardi, 2007). Classroom instruction should model effective ways of applying content knowledge to novel situations. Retrieval practice has been shown to increase long-term retention of content (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Einstein, Mullet, & Harrison, 2012; Karpicke & Roediger, 2008; Roediger, Agarwal, McDaniel, & McDermott, 2011; Roediger & Karpicke, 2006). Retrieval practice also shows promise as to its effect in increasing a student's ability to transfer the knowledge in meaningful ways to novel problems (Rohrer, Taylor, & Sholar, 2010). Nguyen & McDaniel (2016) posited that retrieval allowed students' restudy time to be more effective because it not only permitted them to identify parts of a text they did not remember (in the context of their study), but it also enhanced their learning by providing an opportunity to place the information in a situation model. Dijk & Kintsch (1983) described the situation-model processing as mental representations of verbally described situations. In their study, quiz questions were used to provide optimum opportunities for students to retrieve information and also continue to build their own mental representations of the information in their long-term memory, thus providing a base-line of knowledge from which to draw to solve a real-world problem.

The purpose of this study was to investigate whether the use of retrieval practice in the form of short quizzes given over the course of an instructional unit increased long-term retention of biological content knowledge, and the ability to apply that knowledge to solve a novel, real-world problem. The study was conducted within the context of a high school biology classroom. The specific study questions were:

1) Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course?

2) Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

The researcher's hypothesis stated that the addition of retrieval quizzes given over the course of a unit of instruction would increase high school students' long-term retention of content knowledge, and also improve their ability to transfer their knowledge to solve a novel problem.

Methodology Summary

The researcher began the use of retrieval quizzes prior to the unit of study during the 2019 -2020 school year. Acclimating students to this methodology provided a way to decrease the negative effect that might result from students' perception of quizzing as a means of influence on their grades in the course. A pre-test was administered to all students at the beginning of the unit of study. The pre-test was used to provide information on the initial differences found between the groups, the magnitude of the differences, and aided in statistical analysis of the post-test data (Shadish, Cook, & Campbell, 2002). The researcher randomly divided five blocks of students into two groups: the treatment group received two retrieval quizzes, one each week during the unit of study. The control group studied a review sheet covering the same information as that found on the quizzes. A summative test was given at the end of the unit to both groups of students. The exam used for both the pre-test and post-test consisted of 15 multiple-choice questions and one graphing/analysis free response question. The multiple-choice test was used to gather data to answer the first research question: Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course? The short answer graphing/analysis free response question was used to gather data to answer the second research question: Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

Students from a public high school located in eastern Kansas who were previously enrolled in the researcher's biology courses for the 2019-2020 school year were recruited as participants in the study. Student placement into the researcher's blocks was a part of the ordinary enrollment procedure in the high school in which the study was conducted. Student scores were excluded from the study if students were not present to take both the pre-test and the final post-test assessment, or if they did not return the consent form signed by their guardians. In addition, students in the treatment group were excluded if they did not take both of the retrieval quizzes given over the course of instruction.

The experimental design of choice was a nonequivalent comparison-group design. The quasi-experimental design is weaker than one using random assignment, however, it was chosen because random assignment was not possible in the research setting in which the study took place. Both groups were given a pre-test prior to the beginning of the instructional unit which was the focus of the study. The pre-test consisted of a 15-question multiple-choice exam (to gather data to answer the first research question), and a graphing/analysis question (to answer the second research question). Over the course of instruction, the experimental group participants were given two quizzes, while the control group participants were given equivalent time to study a review sheet. The purpose of providing the review sheet to the control group was an effort to ensure an equal engagement opportunity with the unit's material. Thus, if a difference in the scores between the two groups occurred, the influence of the independent variable could be supported, rather than learning gains occurring solely because of additional exposure to the content (Slamecka & Katsaiti, 1988). Both groups were given a post-test (the same test as the pre-test), which was used for comparison between the groups after scores were adjusted for differences from their pre-test scores using analysis of covariance (ANCOVA).

The design of this study was uniquely nested within the construct of an authentic high school setting and as such, has several important components. First, the participants did not receive compensation or extra course credit for participating. High school students in this study were taking part in a required course (biology) for graduation, and the tasks and skills they were asked to master were a natural component of the course. Second, the final criterion test given at the end of the unit was a natural finale to the unit, and directly related to the content presented in the course. Lastly, the scores on the final assessment were used as part of the calculation of the students' grade in the course. These differences are significant because the environment of this study was an authentic high school classroom setting; therefore, the results provided additional information on effective classroom pedagogy that could be relevant to stakeholders within a public-school setting.

The design used in the study has a number of potential biases, including but not limited to selection bias, selection-maturation, selection-instrumentation, selection-testing, selection-regression, selection-history, and differential attrition (Johnson & Christensen, 2014). A pre-test was given to all participants as a means of reducing differential selection bias, allowing the exploration of the possible size and direction of the bias. In addition, the post-test scores were adjusted for any pre-test differences using analysis of covariance (ANCOVA). Students did not self-select the group in which they participated as a way to reduce selection maturation, selection testing, and selection history; the presumption being that the number of students who were influenced by any of these biases was relatively the same in both groups. Johnson & Christensen (2014) suggest that these two steps, the lack of self-selection and adjusting for pre-test differences statistically, will reduce bias, and result in a study with results that would be a closer approximation of those that would be reached using a stronger experimental design.

Significance of the Study

The purpose of the study was to investigate whether the use of retrieval practice improves high school students' long-term retention of biological content, and also improve their ability to transfer that knowledge to solve a novel problem. Results from the study could be significant for several reasons. First, students need to remember what they learn if the purpose of education is to provide foundational knowledge on which to make informed decisions. The goal of teaching is enhanced when the pedagogical approach used by educators is effective. Second, the ability of students to improve their capacity to learn would be heightened by using techniques that provide the greatest learning gains. If teachers use effective teaching and learning techniques, then students would have deeper experiences on which to draw when faced with the need to proactively engage their own learning. Third, persons use past experiences and knowledge to solve novel problems when encountered. Including teaching and learning practices that enhance the ability to apply foundational knowledge to solve a novel problem would greatly benefit student lives as they traverse the inevitable trials they will encounter as they move through life. Finally, teacher preparation programs would benefit by informing upcoming educators as to the most effective teaching and learning techniques that can be used in their own classrooms as a means to improve student learning.

Limitations of the Study

The potential methodological limitations in this study are provided in the following discussion. The project was conducted in a 6A public high school in eastern Kansas, and the data collected may not reflect data from the population of other high schools, especially ones in different geographic areas or those whose demographics differ from the study population.

The participating biology students were all underclassmen, and recruitment was limited to the pupils enrolled in the researcher's biology course. The small participant pool introduced several limitations. For example, the participating students may not have reflected the entire student population of the high school in which the study took place. The small number of participating students might have statistically limited the ability to find significant differences in the treatment group, if a difference existed. Finally, the inclusion of students from the researcher's classes (versus students from other teachers' biology classes) could have biased the results due to the researcher's explicit knowledge of the research protocol, including which blocks of students were receiving the treatment.

The short time period of the study (approximately three weeks covering one unit of content information) would suggest that the results may not be generalized to longer term learning over the scope of a semester, year, or beyond the unique setting of the sample. It is possible that the influence of the treatment could have taken longer than one unit to be realized. Students might have significant improvement in their scores if provided more time to practice retrieval. Although the time period was relatively short, a three-week time frame potentially allowed for history selection and differential attrition. Students' environments outside of the classroom might influence their response to retrieval quizzes and ultimately their post-test scores.

The time of year in which the study was conducted might have increased the stress of students, as it took place at the end of the first semester. The holiday season can insert additional stress for students, which could impact their learning potential. The lack of random assignment could have inadvertently placed more students with higher stress levels into one of the study's groups, skewing the results. The researcher acknowledges environmental biases that might have been introduced in the time-frame of the study. However, the design of the study included the typical schedule of many K-12 schools, providing results that could potentially be applicable to schools currently following a relatively strict adherence to a prescribed school calendar.

Lack of random assignment imposed an additional limitation to the generalizability of the results of the study. While students did not self-select into groups, background differences between groups possibly introduced selection bias that could not be corrected statistically. Using ANCOVA to analyze the post-test scores allowed the researcher to adjust the post-test scores on the pre-test differences found between the groups. However, authentic randomization was not possible because the students were placed in their biology classes by the typical enrollment procedure used by the school.

Using a testing instrument that has been evaluated for its reliability influences the degree to which a researcher can make claims on the results of the study. The multiple-choice exam that was used to gather data to answer the first research question (Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course?) was designed by the researcher using questions gathered from a variety of sources, such as published text-books and on-line state assessment items. The exam was not evaluated for reliability prior to its administration to students in the study. However, two published research projects used standard classroom materials for data collection

(McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011). Both of these prior studies demonstrated significant learning gains for students using retrieval practice, thus providing the reasoning behind the decision to use the test designed by the researcher.

The short length of the test (15 multiple-choice questions) might have placed an additional limitation on the study. Roediger, Agarwal, McDaniel, & McDermott (2011) found highly variable results on their end-of-the-year exam due to the small number of items on the exam, leading to non-reportable results. It is possible that the low number of test items did not provide enough data to measure learning gains if they occurred.

Students' responses to retrieval quizzes and their perceptions that the quizzes impacted their grade acquisition in the course introduced potential limitations. Steps were taken to decrease this limiting factor; an explicit explanation was given regarding the purpose of the quizzes as a study aid, and students had been following the quizzing protocol in the previous unit as a way to decrease selection-testing bias. Also, the points assigned to the quiz were very low relative to the number of points found on the post-test. Regardless, students may have performed differently due to their differential responses to quiz-taking.

Outside factors not related to the retrieval process might have contributed to the gain or lack of gain of knowledge. Students might have discussed the content at home or with peers from other blocks, even though they were explicitly instructed to avoid discussions or review of material with anyone outside of class. The possibility that students did additional studying beyond their normal study habits because of their interest in the topic was also present. Conversely, they might have done less studying than their normal study habits because of their lack of interest in the topic. Finally, a myriad of school factors might have influenced their time devoted to learning the content. Some examples of typical school activities include involvement in sports practices and events, fine arts practices, other coursework loads, or even interruptions in the normal school schedule with things such as fire and tornado drills.

Definition of Terms

The following definitions are given as a way to clearly communicate the purpose, methodology, and results of this study:

Memory consolidation: the process by which memory becomes enduring over the long-term; strengthening mental representations for long-term memory (Brown, Roediger III, & McDaniel, 2014). A hypothetical family of processes that take place both during wakefulness and during sleep at multiple levels of organization and function in the brain, from the molecular to the behavioral, and over a temporal spectrum ranging from seconds to months and years. The relatively fast molecular, synaptic, and cellular local mechanisms likely serve as repetitive subroutines in the mechanisms that embody slower systems consolidations, in which the experience-dependent information redistributes over brain circuits (Dudai, Karni, & Born, 2015).

Episodic memory: receives and stores information about temporally dated episodes or events, and temporal-spatial relations among these events; it is always stored in terms of its autobiographical reference to the already existing contents of the episodic memory store (Bower, Donaldson, & Tulving, 1972).

Long-term memory: memory for information that has not been in conscious awareness for at least several tens of seconds, but possibly for as long as several decades, prior to its retrieval (Postle, 2016).

Metacognition: students' own awareness of their learning.

Retrieval practice: recalling facts or concepts or events from memory (Brown et al., 2014). In the context of this study it refers to short, low-stakes quizzes. It is also known as the testing effect or test enhanced learning.

Schema: the general representation of knowledge a person has of a particular domain. A schema allows for encoding, storage, and retrieval of information related to that domain (Alba & Hasher, 1983).

Semantic memory: organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of these symbols, concepts and relations. Content words represent linguistic translations of information retrieved about general concepts and their interrelations (Bower, Donaldson, & Tulving, 1972).

Situation models: integrated mental representations of a described state of affairs (Zwaan & Radvansky, 1998); mental representations of verbally described events (Dijk & Kintsch, 1983).

Transfer: the influence that past learning has on new learning, and the degree to which the new learning will be useful in the learner's future (Sousa, 2011).

Conclusions

Education in the 21st century and beyond must meet the needs of students who need critical thinking and problem-solving skills to be successful (Lombardi, 2007). Teaching strategies that support this goal could be helpful to practicing educators and students, especially if those strategies work within the time frame of the rigorous schedule found in K-12 schools. The design of the study investigated the effectiveness of using retrieval practice in the form of quizzing in developing long-term memory and problem-solving skills in high school students enrolled in a biology course. The uniqueness of the study includes its conduction in an authentic learning environment of a public high school found in eastern Kansas over the time frame of one learning unit. The grade for the unit included the score students earned on the post-test assessment, and the material on the post-test was directly related to the unit of study. The data collected from the study could be beneficial because it could inform the pedagogy of classroom instructors, providing them a mechanism easily incorporated into classroom instruction while providing large learning benefits for their students. The study might also provide evidence to students that retrieval practice is an effective way to study, oftentimes as a way to meet their own learning goals.

Following this general overview, the remainder of this dissertation will discuss the prior research that has been conducted on this topic, the methodology used in this particular study, the results from this study, and finally, a discussion related to the findings of the study.

Chapter 2 - Literature Review

Learning in the 21st century and beyond must ensure that students acquire skills that will increase their abilities to be successful. Lombardi (2007) described authentic learning as that which focuses on real-world, complex problems and their solutions. Many practicing K-12 educators strive to incorporate pedagogy that supports authentic learning. The purpose of the study was to investigate a teaching and learning technique that has been demonstrated to be effective, and investigate the outcome on a learner's ability to remember biology content and then transfer that content to assist in solving a real-world problem. The following literature review provides instrumental background information used in supporting and defining the study and is divided into two sections. The first section is a description of memory formation; this segment will explain current theories describing how long-term memories are formed. The second section defines retrieval practice and identifies studies demonstrating its potential to be an effective teaching and learning tool.

Memory Formation

Long-term retention of information is important to provide the scaffolding upon which new information can be built. Memory is the capacity of an organism to acquire, store, and recover information based on experience (Josselyn, Köhler, & Frankland, 2015). Current theories explaining memory formation explain that memories are encoded as enduring physical changes in the brain, sometimes referred to as engrams (Josselyn, et al, 2015). An engram is not static; using retrieval causes a destabilization of a previously encoded engram, and then initiates a new consolidation cycle. As a matter of point, Josselyn, et al. (2015) demonstrated that activating engram neurons at times different than that of the time of encoding information is an effective way to induce memory. If indeed this process can be replicated in the framework of teaching, more specifically, if a teacher asks a student to retrieve information after time for forgetting has passed, then synaptic connections could be strengthened, thus increasing the probability that a student can recall that same information at a later time.

In contrast, some researchers suggest that retrieval practice can inhibit later recall of competing, non-tested material, a phenomenon referred to as retrieval-induced forgetting (Wimber, Rutschmann, Greenlee, & Bäuml, 2008). Chan (2009) investigated variables influencing the movement of information to determine whether cognitive thoughts are more likely to move into long-term memory or are forgotten. Chan concluded that long-term memory could be influenced by the level of integration achieved during coding, and the length of delay between retrieval and a test. Students are able to incorporate information into the same situation models if it is highly integrated, thus reducing competition. In addition, providing a delay before administering the final test seems to eliminate retrieval-induced forgetting because inhibition is transient (Bjork, Bjork, & MacLeod, 2013). There is conflicting information however, on the length of time in which retrieval inhibition occurs (Garcia-Bajos, Migueles, & Anderson, 2009). Taken together, teachers who implement the practice of retrieval might consider efforts to incorporate their instruction into current student schemas, and to space retrieval so it is not practiced immediately before a summative exam. Chan (2009) suggests that following this practice as a teaching tool could allow educators not only to determine what students know, as is a traditional approach to quizzing, but also change what students know, by improving the long-term memory of the concept, thus increasing the value of instruction.

Memory Systems

Various memory systems can be identified by the kinds of information they process and the principles by which they operate (Squire & Dede, 2015). Overall, memory systems are separated into two complementary learning systems: short-term memory (working memory) and long-term memory. Sousa (2017) described an information processing model explaining the flow of information between the two systems. His model includes seven operations involved in memory formation: 1) Selection: the brain selects what sensory information is important for processing, 2) the brain processes the sensory information and then 3) encodes the information. If the information is important and/or relevant, then in the 4^{th} step there is movement of the information into long-term storage. Once the memory has been disaggregated and stored, it can be retrieved (5^{th} step), and then used to decide how to act (6^{th} step). The 7^{th} step is described as forgetting, but this can happen at most any point in the sequence. The researcher built the framework for the current study within the context of this model.

Working Memory

Working memory has largely replaced the term short-term memory (Squire & Dede, 2015) and refers to the limited information that can be maintained in the mind. Sousa (2017) defines working memory as a temporary memory where conscious rather than subconscious processing occurs. The memories in working memory's activity occur primarily in the frontal lobes and allow for information to be built, taken apart, or reworked for storage in long-term memory. The capacity for working memory is limited (Mellanby & Theobald, 2014) and varies by age (Zhang, Zhao, Bai, & Tian, 2016). New information interacts with the pertinent information already stored in long-term memory as knowledge becomes available for consideration of memory storage (Mellanby & Theobald, 2014). Working memory's data is sent to the cortex for long-term storage, but this movement is influenced by whether the information makes sense (is more familiar to the learner) or has meaning; when new information exhibits both components, there is more cerebral activity and improved retention (Poppenk &

Moscovitch, 2010; Sousa, 2017). Past experiences can actually facilitate the encoding of new memories, perhaps because of repetition attenuation (Hutchinson, Pak, & Turk-Browne, 2016). Repetition attenuation for old information may be one process that increases a higher probability of new information moving into long-term memory.

Long-term Memory

The formation of long-term memory involves two brain components: the hippocampus and the neocortex. (See figures 2.1 and 2.2.)

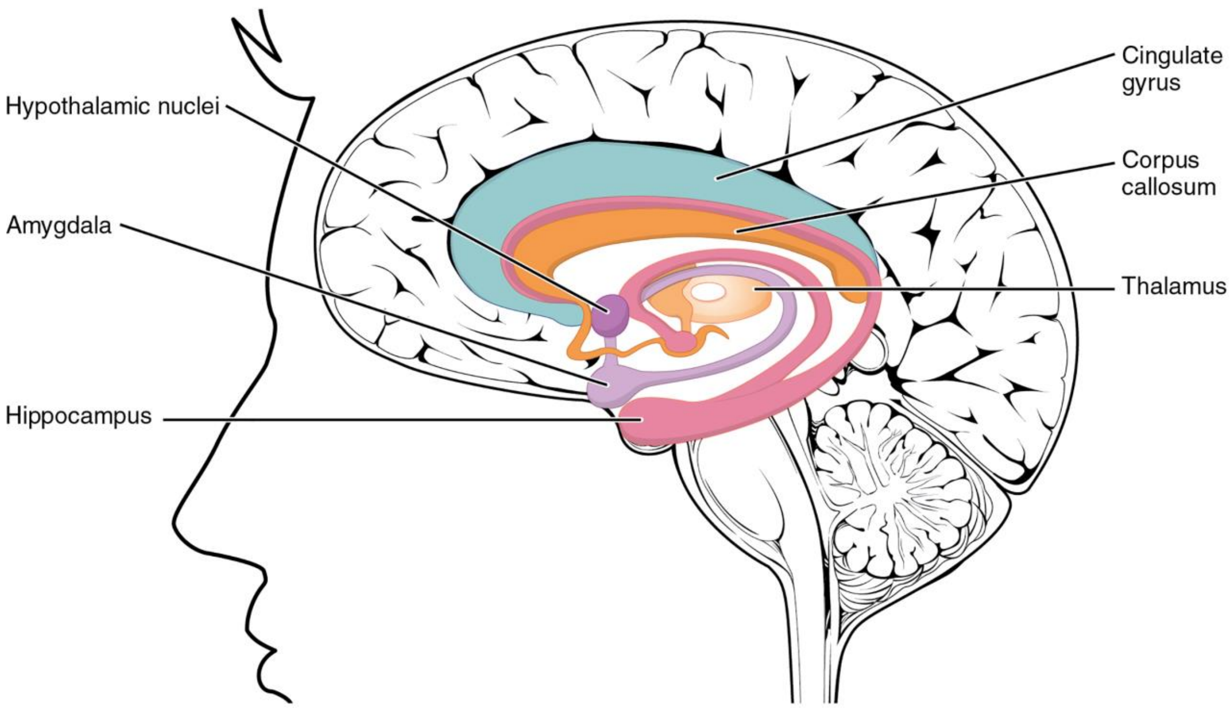


Figure 2.1 ***The Limbic Lobe***

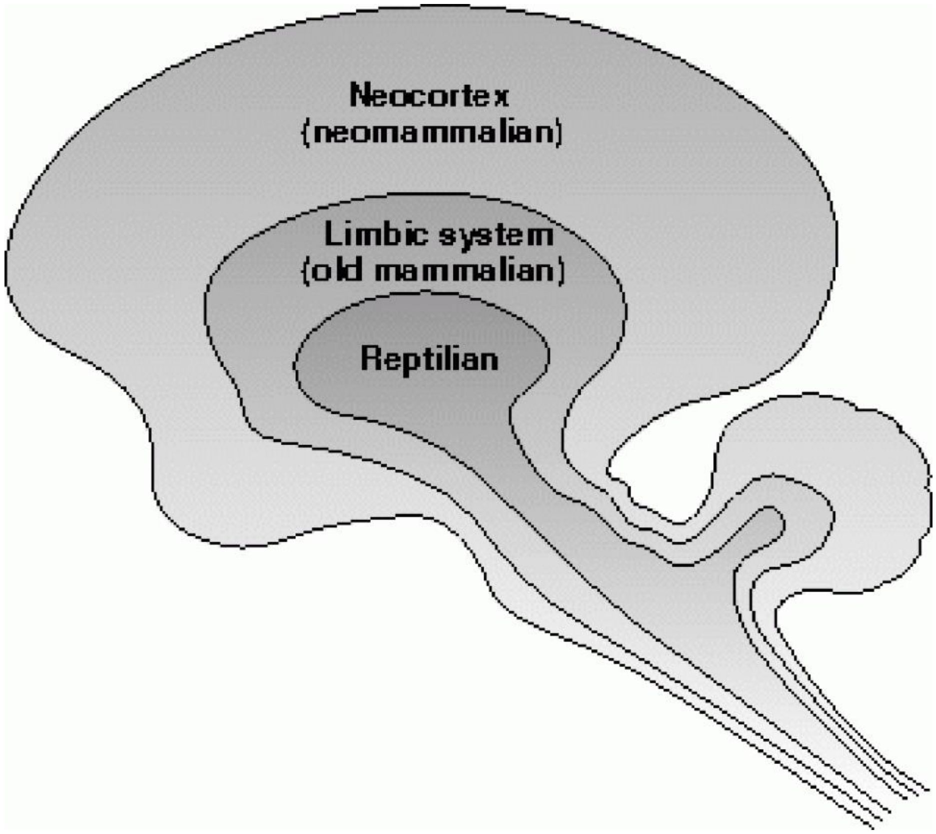


Figure 2.2 ***Neocortex in Relation to Surrounding Brain Structures***

The hippocampus rapidly encodes and binds together cortical associations, while the neocortex works slowly to store experiences. The interaction of these two systems allows for the formation of stable long-term memories (Antony, Ferreira, Norman, & Wimber, 2017). Studies in persons experiencing amnesia caused by brain damage or disease have revealed that the output of the hippocampus is relayed to many areas of the neocortex, areas involved in the processing and storing of information (Squire & Zola-Morgan, 1988). The hippocampus plays a major role in consolidating information and then converting it into long-term memory (Sousa, 2011). McClelland, McNaughton, & O'Reilly (1995) argued that the hippocampus and supporting structures actually allow the retention of content without the interference of previously acquired knowledge held in the neocortex. This allows for a memory stored in the hippocampal system to be reactivated and reinstated in the neocortex, providing information that can be used in behavioral responses. Dudai, Karni, & Born (2015) suggest that consolidation is comprised of the dissention between novel information and previously acquired and stored knowledge. Experiences that align with the knowledge schema that has already been stored may actually consolidate faster. In addition, the recall of information adjusts neocortical connections, allowing memories to gradually become independent of the hippocampal system. Multiple factors influence the relation between working memory capacity and long-term memory (Unsworth, 2016), including coding strategies (both effective and less effective) and study time allocation. While all factors have not been demonstrated to relate to working memory capacity, they are at least related to individual differences in long-term memory.

Long-term memory can be subdivided into two memory systems: an implicit memory system and an explicit memory system. Implicit memory is unconscious memory and is now referred to as nondeclarative memory (Squire & Dede, 2015). Nondeclarative memory system is involved in the development of skills, habits, conditioning, and other basic procedural tasks. Nondeclarative memory systems do not require intentional recall of experiences and are found primarily in the basal ganglia and cerebellum.

Declarative memory, also known as explicit memory, is conscious memory; this is the memory to which is usually referred when used in everyday language. Squire & Zola-Morgan (1991) found that this memory is primarily located in the hippocampus and structures which make up the parahippocampal gyrus. Bower, Donaldson, & Tulving (1972) distinguished two subdivisions of declarative memory: episodic and semantic, and described them as two information processing systems, both of which receive, retain, and transfer information to other systems. They differ in several ways, but generally, episodic memory is the memory of personally experienced unique episodes, such as an important birthday party or reception of an award. Semantic memory, on the other hand, stores general knowledge about the world that does not require a specific personal experience, such as knowing that Tallahassee is the capital of Florida, even if one has not actually been to Tallahassee. Retrieval from either system is usually entered as an episode into episodic memory, which may lead to changes in the contents and retrievability of those contents (Bower, et al, 1972). Retrieving information from episodic memory (conscious recollection) is contingent on the establishment of a special mental set, referred to as episodic retrieval mode (Tulving, 2002). Episodic memory's brain regions overlap and goes beyond the networks serving other memory systems. Even though the current model of memory systems separates declarative and nondeclarative memory, these two systems interact during memory consolidation (Albouy, Fogel, King, Laventure, Benali, Karni, Carrier, Robertson, & Doyon, 2015).

The Role of Forgetting

While one goal of education is to build long-term memories that can be recalled and used to make decisions, forgetting is actually an important component of a healthy memory system. Richards & Frankland (2017) describe the importance of balance between the development of long-term memory (to which they refer as persistence) and the necessity of forgetting (they term transience) in the ability to generalize information and adapt behavior to new experiences. In an on-going changing environment, the ability to remodel the neuronal connections in the brain to fit with environmental changes is an important component required to make relevant decisions. Therefore, long-term memory is significant only when that memory is able to provide a foundation on which to make those decisions.

Memory formation involves reactivation of neuronal patterns that are established in the initial encoding process (Richards & Frankland, 2017). If this holds true, then forgetting would involve a break between those same neuronal patterns. Balance between remembering and forgetting is important so that the brain eliminates outdated and non-useful information. Dong, He, Wang, Wang, Chen, & Zhong (2016) provided evidence for the value of forgetting in their study of autism susceptible genes in fruit flies. They found that the five genes affected not only the regulation of behavior, but also long-term memory formation. The defects in those genes caused the inability to forget, and their study suggested that the inability to forget is a cognitive mechanism manifested in autism. Therefore, memory formation is important when it supports purposeful and relevant decision-making skills, but is balanced with adaptive forgetting; effective instruction maintains this balance. The misunderstanding of the role of forgetting, and the techniques that recognize the need to interrupt forgetting such as quizzing and asking questions in class, often leads both teachers and students to use less effective learning strategies (Bjork & Bjork, 2019).

Some researchers disagree on the role that retrieval plays in inducing forgetting. Some studies suggest retrieval practice can actually impair recall of related materials because the practice of retrieval adds new traces into memory, competing with similar associations (Anderson & Bjork, 1994). Anderson (2003) provided evidence that interference is resolved by an organism's ability to recruit executive control processes; those processes that allow an organism to exercise control over thought processes. Thoughts have a response override in which an organism can select a weaker, but more appropriate response even in the face of interference from competing memory traces. Thus, forgetting is an active mechanism which is a response to prevent awareness of a distracting memory, and not necessarily a result of new associative learning. The act of retrieval, for example quizzing over new content material, does not necessarily compete with preexisting information, but rather can strengthen a selected response, even if it has a weaker association.

Chan (2009) conducted a series of experiments with the goal of determining the conditions under which retrieval practice induces forgetting versus enhancement of memory. His conclusion supported two key factors influencing memory: first, the level of integration during which initial encoding occurred, and second, the length of delay between retrieval practice and the final test. Students either update their current situation model, or they create a new one which will accommodate new information when encountering new material. Retrieval competition can be minimized by integration because it strengthens the association between the target and its retrieval competitors; in other words, students can use the target item as a retrieval cue to recall its related item (Postman, 1971). Evidence suggests that pieces of information stored in the same situation model do not interfere with each other (Radvansky & Copeland, 2006). Therefore, if new material is more easily integrated into students' existing schema, then fewer situation models are created, thus reducing competition.

Incorporating a delay between a retrieval event and a final exam is also an important component in determining the level of forgetting induced by retrieval. Bjork, Bjork, & MacLeod (2013) proposed that retrieval inhibition is a transient and flexible mechanism making the recall of target information temporarily less retrievable. Bjork, et al. (2013) suggest that retrieval inhibition shows a gradual decline over time, and Chan (2009) explicitly incorporated time differentials into his study. Chan found student performance increased when a delay occurred between the retrieval event and the final test, although his conclusion was based on the cumulative effect of both integration and time delay.

Wheeler, Ewers, & Buonanno (2003) provided strong evidence that retrieval practice slowed the rate of forgetting when compared to repeated study conditions, reinforcing the value of quizzing as a strategy to reach learning goals. Karpicke & Roediger (2008) also found that forgetting is largely influenced by the type of practice involved (effective versus ineffective), and repeated retrieval produces larger benefits for long-term retention. Conflicting information has been generated by some studies which found evidence for a long-term forgetting effect (Conroy & Salmon, 2005; Migueles & Garcia-Bajos, 2007. The conflict over retrieval's influence on forgetting is an area for continued research.

Episodic Context Theory

While the underlying mechanism of the influence of the retrieval effect is widely unclear, one proposed theory is referred to as the episodic context theory. An episodic retrieval mode is considered as a cognitive state in which a person consciously thinks of the past when he/she encounters a potential cue (Tulving, 2002). If retrieval includes the process of using retrieval cues to initiate recall of prior context information, then the episodic context theory suggests that the value of recollecting a previous episode improves the student's ability to distinguish one particular target from others (Gao, Rosburg, Hou, Li, Xiao, & Guo, 2016). Liu, Rosburg, Gao, Weber, & Guo (2017) concluded that episodic recollection occurring during retrieval practice is crucial for the retrieval practice effect. The 2017 study emphasizes the need for students and teachers to understand the value of the student's episodic processing involved in the testing effect, and how that might improve long-term retention. Chan (2009) also supported the need for integration of information into students' schema in order to reduce the number of situation models. Therefore, there is great value for teachers to find ways to incorporate any content information into student schemas when introducing new learning experiences.

In summary, evidence suggests that the practice of retrieval, for example in the form of intermittent quizzing, can aid in the movement of content information from working memory into long-term memory. Research supports the power of retrieval in providing the transfer of information to assist in the application of this knowledge to solve a novel problem (Butler, 2010). The current study seeks to investigate both phenomena within the context of a high school biology classroom.

Retrieval Practice

Active learning is important in science education; a robust science education enhances the skills necessary for students to succeed in a highly changing, variable world. The Next Generation Science Standards (NGSS) were written in response to a report from the National Research Council's document A Framework for K–12 Science Education (NAEP, 2008). The key reasons for developing a new science framework in 2009 were listed as: 1) publication for the first time of national science standards, 2) advances in science research that have increased knowledge and, as a consequence, have influenced the school curriculum in the fields of physical, life, and Earth and Space Sciences, 3) advances in cognitive research which have provided insights into how students learn about science, 4) growth in the prevalence of science assessments nationally and internationally, 5) growth in innovative assessment approaches, and 6) increased inclusion of formally excluded groups. NGSS now focus on crosscutting concepts, scientific and engineering practices, and disciplinary core ideas. NGSS standards suggest that the focus of instruction should allow teachers to teach in a way in which students learn best, which is a hands-on, collaborative, and integrated environment rooted in inquiry and discovery (KSDE, 2013).

David A. Sousa (2017) proposed that teachers who want their students to have retention of new information or skills beyond the immediate lesson must include rehearsal as a consistent part of student processing. In order to perform and master the scientific skills required within the context of NGSS, students must be able to do more than mere recall. Students must incorporate the information into their current schema so as to make sense of this new information. "The more you can explain about the way your new learning relates to your prior knowledge, the stronger your grasp of the new learning will be, and the more connections you create that will help you remember it later" (Brown, Roediger, & McDaniel, 2014, p. 5). One of the objectives of education is to maximize long-term knowledge retention and transfer (Hays, Kornell, & Bjork, 2010). The act of recalling information from memory, and then elaborating on that information and connecting it to prior knowledge enhances a student's ability to perform the complex degree of thinking that might meet the goal of improving students' foundational long-term memory. Memory provides knowledge from which to draw when solving real-world problems.

The foundational concept of retrieval practice (test enhanced learning, or the testing effect) stems from the construct that small, low-stakes tests are useful as learning tools rather than only assessment devices (Larsen, Butler, & Roediger, 2013). Quizzing serves at least two important functions: 1) it enhances the learning of the material, and 2) it informs students on material not mastered so students can focus future study time appropriately (Roediger, Agarwal, McDaniel, & McDermott, 2011).

Time availability constraints for learning are a reality in classrooms where lessons must be paced, and where students have limited time (or desire) to study. Therefore, the strategies that provide the greatest impact for learning should be considered/determined, and then this information can be disseminated to those in the educational field. For example, Hays, Kornell,

& Bjork (2010) demonstrated that learning was enhanced when instructional time was assigned to retrieval events instead of immediate feedback on correct responses. Many students and teachers still work under the assumption that testing only measures the learning of specific content material (Karpicke & Roediger, 2008) and use less effective strategies while learning new concepts For example, college students tend to use note-taking, highlighting, and rereading as their key study strategies (Karpicke, Butler, & Roediger, 2009) even though using some kind of test such as low-stakes quizzes might provide students with more accurate feedback related to what they do and do not understand, leading to a more effective use of subsequent study time. When undergraduate college students (from Washington University in St. Louis) were provided surveys and asked to identify study strategies they currently use, they reported that repeated reading was the most frequently used strategy; 84% of the students surveyed used this as their key tactic. In addition, only 11% of students used retrieval practice as a strategy, and only 1% used recall as their key study strategy (Karpicke & Roediger, 2009).

Even though spending extra time maintaining or holding content knowledge in memory does not by itself promote learning has been demonstrated (Craik & Watkins, 1973), many students continue to use short-term memory strategies as their key study approach. Students using these less effective strategies sometimes have the illusion of competence because they develop familiarity with the text, but not necessarily a deep understanding of the content; they base their assessment of learning on their fluency with the text, even though it does not assess their long-term retention of the knowledge (Karpicke & Roediger, 2009). Soderstrom & Bjork (2014) refer to this misrepresentation of long-term memory as the foresight bias and define it as an overestimation of one's future memory performance because of the difference between study and test situations. Roediger & Karpicke (2006b) demonstrated this lack of metacognitive awareness and supported the idea that students tend to be poor judges of their own learning without some sort of test providing information which is a better predictor of their knowledge. In addition, Nelson & Leonesio (1988) found that the *feeling* of knowing is related to study-time allocation; if a person can recall an item from short-term memory, that person might conclude that the information has been mastered. However, when tested at a later time, the person might be incorrect, even if an item could be recalled from short-term memory. Nelson & Leonesio's study (1988) supports the idea that persons do not necessarily understand the difference in memory processing relative to short-term or long-term memory. Even students who do use retrieval as a study method typically do so to assess what they have or have not learned, and not as a way to improve their long-term retention of the content (Dunlosky, Rawson, & McDonald, 2002).

Many variables influence the power of the testing effect. For example, Pyc & Rawson (2009) found that the degree of difficulty of recall practice is related to the result on the final test; the more difficult successful retrievals led to higher final scores when compared to easier successful retrievals. Additional studies have used different formats of test questions (free recall vs. multiple-choice), different study/test arrangements in their experiments, or cued recall tests (for a review see Eisenkraemer, Jaeger, & Stein, 2013). Regardless of the format, greater retention of content is achieved when a student recalls and then recodes the information as supported by most research.

Conflicting information on the testing effect exists. Two studies (Von Gog & Sweller, 2015; Leahy, Hanham & Sweller, 2015) suggested the testing effect is minimized, or even disappears, as the complexity of the content increases. Their deduction stated the testing effect only occurs when context has low-interactive materials (information learned without reference to the other elements in the task), and significantly decreases with highly-interactive materials, which they argue is the type of complexity required in most learning. Their proposed explanation for this phenomenon was the heavy load on working memory when confronted with highly-interactive materials, decreasing the movement of information into long-term memory. (For a response providing an alternative conclusion to Von Gog & Sweller's study see Karpicke & Aue (2015)).

Retrieval Practice in Biology

Prior studies have demonstrated that quizzing is an effective learning tool in undergraduate introductory biology courses (Carpenter, Rahman, Lund, Armstrong, Lamm, Reason, & Coffman, 2017; Orr & Foster, 2013; Walck-Shannon, Cahill, McDaniel, & Frey, 2019). The correlation between success on summative exams and participation in quizzing provides a way to measure the effectiveness of the quizzes as related to improving student understanding and memory of course material. For example, Carpenter, et al. (2017) provided undergraduate biology students optional on-line review questions that could be accessed as either test questions in which the question had to be answered prior to receiving the correct response, or read questions in which the question and the answer were provided up front. Students who accessed the review in the test format had higher scores on subsequent exams compared to the students who accessed the review in the read format, or those who did not access the review at all. Participation in quizzing improved summative test scores (Orr & Foster, 2013; Carpenter, et al., 2017; Walck-Shannon, et al., 2019) supporting the value of quizzing as a learning tool.

Explaining the worth of quizzing to students furthers the effectiveness of quizzing as a learning tool. Conceptually, this was demonstrated by Carpenter, et al. (2017) when student participation in quizzing increased after the instructor provided data displaying an increase in test scores for students who participated in the quizzes. Providing evidence for an effective learning tool is important, because students tend to otherwise rely on less effective study strategies, such as rereading, highlighting, and underlining (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Einstein, Mullet, & Harrison, 2012).

In order to implement quizzing in introductory biology classes, manageable means of delivery must be identified that can feasibly be implemented in such large enrollment settings. One such method of quizzing a large number of students without direct instructor oversight is to provide online access to quizzes over content presented in lecture. For example, as noted above, Carpenter, et al. (2017) illustrated that online quizzing could be used as a means to improve summative test scores. Thus, online quizzing may be a promising learning tool for introductory biology classes.

The researcher was interested in investigating the use of quizzes within the context of a high school biology classroom while studying the complex phenomenon of photosynthesis. There are key differences between high school and college students that might impact the effect of retrieval quizzes. The environment of a high school classroom is quite different than that found in an undergraduate institution: students are younger, student attendance is required, students typically have more adult oversight, and passing biology is required for high school graduation. An additional area of research is important to investigate the effectiveness of quizzing in a high school biology course, which could provide valuable information for high school instructors whose goal is to improve student retention of content material.

First Research Question

The first research question was designed to investigate retrieval practice's effect on the long-term memory of students over their understanding of photosynthesis. Students do not have a large background of photosynthetic knowledge is the assumption for high school students who are in a biology classroom. Using an effective teaching tool could benefit students' learning process as photosynthesis is typically a difficult concept for entry level biology students. The phrase backward testing effect describes the process of retrieval practice and its enhancement of long-term retention for target material; for examples of this effect see Roediger & Karpicke (2006a). The phrase backward testing effect has been used synonymously with testing effect or retrieval in the context of this study.

Second Research Question

Recent studies have suggested that there is an additional benefit of testing which is that retrieval of prior information facilitates learning of new content as well. The phrase forward testing effect is defined as the improvement of encoding and retention of new information due to the testing of previously studied information (Yang, Potts, & Shanks, 2017). The additional value of the forward testing effect is its influence on the learning of information not necessarily related to the previously tested material (Pastötter & Bäuml, 2014). The forward effect possibly generalizes to different kinds of information, including educational learning environments. At least two theories explain the forward testing effect as listed in a review of recent studies (Pastötter & Bäuml, 2014), and hypothesize that 1) a test permits students to use list-specific context cues and create a more focused memory search which may reduce or even eliminate interference from non-target material, or 2) recall testing improves encoding of the target material presented after the test; it allows for a reset of the encoding process, either making the encoding of later lists as effective as the earlier ones, or a change in student encoding strategy.

Studies supporting the reset of encoding also show that recall tests help students maintain a higher level of attention to encoding over the course of a lecture by encouraging task-relevant activities like note taking and discouraging task-irrelevant activities like mind wandering (Szpunar, Khan & Schacter, 2013). In addition, Yang & Shanks (2018) provided strong evidence that retrieval tests given over the course of instruction also improve inductive learning, the process in which people generalize their previous experience when making inferences that go beyond direct experience. Thus, interim testing might enhance the encoding of exemplars given following the test, and the testing of exemplars serves to consolidate them, providing both a forward and backward learning benefit, greater than that of restudy (Lee & Ahn, 2018; Cho, Neely, Crocco & Vitrano, 2016).

The second research question (Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?) was investigating the forward testing effect in the context of this study. Butler (2010) demonstrated that repeated retrieval of information produced superior transfer to several different types of questions, leading to investigate whether the same transfer occurs in high school students in a biology classroom. lThe testing effect could carry high significance to teachers and students alike as the teaching community seeks to provide students with tools to use to solve unique, real-world problems,

A final consideration for the value of the testing effect, both backward and forward, is found in the advantages provided to learners regardless of their age. Meyer & Logan (2013) studied students outside of the university population and explicitly considered three populations of people: younger university students aged 18-25, younger community adults aged 18-25, and older community adults aged 55-65. While the performance between the groups was different, all three groups benefited from testing similarly, suggesting that the testing effect is not isolated to a particular age group. The evidence that the testing effect is a powerful learning tool in different age groups might provide teachers evidence that this tool would be valuable to implement, regardless of the age of the students they teach.

Chapter 3 - Methodology

Chapter three explains the methods used in carrying out this quasi-experimental quantitative study. The research site and participants are clearly identified, as well as an explanation for the convenience sampling method used to recruit participants. Two separate tests were developed to provide data to answer the two research questions, and the processes used in test development are provided in this chapter. Finally, detailed procedural protocols are listed for both the experimental design and data analysis.

The Research Design

The purpose of this study was to investigate whether the use of retrieval practice in the form of short quizzes increased long-term retention of biological content knowledge, and the ability to apply that knowledge to solve a novel, real-world problem. The specific study questions were:

1) Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention for high school students in a biology course?

2) Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

Random assignment is the best technique for equating the comparison groups on all variables at the start of an experiment. Random assignment is the key criterion that allows one to make a strong claim of cause and effect in an experiment (Johnson & Christensen, 2014). However, in the context of a typical high school enrollment process, students are assigned into classrooms before the beginning of a school year. The enrollment procedure does not allow the flexibility needed to randomly assign students into groups. It was for this reason that the design used in this study was quasi-experimental, specifically, nonequivalent comparison group design (Johnson & Christensen, 2014). The researcher was unable to assign students to a specific instructional block, but each of the five blocks of students in the study (nonequivalent) was randomly assigned to either the treatment or control group as an effort to control potential biases in the nonequivalent comparison group design. The research design's protocol specified that both the experimental and control groups were given a pre-test. A post-test was administered to both groups after the treatment was applied to the treatment group. The post-test data was analyzed using Analysis of Covariance (ANCOVA) as a way to statistically adjust for pre-test differences on post-test scores, because of the lack of random assignment of students into each of the two groups.

Results from a quasi-experimental design must not be given the same weight as those from a stronger experimental design which includes random assignment of participants. A weaker experimental design that does not include random assignment reduces internal validity of the study. Additional limitations to the study included:

-short time frame of study (one instructional unit)

-time of year in which the study took place

-potential cross talking between students in control and treatment groups

-low number of participants

-using an instrument with unknown reliability

-short length of the multiple-choice test

-using the researcher's students as participants

The Research Site

The district in which this study took place is one of the major school districts located on the edge of a metropolitan area in eastern Kansas. One corner of the district includes a small slice of one of the suburbs of the city, while the opposite corner includes a small town whose main industry is agriculture (farming and ranching). The district was comprised of one high school, three middle schools and seven elementary schools at the time of the study. Specifically, this study took place at Washington High School, which included grades 9-12 and had an enrollment of 1739 students. (Washington High School is a fictitious name, used to preserve confidentiality.) The attendance rate for the 9^{th} grade students in this school was 95.2685%, while the rate for 10^{th} grade students was 93.7520%. (These numbers include attendance from 8/13/2019-12/19/2019, encompassing the dates of the study.) The administrative structure of WHS included a principal, four assistant principals and one principal intern. The principal and all assistants had divided responsibilities of discipline of students and evaluation of the teaching staff. Four counselors carried a caseload which included a portion of the student body; each was assigned a specific grade level and then worked with the same students all year. The school also employed a head counselor who oversaw the counseling program and coordinated all of the student support services. Two social workers were on staff, along with one school psychologist. Staff members totaled approximately 147, including all of the administration, teachers and support staff.

The Research Participants

Washington High School (WHS) had an enrollment of approximately 1739 students at the time of the study. The researcher used convenience sampling as a means of procuring participants, offering students who were enrolled in her biology courses the opportunity to participate. Each block of participating students was randomly assigned to the groups to be studied in the experiment. The purpose of randomly assigning blocks was to produce groups that were as similar as possible prior to the introduction of the treatment. Random assignment is the key factor that allows one to make a strong claim of cause and effect in an experiment (Johnson & Christensen, 2014).

The participants in this study were from the 9^{th} and 10^{th} grades; total enrollment at WHS for the 9^{th} grade was 491 and the 10^{th} grade was 438 students. The pupils in this school self-reported their ethic identification (the school registrar provided the categories). The 9^{th} grade included six Asian students, 18 Black students, five Hispanic students, two multi-cultural students, and one American Indian; all other students were White. Therefore, the percent of minority students in 9^{th} grade was 6.5%. The self-reported ethnic identification in 10^{th} grade was three Asian students, 13 Black students, six Hispanic students, and one multi-cultural student; all other students were self-reported as White. The percent of ethnic minority students in 10^{th} grade was 5%.

Special education students were streamed into regular education classes when possible at WHS. The number of special education students in the 9^{th} grade for the time-frame of the study was 58 students, and 10^{th} grade had 55 students. In the classes in which the study took place, only one student had an IEP on file identifying the student as requiring special services. This student, however, did not return the consent form, therefore, his score was not included in the study. Four students in the study had an IEP on file identifying them as gifted; two were in the treatment group, and two were in the control group. However, one of the students in the control group did not take the free response portion of the pre-test, and therefore his scores were not included in the final data set. Students who had health issues that could impact learning, such as asthma or ADHD, were identified as well. Thirty students had some kind of health alert in the classes in which the study took place.

The number of students who took advantage of free and reduced lunches provided information on the socioeconomic status (SES) of the school population. In order to receive free/reduced lunch services in the National School Lunch and/or School Breakfast Program, families must meet certain income guidelines. For example, if a family of four had a household income of $33,475 or lower, they would qualify for free benefits; if their income was $47,638 or lower, they would qualify for reduced benefits (KSDE.org). The number of students attending WHS who received free lunch services was 292, while the number of students who received reduced lunch services was 204. The students participating in the study who qualified for free/reduced meals were not specifically identified, but one can infer that the percent of students included in these programs is similar as that found school wide; approximately 16.79% of the students received free lunches, while 11.7% received reduced fee lunches. The SES status is relevant because student poverty in the home and community is the single most important predictor of science teaching and learning (Wiseman, 2012). However, SES status was not used as a variable in the study, the SES statistics provided are a means of providing general context of the participants in the study.

Washington High School operated on an alternating block schedule; students attended four different blocks on an every-other-day basis; routinely, each block lasted for 83 minutes. The study encompassed five blocks of biology students based on the number of blocks that were assigned to the researcher during the time-frame of the study. Biology courses at WHS included both freshmen and sophomores; in this study there were seven freshmen and 27 sophomore students who were in the treatment condition, and 12 freshmen and 17 sophomores who were in the control condition. The number of participants for both groups were the students who met the requirements of inclusion in the study; specifically, they returned signed consent forms, they were present to take the pre and post-test, and those in the treatment group were present to take both of the quizzes given over the course of the unit of instruction. The placement of the students into each block occurred during the enrollment process prior to the beginning of the school year, and as a result, the ratio of freshmen to sophomores was random. An on-line random order generator was used to determine to which condition each block was assigned; thus, the freshmen to sophomore ratio was beyond the researcher's control.

Evaluating ACT scores can provide a general sense of the level of educational learning at WHS. The ACT consists of curriculum-based tests of educational development in English, mathematics, reading, and science designed to measure the skills needed for success in first-year college coursework (www.act.org). The composite average ACT score for the last five years are provided Table 3.1.

Table 3.1.

WHS Composite ACT Scores 2015-2019

2015	23.5
2016	23.6
2017	22.7
2018	22.5
2019	22.0

College readiness benchmark scores refer to the minimum score needed to indicate a 50% chance of obtaining a B or higher, or about a 75% chance of obtaining a C or higher in the corresponding credit-bearing college course. The ACT scores from WHS indicated that 47% of students who took the ACT in 2019 met or exceeded the benchmark score of 23 on the science portion of the ACT.

Student Recruitment

Washington High School had a total of five instructors who taught biology for at least a portion of their day. The instructors were not required to have equal pacing of content material, nor were they required to use the same instructional materials. The researcher sought to investigate the effect of retrieval practice implemented over the course of an instructional unit on student learning. The study's design necessitated similar instruction, similar pacing, and similar activities to occur for five blocks of students who met at different times during the school day as a way to reduce the influence of confounding variables. The researcher could not control the instruction in classrooms outside of her own, and therefore the participants were recruited from the pool of students assigned to the researcher's classroom via the typical school enrollment process. In some research situations, the use of a researcher's students is integral to the research, particularly in the area of teaching methods, curricula, and other areas related to the scholarship of teaching and learning (https://www.boisestate.edu/research-compliance/irb/guidance/own-students-research-subjects/). The potential bias that might exist with students who are asked to participate in their instructor's research includes the feeling of forced compliance as related to earning a grade in the course. Research using the researcher's students is acceptable if students are provided a signed consent form clearly stating the voluntary nature of participation, the confidentiality of test scores, and the ability to withdraw at any time. The research was approved by the review board of Kansas State University, the university in which the study took place.

Consent Forms

The researcher introduced the research's protocol and explained the content of the consent form to all of the students who were enrolled in her courses. Students were asked to take the consent form home, explain the research to their parents or guardians, obtain a signature acknowledging their consent for their student's participation, and then return the form to the researcher. The researcher sent an email to all parents/guardians who had email addresses on file with the school to ensure that they were aware of the study and explained the availability of the researcher to answer questions or concerns. In addition, the researcher asked students to sign the consent form indicating their approval for their test scores to be included in the research results. Students were excluded from the study if they did not return the consent form, if they or their parents/guardians chose not to allow their participation, or if they did not take both the pre-test and the post-test. Finally, students in the treatment condition were excluded if they were not present to take both of the retrieval quizzes. These criteria excluded 20 students from the treatment group, and seven students from the control group, decreasing participation from the total class enrollment by 60% in the treatment group and 24% in the control group.

Instrumentation

The researcher developed two instruments used for data collection: a multiple-choice test and a graphing/analysis free response question. The purpose of the multiple-choice test was to gather data to answer the first research question: Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course? The graphing/analysis question was designed to gather data to answer the second research question: Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

Photosynthesis was the topic of study over the course of the project, and the pacing of the researcher's biology class placed it as the first unit related to the topic of matter and energy in organisms and ecosystems. The researcher identified the state standards related to the information in the study, and they were used to build an appropriate assessment for the depth and breadth of content taught. The aforementioned standards were: S-LS1-5: From Molecules to Organisms: Structures and Processes; use a model to illustrate how photosynthesis transforms light energy into stored chemical energy. Emphasis on illustrating inputs and outputs of matter and the transfer and transformation of energy in photosynthesis by plants and other photosynthesizing organisms. Examples of models could include diagrams, chemical equations, and conceptual models. HS-LS2-4: Ecosystems: Interactions, Energy, and Dynamics; use mathematical representations to support claims for the cycling of matter and flow of energy among organisms in an ecosystem. Emphasis on using a mathematical model of stored energy in biomass to describe the transfer of energy from one trophic level to another, and that matter and energy are conserved as matter cycles and energy flows through ecosystems. Emphasis is on atoms and molecules such as carbon, oxygen, hydrogen, and nitrogen being conserved as they move through an ecosystem. HS-LS2-5: Ecosystems: Interactions, Energy, and Dynamics; develop a model to illustrate the role of photosynthesis and cellular respiration in the cycling of carbon among the biosphere, atmosphere, hydrosphere, and geosphere. Examples of models could include simulations and mathematical models (https://community.ksde.org/LinkClick.aspx?fileticket=r_hKuRCOzkY%3d&tabid=5785&mid=14106). (Kansas State Science Standards use the following abbreviations for identification: HS

= high school, LS= life science followed by a number which identifies the topic being studied. The final number in the code refers to the specific skill to be mastered within the topic.)

Multiple-Choice Test Development

The purpose of the multiple-choice test was to gather data to measure students' long-term memory over a unit covering the topic of photosynthetic. The research question was: Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course? The researcher drew upon a learning framework that was developed by Barnett & Ceci (2002) to identify questions that would appropriately measure student learning.

Barnett & Ceci (2002) developed a taxonomy describing characteristics of transfer that are broken down into two main factors: the content and the context.

Content refers to what is transferred and is further divided into three dimensions:

-the specificity-generality of the learned skill: is the skill a fact or learned procedure, or is it a higher problem-solving heuristic or principle

-the nature of the performance assessed: the measure against which improvement is expected

-the memory demands of the task: does the transfer require execution of a learned activity, or does the student need to select the appropriate approach

Context refers to when and where learning is transferred from and to, and is divided into six dimensions:

-knowledge domain: refers to the knowledge base to which the skill is to be applied, for example, Physics to chemistry

-physical context: macroenvironments refer to the location of the learning and transfer (school), while the microenvironment refers to the same room, researcher, etc..

-temporal context: the time between the training and testing phases

-functional context: the nature of the learned skill as purely academic, or is it positioned as something that has purpose in the "real world"

-social context: is the task learned alone or collaboratively with others

-modality: macrolevel is the form of the task (for example, visual or auditory), and microlevel the form of the task, such as multiple-choice, essay, etc..

In this study, near transfer refers to transfer of knowledge to a similar context, specifically, the students' ability to recall the information on photosynthesis at the end of the three-week unit. The content and contextual factors are not either/or, but rather are found on a continuum in the Barnett & Ceci (2002) taxonomy. Therefore, the researcher considered each dimension separately as she sought questions to assess the near transfer of information for her students.

The multiple-choice questions more closely aligned with near transfer when considering the content of transfer. The multiple-choice questions varied in their complexity. For example, some questions asked about specific photosynthetic facts (What are the products of photosynthesis?), while others asked students to select a broader approach from the choices given within the question (The reactions of the Calvin Cycle are not directly dependent on light, but they usually do not occur at night. Why?).   Five of the 15 multiple-choice questions tested information that was not explicitly taught by the researcher. Students had to apply what they had been directly taught in order to arrive at the correct answer. The majority of the questions (10/15) tested information that was explicitly taught during instruction.

Four of the six contextual dimensions for the multiple-choice test fell into the category of near transfer: the topic on all tests was photosynthesis (knowledge domain), the study took place in the same room (physical context), was always academic (functional context), and focused on individual versus group learning (social context). The final two dimensions moved one rank toward far transfer because of the three-week delay from the beginning of the unit and the final test (temporal context), and the test structures between the quizzes and the final assessment were slightly different (modality).

The final multiple-choice test contained 15 questions (Appendix D). Questions 3, 4, 5, 7, 12, and 13 were used to measure NGSS HS-LS 1-5; questions 9 and 11 measured NGSS HS-LS 2-4; and questions 1, 2, 6, 8, 10, 14, and 15 measured NGSS HS 2-5.

Far Transfer as Measured by the Graphing Question

In this study, far transfer refers to information transfer to a dissimilar context (Barnett & Ceci, 2002). The research question was: Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem? The taxonomy that was used to select multiple-choice questions provided the guidance in selecting the graphing questions as well. The graphing/free response question's content was more generalized in that it was a problem-solving question in which students selected the most appropriate approach to graphing a data set. Students were asked to analyze the slopes of the graph and then use their analysis to answer application questions (For example, one question asked the following: Since velvetleaf is a weed invasive to cornfields, predict how increased CO2 concentration may affect interactions between the two species). The structure of this question more generally fit the parameters of far transfer when considering content.

Four of the six dimensions within the context factor more accurately measured far transfer. Altering the topic from the specificity of the photosynthetic process to the more general topic of botany increases transfer within the knowledge domain. The temporal context includes the three-week delay in the final exam, while the functional context changes from purely academic to an actual real-world problem. The test format changes significantly, altering the modality. Therefore, the graphing/free response question measured students' ability to apply their knowledge of photosynthesis to solve a novel, real-world problem, and more generally measured far transfer of student learning.

The final graphing/analysis question can be found in Appendix D.

Validity of the Multiple-Choice Test

Ensuring that a test used in a study measured what the researcher intended for it to measure is important. The inferences must be accurate when making them based on test scores. The researcher sought the aid of three biology instructors as a means of measuring the validity of the multiple-choice test that would be used to infer memory retention in the study.

The final multiple-choice exam included 15 questions, which were given to three biology instructors: two from the building in which the researcher worked, and one from outside the district in which the study was conducted. One instructor had 18.5 years teaching experience, one had three and a half, and one had five and a half. The questions were randomized and given to each instructor along with a list of the three state standards which were the focus of the study. The instructors were asked to analyze each question, and then align the question with the state standard it was testing. For inclusion, two of the three instructors had to provide agreement on the standard being tested; eight of the questions had agreement between all three instructors, while seven of the questions had two of three in agreement.

Developing the Graphing/Analysis Test

A graphing/analysis question was developed from a question found in an Advanced Placement Biology text (Urry, Cain, Wasserman, & Minorsky, 2016) and was used to collect data to answer the second research question. It was specifically chosen to test students' ability to graph and analyze data to answer questions related to a real-world problem. The graphing question was not given to the biology instructors for validation, as this particular question came from a text which explicitly stated this as a question which would test students' ability to demonstrate the specific scientific skill analyzed in the second research question. The scientific skills exercises from which this question was modified, are related to an experiment used to test information found within the chapter content, in this case the chapter on photosynthesis. Those skills include data analysis, graphing, experimental design, and math (Urry, Cain, Wasserman, & Minorsky, 2017), aligning with the purpose of the study's second research question to determine the ability of students to analyze data and transfer content knowledge to a real-world problem.

Procedure

The researcher introduced the study to the students enrolled in her biology classes in November, 2019. The general process for the study was explained, and the consent form was given to each student, and then read out-loud with time and opportunity for students to ask questions. The researcher emphasized the optionality of participation, and explained that data would be disaggregated so no identifying information for individual scores would be used in the final analysis. Students were asked to take the consent form home and explain the project to their parents or guardians. They were encouraged to contact the researcher with any questions or concerns, and contact information was provided for the researcher, the Principal Investigator, and the Chair of the Institutional Review Board.

The researcher's school followed an alternating block-schedule, meaning that students attended four blocks on "blue" days, and four different blocks on "white" days. The researcher had one block every day in which to plan, teaching a total of six blocks, five of which were biology. An on-line random number generator determined which blocks were assigned to the treatment group, and which blocks were assigned to the control group (https://www.google.com/search?q=random+number+generator&rlz=1C1CHZL_enUS712US713&oq=rand&aqs=chrome.0.69i59j69i57j0l6.2416j0j7&sourceid=chrome&ie=UTF-8). The numbers entered in the number generator were one and five to represent the five blocks of students included in the study. The first two numbers that were generated determined the control group, and the remaining three blocks were assigned to the treatment group.

A pre-test was given to each student in the block following the introduction of the study (11/20/19 and 11/21/19). The administration of the pre-test included a 15-question multiple-choice test delivered using Google Forms. The students were then given a graphing/analysis question using a paper copy because this allowed students to graph and write their answers to the analysis questions. The researcher did not review either portion of this quiz or provide answers to any questions upon its completion in an effort to decrease the likelihood of students memorizing either test questions or test answers.

The general pacing and instruction for the photosynthesis unit occurred over the course of three weeks (November 20 – December 13, 2019). The researcher used the same lesson plans and the same pacing for each block of students. A quiz containing questions specific to instructional material presented during the first week of the unit was given to the treatment group (12/6/19, 12/9/19) and was timed at 20 minutes. The researcher displayed a stopwatch on the board, allowing students to pace themselves within that time frame. The researcher reviewed the questions on the quiz to identify the correct answers, and provided feedback to student questions once the students finished the quiz. The control group was provided the same amount of time to work on a paper copy of the study guide over the same material (12/6/19, 12/9/19) . Again, the researcher projected a stopwatch on the board to ensure the provision for the same amount of time of interaction with the content material for both the treatment and control groups. The purpose of timing the quiz and study guide was to decrease the likelihood for any significant learning gain differences between the treatment and control groups to be the result of the treatment itself, and not merely because quizzing increased overlearning of the tested information (Slamecka & Katsaiti, 1988).

The same procedure was followed after the second week of instruction (12/10/19, 12/11/19), with the only difference being the quiz questions administered to the treatment group reflecting the material covered in that week. The quizzes were assigned ten points each, and the scores were entered into the gradebook for students in the treatment groups. Evidence from Adesope, Trevisan, & Sundararajan (2017) provided reasoning of assigning few points to the quizzes, stating that low-stakes quizzing can be used as a learning tool. The information from the quizzes provided an opportunity for students to study, test, and then use information from the quizzes to improve their own learning. Again, the researcher reviewed the questions for the second quiz and provided feedback as necessary. The control group was provided the same amount of time to work on a paper copy of the study guide over the same material. The researcher also answered and clarified the control group’s questions as they were working on the review sheet.

A post-test was given to all students (12/12/19, 12/13/19) upon completion of the instructional unit. The post-test scores were entered into the gradebook because they were used as a partial indicator of mastery of content.

Data Analysis

The purpose of this study was to investigate whether the use of retrieval practice in the form of short quizzes increased long-term retention of biological content knowledge, and the ability to apply that knowledge to solve a novel, real-world problem. The specific study questions were:

1) Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course?

2) Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

Participants were given a pre-test prior to introducing content material for the unit of study. The treatment group was provided two quizzes, one each week, over the course of the unit, while the control group was provided a study guide on which to work for the same time frame. All students were given a post-test at the conclusion of the unit.

The following procedures were followed as a way to analyze the resulting test data for both research questions: 1) A Shapiro-Wilk test was used to compare the scores in each sample to a normally distributed set of scores with the same mean and standard deviation. Tests of normality directed the choice for the appropriate statistical test used when analyzing pre-test and post-test data. 2) Pre-test differences between the treatment and control groups were calculated. If both the pre-tests were normally distributed, then an independent samples t-test was used for the analysis. An independent samples t-test is used to determine if a difference exists between the means of two independent groups if both groups have a normal distribution. However, a Mann-Whitney test was used if the pre-test data were not normally distributed, because it is a non-parametric test equivalent to a t-test. 3) Differences between the treatment group's and control group's post-test scores were analyzed using Analysis of Covariance (ANCOVA). ANCOVA is an appropriate test because it allowed the researcher to statistically control for the influence of the pre-test differences on the post-test scores. The pre-test differences were important to consider because the group in which students were placed was not randomized. It is possible for statistically significant differences to exist in the two groups because of nonrandomization, thus skewing the interpretation of the results. Random assignment greatly strengthens an experimental design, in that it controls for unknown variables (Johnson & Christensen, 2014). However, the school environment in which this study took place did not allow for random assignment of participants into groups, due to the enrollment process used for student assignment into classes. Using a statistical test like ANCOVA, which allows for the control of pre-test differences, strengthened the quasi-experimental design used in this study. 4) Pre-test to post-test differences were statistically analyzed within the treatment group and within the control group. A paired t-test was used if the data sets had a normal distribution, while Wilcoxon Signed Rank Test was used for data sets that did not have a normal distribution. Wilcoxon Signed Rank Test was an appropriate test to use because it is nonparametric test that looks for differences between two independent samples; it can be considered as the nonparametric equivalent to the paired samples t-test. The final analysis allowed the researcher to determine if there were significant learning gains within each of the two groups.

The significance level was set at an alpha level of .05 for all analyses. SPSS (version 26.0.0.0) was used to generate all statistical outputs.

Chapter 4 - Results

One proposed goal of education is developing students' abilities for critical thinking and problem-solving; pupils need a baseline of knowledge from which to draw, and then the ability to apply that knowledge to novel situations in a meaningful way. The purpose of this study was to investigate the use of retrieval practice in the form of low stakes quizzes as a means of increasing high school student retention of biological content knowledge, and an increased ability to transfer the knowledge to solve a novel, real-world problem. Chapter four is organized in terms of the two specific research questions posed in chapter one, and will provide the results from the collection of statistical data analysis for both research questions. The results from the multiple-choice test will be presented, followed by the results from the free response/graphing question. Finally, it will summarize the results and look ahead to the discussion found in chapter five.

Statistical Results

The data gathered from the post-tests was designed to answer the following two research questions:

1) Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course?

2) Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

Two separate tests were developed to answer these questions; a 15-question multiple-choice test provided data to answer the first research question, and a graphing/analysis free response question provided data for the second research question. The researcher administered a pre-test to all students consisting of both the multiple-choice questions and the graphing question. The treatment group was given two low stakes quizzes with feedback over the course of the three-week content delivery, while the control group spent an equal amount of time working on a paper copy of a review sheet covering the same content material. All students were administered a post-test at the end of the instructional unit.

First Research Question

The researcher developed a 15-question multiple- choice test to gather data which would be used to measure high school students' long-term memory recall over the process of photosynthesis. The questions were gathered from various sources, including published text books and state assessments. The completed exam was administered to all students participating in the study as a pre-test. The unit's learning cycle included all of the elements of instruction and took place over the course of approximately three weeks. The exam that was used as a pre-test was given as a post-test upon completion of the unit.

Phi Coefficient Analysis of the Multiple-Choice Exam

The researcher analyzed the post-test data using phi coefficient as a way to measure the strength of association between questions used to test each specific content standard, as a way to measure the reliability of the exam. A student should answer all of the questions measuring the same standard correctly if the student has mastered the standard, or incorrectly if the student has not mastered the standard.

Phi coefficient ranges from -1 to 1, where 0 is no relationship, 1 is a perfect positive relationship, and -1 is a perfect negative relationship. The political science department at Quinnipiac University posted a crude estimate of interpreting the strength of associations based on the Pearson's correlation coefficients, which can also guide the interpretation of the phi coefficient. The estimates are: 0.01 - 0.19: no or negligible relationship, 0.2 - 0.29: weak positive relationship, 0.3 - 0.39: moderately positive relationship, 0.4 - 0.69: strong positive relationship, and 0.7 and higher: very strongly positive relationship (Glen, 2019). Haldun Akoglu (2018) explained that the interpretation of the phi coefficient values differs among diverse scientific research areas, with no absolute rules of interpreting strength. Therefore, researchers should be careful not to overestimate the value of the strength of associations. It is for this reason that the strength of the associations between the questions used to test the three science standards are considered as a rough estimate of the reliability of the multiple-choice test.

The phi coefficients for questions testing the first state standard (NGSS Standard HS-LS-1-5: Use a model to illustrate how photosynthesis transforms light energy into stored chemical energy) revealed a small cluster of questions that had associations between each other of 0.2 or greater. Consistent coefficient values between all four questions suggest that the core set of questions (13, 4, 5, 12) measured a central characteristic within the standard. The remaining two questions (7 and 3) had more erratic associations. For example, questions 7 and 4, and questions 7 and 5 had associations of 0.2 and 0.28 respectively. However, questions 7 and 13, and questions 7 and 12 had very low associations (< 0.2). Question 7 did not have a correlation coefficient of at least 0.2 or greater with all of the questions from the core set, suggesting that it did not belong in the assessment as a way to measure this standard, if the core set of questions did measure the standard. It is possible that question 7 might be measuring content outside the purview of this particular standard. It is also feasible that the question had vocabulary that students did not understand, or misunderstood, and guessed on the answer. If this were the case, the data from question 7 was random and not accurately measuring student knowledge of the content. Question 3 had the same inconsistent data; it had high associations with questions 4 and

12 (0.25 and 0.46 respectively), but lower associations with questions 13 and 5 (0.148 and 0.149 respectively). A general overview of the data suggests that the full set of questions used to measure the first state standard were not related to each other in a way that provided reliable data.

The researcher analyzed the associations between the core set of questions from the first standard (13, 4, 5, 12) to determine which met a threshold of significance to further scrutinize the data. Significance is defined here as having a probability at or below 0.05 of measuring an association as strong (or stronger) as the one observed in a sample from a population in which the true association between these two variables is zero. Every pair of associations between the core set of questions (13, 4, 5, 12) was significant (< 0.05).

The phi coefficient from the two questions measuring the second state standard (NGSS Standard HS-LS-2-4: Use mathematical representations to support claims for the cycling of matter and flow of energy among organisms in an ecosystem) was 0.291, signifying a weakly positive relationship between questions 9 and 11. The data suggests that the two questions used to test the second state standard provide weakly reliable data. The probability of getting an association this strong (or stronger) in a sample like this if the true association between these two variables in the population is zero is 0.001.

The phi coefficient from questions testing the third state standard (NGSS Standard HS-LS-2-5: Develop a model to illustrate the role of photosynthesis and cellular respiration in the cycling of carbon among the biosphere, atmosphere, hydrosphere, and geosphere) revealed little to no correlation. A subset of questions did not emerge as the data were analyzed. The lack of correlation could have resulted from multiple causes. The standard itself encompasses two of the most important and complex chemical processes taught in biology. To find test questions that test student understanding of the role of complex and interrelated processes and relate those processes to large sections of the ecosystem is difficult. It is possible that a student could have had a general understanding of the cycling of carbon, but did not demonstrate that understanding with the small subset of questions on the exam. In addition, it is possible that the questions did not appropriately measure the construct within this standard. The data from the questions testing this standard is likely unreliable. The researcher did not consider the criteria of significance for any of the correlations because a set of questions that collectively measured this standard did not emerge.

The data collected from the phi coefficient analysis suggests that the instrument developed to measure the long-term memory of students was unreliable. Discussion of the impact of using an unreliable instrument in this study can be found in Chapter 5.

Analysis of Data from the Multiple-Choice Exam

Each question that was answered correctly on the multiple-choice exam was given one point. The researcher graphed the raw data as a way to observe its general distribution upon completion of its administration.

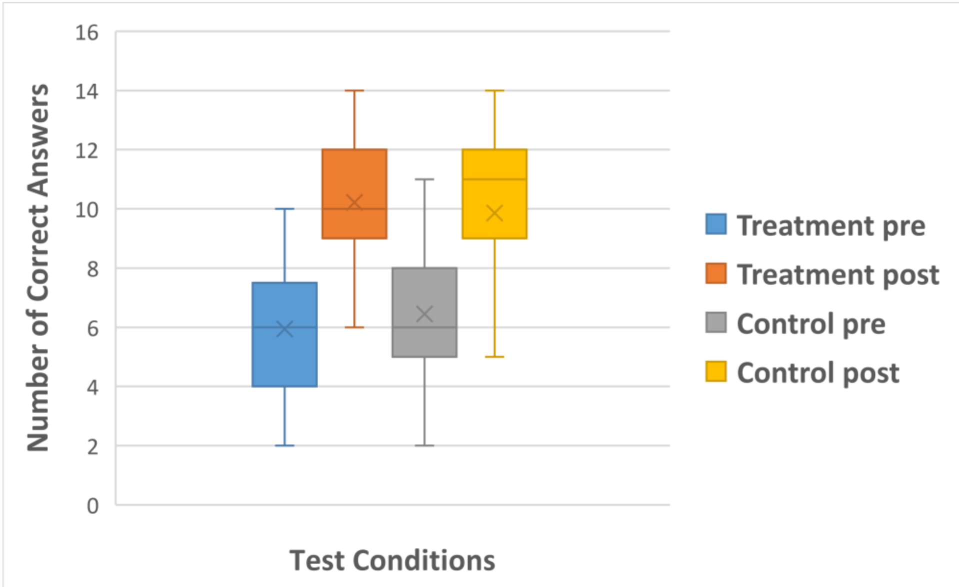


Figure 4.1 ***Multiple-Choice Pre and Post-Test Data***

The researcher sought evidence that the parameter estimates had a normal distribution, providing valuable information as to which statistical tests would appropriately support differences between the groups once the raw data was gathered. Applying statistical tests for data with a normal distribution suggests that the test statistics and p-values will be accurate (Field, 2017). Shapiro-Wilk test is defined as a test used to determine whether a distribution of scores is significantly different from a normal distribution (Field, 2017) and was the test used in this instance. If the test is non-significant ($p > .05$) it indicates that the distribution of the sample is not significantly different from a normal distribution. If, however, the test is significant ($p < .05$) then the distribution is significantly different from a normal distribution. Data analysis concluded that the pre-test scores were normally distributed for both control (0) and treatment (1) groups, as assessed by Shapiro-Wilk's test ($p > .05$). Post-test scores were normally distributed for the treatment group, but not the control group, as assessed by Shapiro-Wilk's test ($p < .05$).

Table 4.1.

Shapiro-Wilk: Multiple-Choice Pre and Post -Test Data in Control (0) and Treatment (1) Groups

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
Pre	0	.197	29	.006	.955	29	.246
	1	.124	33	.200*	.966	33	.379
Post	0	.196	29	.006	.909	29	.016
	1	.115	33	.200*	.959	33	.244

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Determining the normality of distribution for each group allowed the researcher to select the appropriate statistical test to compare the pre-test differences between the two study groups (treatment and control) prior to the application of the treatment. Analysis of pre-test differences that exist in the groups prior to the treatment is critical to determine the potential effect of the treatment. Pre-test data from the treatment and control groups were analyzed using an independent samples t-test. The t-test was appropriate because it is used when there are two experimental conditions and different participants are assigned to each condition. Data provided from the analysis are mean ± standard deviation, unless otherwise stated. There were 29 control and 33 treatment participants; the pre-test scores were higher for the control (m = 6.41, sd = 2.16) than for treatment participants (m = 5.94, sd = 2.0), $t(60) = 0.96$, $p = 0.34$, 95% CI (-1.57, 0.55), $d = 0.245$.

The instructor (the researcher) used the same lesson plans, activities, and pacing over the course of instruction. The treatment group took two low stakes quizzes (timed) with feedback, while the control group studied a paper copy of a review sheet covering the same material; both groups were given a post-test at the conclusion of the unit. The post-test scores were examined using an Analysis of Covariance (ANCOVA). It was important to use ANCOVA because it statistically allowed the researcher to determine whether there were any statistically significant differences between the adjusted population means of two or more independent (unrelated) groups, in this case, the treatment and control groups, by controlling for pre-test scores. The lack of random assignment of students into groups could have led to unintended differences in the two groups prior to the administration of the treatment. For example, scores on the post-test could be skewed if the general academic level of either group tended to be higher or lower than the general academic level of the other. If some of the unexplained variance could be explained in terms of covariates, then the error variance is reduced and more accurately assesses the effect of the independent variable (Field, 2017).

An ANCOVA determined the effect of retrieval quizzes on post-test scores after controlling for pre-intervention differences. After adjustment for pre-intervention differences, there was not a significant difference in post-test scores between treatment and control groups, F (1,59) = 0.699, p = 0.407.

The lack of significant difference between the post-test scores of the treatment and control groups after controlling for pre-test scores led the researcher to question whether either group had significant learning gains. The researcher used statistical analyses as a way to further analyze the data. Considering the 29 participants in the control group, 25 demonstrated an improvement in post-test scores, three saw no improvement, while one did worse. Wilcoxon

Signed Rank Test is defined as a non-parametric test that aids in the analysis of differences between two related samples (Field, 2017) and is therefore an appropriate test to determine whether there were significant differences in pre-test and post-test scores within the control group. A Wilcoxon Signed Rank Test determined there was a statistically significant median increase in post-test scores compared to pre-test scores, $z = -4.383$, $p < .001$.

In a similar analysis, a paired-samples t-test was used to determine whether there was a statistically significant mean difference between the pre-test and post-test scores from participants in the treatment group. Paired-samples t-test was used because it is a test that establishes whether two means collected from the same sample differ significantly (Field, 2017). The data provided are mean ± standard deviation, unless otherwise stated. The assumption of normality was not violated, as assessed by Shapiro-Wilk's test ($p = .244$). Participants scored higher on the post-test ($m = 10.21$, $sd = 2.18$) compared to the pre-test ($m = 5.94$, $sd = 2.0$), a statistically significant increase of -4.27, $t(32) = -8.67$, $p < .001$, 95% CI (-5.28, -3.27), $d = 2.04$.

Research Question Two

The researcher developed a graphing/analysis free response question to gather data used to measure high school students' skill at transferring their content knowledge about photosynthesis to solve a real-world problem. The question was adapted from an Advanced Placement Biology text book (Urry, Cain, Wasserman, & Minorsky, 2016). The completed exam was administered to all students prior to the instructional unit as a pre-test. Students then participated in the unit's learning cycle, including all of the elements of instruction over the course of approximately three weeks. The exam that was used as a pre-test was given as a post-test upon completion of the unit.

Analysis of Data from the Graphing/Analysis Exam

The researcher graphed the raw data as a way to observe its general distribution upon completion of its administration.

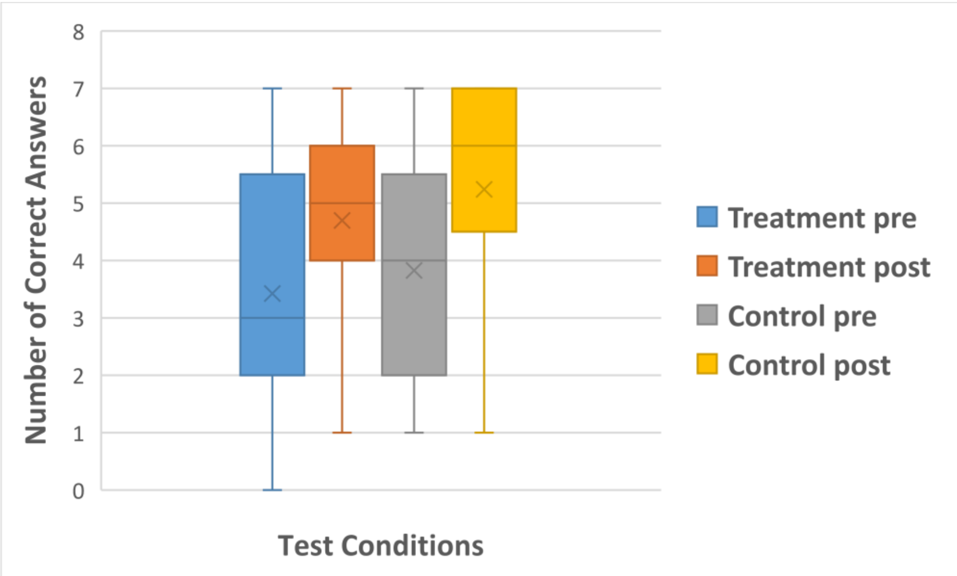


Figure 4.2 ***Graphing Question Pre and Post-Test Data***

The researcher sought evidence that the parameter estimates had a normal distribution once the raw data was gathered; distribution provided valuable information as to which statistical tests would appropriately analyze post-test differences between the groups. The researcher used Shapiro-Wilks test and analyzed the distribution found in the pre-test and post-test scores of both the control and treatment groups. Shapiro-Wilk test is defined as a test used to determine whether a distribution of scores is significantly different from a normal distribution (Field, 2017), and was the test used in this instance. The distribution of the sample is not significantly different from a normal distribution if the test is non-significant ($p > .05$). The distribution is significantly different from a normal distribution if the test is significant ($p < .05$). Pre-test scores were normally distributed for the control (0) group, but not the treatment (1) group, as assessed by Shapiro-Wilks test ($p > .05$). Post-test scores were not normally distributed for control (0) or treatment (1), as assessed by Shapiro-Wilks test ($p < .05$).

Table 4.2.

Shapiro-Wilk: Graphing Pre and Post-Test Data in Control (0) and Treatment (1) Groups

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
Pre	0	.148	29	.102	.933	29	.066
	1	.229	33	.000	.893	33	.003
Post	0	.202	29	.004	.873	29	.002
	1	.183	33	.007	.926	33	.027

a. Lilliefors Significance Correction

Determining the normality of distribution for each group allowed the researcher to select the appropriate statistical test to compare the pre-test differences of the two study groups (treatment and control) prior to the application of the treatment. The Mann-Whitney Test is a rank-based nonparametric test used to determine differences between the two groups, and it was used because there was not a normal distribution in the treatment group's pre-test data set. Data analysis concluded that median pre-test score for control (median = 33.59) and treatment (29.67) was not significantly different, $U = 418$, $z = -.865$, $p = .387$.

The researcher determined whether a statistically significant difference occurred between the post-test scores of the treatment and control group at the end of the learning unit over photosynthesis. The instructor (the researcher) used the same lesson plans, activities, and pacing over the course of instruction. The treatment group took two low stakes quizzes (timed) with feedback, while the control group studied a paper copy of a review sheet covering the same material; both groups were given a post-test at the conclusion of the unit. The post-test scores were examined using an Analysis of Covariance (ANCOVA), which allowed the researcher to control for the influence of the pre-test differences on the post-test scores. If some of the unexplained variance could be elucidated in terms of covariates, then the error variance was reduced and more accurately assessed the effect of the independent variable (Field, 2017). An ANCOVA determined the effect of retrieval quizzes on post-test scores after controlling for pre-intervention differences. There was not a significant difference in post-test scores between treatment and control groups, $F(1,59) = 1.09$, $p = 0.30$.

There was not a significant difference between the post-test scores of the treatment and control groups after controlling for pre-test scores, leading to question if either group had significant learning gains. The researcher used statistical analyses to determine whether there was a significant difference between the pre-test and post-test scores from the participants in the control group as a way to further analyze the data. There were 29 participants in the control group; 21 demonstrated an improvement in post-test scores, three saw no improvement, while five did worse. A Wilcoxon Signed Rank Test determined that there was a statistically significant median increase in post-test scores compared to pre-test scores, $z = -3.357$, $p = .001$.

In a similar analysis, the researcher determined whether there was a statistically significant mean difference between the pre-test and post-test scores from participants in the treatment group. There were 33 participants in the treatment group; 20 demonstrated an improvement in post-test scores, 12 saw no improvement, while one did worse. A Wilcoxon

Signed Rank Test determined that there was a statistically significant median increase in post-test scores compared to pre-test scores, z = -3.679, $p < .001$.

Conclusions

The multiple-choice test was developed to gather data to answer the first research question: Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course? One group was provided two low-stakes quizzes with feedback, and one group was given a review sheet covering the same content information. There were no significant differences between the pre to post-test scores of the multiple-choice test in the two groups of students after data analysis was complete.

The graphing question was developed to gather data to measure the second research question: Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem? There were no significant differences between the pre to post-test scores of the graphing/analysis question in the two groups of students as indicated by the results provided in this chapter.

Statistical analyses of the data collected over the course of the study are included in this chapter. Each statistical test used in the analysis has been identified, along with the justification for its use. A more detailed summary and a discussion of the findings are presented in the next chapter.

Chapter 5 - Summary and Discussion

The purpose of this study was to investigate the use of retrieval in the form of low stakes quizzes on the learning and transfer of biological content knowledge in high school students. The results indicate that there was not a significant difference in the post-test scores of students who took the quizzes and students who were given a review sheet with which to study. While the results do not align with a body of research supporting the use of retrieval as a valuable way to enhance learning, it provided insights as to the direction future research might take as a way to seek improvements in teaching and learning Teaching pedagogy must continue to meet student needs if it is to be effective, and educational research plays a large role in providing evidence as to which strategies will best meet this goal.

Statement of the Problem

One goal for many practicing educators is to teach students important content to be retained in a way that assists pupils in transferring what they have learned to solve novel problems. Therefore, teachers and students could benefit from evidence as to which strategies support this process. In addition, studies demonstrating effective teaching strategies that will improve the opportunity to meet the educational goals of students could inform the pedagogy of teacher preparation programs.

As explained in Chapter 3, this study set out to investigate the efficacy of implementing low-stakes quizzing during the course of a biological unit of instruction as a means to improve long-term retention of information, and also to improve the ability of students to transfer learned material to solve a novel, real-world problem. The specific study questions were:

1) Does the use of retrieval practice in the form of short quizzes given over the series of a unit affect the long-term retention of high school students in a biology course?

2) Does the use of retrieval practice in the form of short quizzes over the series of a unit affect high school students' ability to transfer content information to assist in solving a novel, real-world problem?

The study selected Kansas State Science Standards aligned with the topic of instruction, and those standards were used to guide both the breadth and depth of instruction and the choice of test questions that would be used to assess both long-term memory and students' ability to transfer content to solve a novel problem. The test questions were selected from published texts and on-line state assessments. In addition, a graphing/analysis free response question was modified from an Advanced Biology text (Urry, Cain, Wasserman, & Minorsky, 2016) and was used to measure high school students' ability to transfer their content knowledge to solve a real-world problem. When completed, the final exam consisted of 15 multiple-choice questions and one graphing/analysis free response question.

Methodology Summary

The research project was conducted in a large, 6A high school in eastern Kansas during the 2019 fall semester. The methods used in the study were:

-Recruitment of students from the pool of enrolled students in the researcher's biology course

-Each block of students was randomly assigned to either the treatment group or the control group using an on-line random number generator

-Students were introduced to the purpose of the study and were asked to have their parents sign consent forms to participate in the study

-All participants were given a pre-test

-The researcher taught a unit over photosynthesis. The same notes, lessons and activities were provided to students during the instructional time frame.

-Two low-stakes quizzes were given to the treatment group, while the control group studied a review sheet covering the same information.

-Both groups were given a post-test at the end of the instructional unit

-The scores for both tests were calculated and entered into an Excel spreadsheet for statistical analysis.

Summary of the Results

A multiple-choice test was used to generate scores providing data to analyze whether including short, low-stakes quizzes over the course of a biological instructional unit improved long-term retention of the content (research question 1). In addition, a graphing/analysis free response question provided data to measure the ability of high school students to transfer their biological content to solve a novel, real-world problem (research question 2).

Research Question 1

An analysis of covariance (ANCOVA) provided analysis of the post-test scores because it allowed for the researcher to adjust post-test scores based on pre-test differences. ANCOVA revealed a non-significant difference between the post-test scores of the control and treatment groups ($p = 0.407$).

Research Question 2

The treatment and control groups took a post-test (the same test as the pre-test) at the end of the instructional unit. An analysis of covariance (ANCOVA) provided analysis of the post-test scores because it allowed for adjustment of post-test scores based on pre-test differences. The data reflected a non-significant difference between the post-test scores of the treatment and control groups ($p = 0.3$).

Discussion of the Results

The researcher was interested in investigating the efficacy of low-stakes quizzes used within the course of instruction in a biology class to increase a high school student's ability to retain information in long-term memory and transfer that knowledge to solve a real-world problem. The researcher hypothesized that quizzing would improve student long-term retention of content, as well as the ability to transfer the information to solve a real-world problem (when compared to students who did not take the quizzes). The results of this study stand in contrast to a body of research that suggests quizzing improves long-term memory storage (Dunlosky et al., 2013; Einstein et al., 2012; Karpicke & Roediger, 2008; Roediger et al., 2011; Roediger & Karpicke, 2006). A meta-analysis of practice testing (Adesope, Trevisan, & Sundararajan, 2017) provides evidence that practice tests improve student performance with respect to retention and transfer-based outcomes. Adesope et al.'s study included a total of 118 articles involving 15,427 participants. The study presented results from low-stakes tests only, which aligns with the researcher's methodology for the current research, and provides evidence contradictory to the results. In addition, the results do not align with research indicating that retrieval quizzing improves the ability of students to transfer information (Lee & Ahn, 2018; Cho, Neely, Crocco & Vitrano, 2016; Yang & Shanks, 2018). It was the researcher's goal to contribute to this body of knowledge, as Adescope et al.'s, (2017) meta-analysis found only 11 studies that examined testing effects on transfer outcome, suggesting that more research was needed in this area.

Methodological Considerations for Nonsignificant Results

The statistically nonsignificant gains between the treatment and control group might have resulted from one or more of the methodological limitations of the study. The first possibility is the short time frame of the study. A single 15-question exam given at the end of one learning unit might not adequately capture student learning gains in an accurate and measurable way. Student test results can vary widely on any given day for a myriad of reasons. Environmental influences, such as what is happening at home, the availability of food, sleep patterns, or the stress level of the student can create a swing in a test score that is unrelated to the knowledge of the content, and would provide an inaccurate measure of student mastery. The second possibility is that one unit might not hold the interest for certain students, and students' willingness to actively engage in the retrieval quizzes could vary based on their interest in the particular unit. Generally, students will spend more time engaged in learning and studying topics of interest, while the opposite is true of topics of uninterest. A third possibility is that students might need time to acclimate to the retrieval quiz protocol to reduce potential quiz anxiety. Familiarity tends to provide a degree of stability, thus reducing stress and providing a better opportunity for students to reliably demonstrate their knowledge. Roediger, Agarwal, McDaniel, & McDermott (2011) studied the testing effect in sixth grade students over four chapters of material and found that students improved from roughly a B- average on non-quizzed items to an A- average on quizzed items. The gains in scores might have resulted, at least in part, from the larger span of time used in the study. A project that spans an entire semester, or even a school year could offer thought-provoking information to educators who are interested in providing enduring knowledge for their students.

Each of these identified potential methodological limitations may be mitigated by expanding the study to span a greater length of time and more learning units. For example, expanding the study to include multiple content units would provide a broader view of the learning pattern for each student, as the additional data would be more likely to capture the student's average interest and engagement in the content and be less drastically impacted by a student's potential disinterest (or interest) in one particular topic. Further, providing students an opportunity to become more familiar with the quiz protocol may provide a more stable measure of their learning that is less impacted by discomfort with the particular protocol.

Multiple design flaws might have also influenced the student results. Possible examples include a small sample size, the use of an unreliable testing instrument with too few questions, and the inclusion of an active control group. The flaws would be exasperated if the effect size of the treatment was small to moderate, which is likely. A meta-analysis conducted by Sotola & Crede (2020) found a large array of reported results from retrieval practice studies, varying from a strong positive effect to a weak or negative effect. The purpose of the meta-analysis was to synthesize the results available as a means of determining the effect size of retrieval quizzes. The analysis provided a synthesized effect size from the accumulated studies (including a total of 7864 students), and concluded a moderate association of quizzes with academic performance ($d = .42$). It is less likely for students in one group to score significantly higher than another group when the sample size is small, if the effect of the treatment is small to moderate.

An overarching theme in statistics is the collection of data from a sample and the use of that data to infer things about the population as a whole; the larger the sample, the more likely it is to reflect the whole population (Field, 2017). The researcher did not have 100% participation for her enrolled students because of the factors that were required for inclusion of student scores.

The criteria for the inclusion of scores were: 1) the return of a signed consent form, 2) the presence of students for both the pre-test and post-test (including completion of all portions of the pre-test and post-test), and 3) the presence of students in the treatment group for both of the quizzes. These criteria excluded 20 enrolled students from the treatment group, and seven enrolled students from the control group, decreasing participation by 60% in the treatment group and 24% in the control group. In conclusion, there were 33 participants in the treatment group, and 29 in the control. The students whose scores were not included may or may not have represented a random sample; a non-random sample could have skewed the scores in a way that negated the positive learning gains that actually occurred, considering the relatively low number of participating student scores. The effect of the treatment might have been missed because it was small and difficult to detect with the small number of participants.

Reliability of the instrument is important to ensure that it is consistently measuring test scores (Johnson & Christensen, 2014). An instrument with unmeasured reliability would not dependably demonstrate the learning gains of students. The methodology used in the current study aligned with that used by Roediger, Agarwal, McDaniel, & McDermott (2011), whose research occurred in a sixth-grade social studies classroom, and used multiple-choice questions developed by the teacher to measure the learning gains of the students. Roediger et al. (2011) found that quizzing integrated into the course of a school year improved academic performance, supporting the inclusion of classroom materials as an appropriate data collection tool. However, an instrument that has been tested and shown to be reliable increases the confidence that the inferences made from the results are valid (Johnson & Christensen, 2014). The importance of using a reliable instrument decreases the noise that might be found in nonreliable instruments, improving the results. The instrument used in this study had not been tested for reliability prior to its use. The phi coefficient analysis on the post-test data suggested that the instrument was not reliable and the inferences that can be made from such a test do not carry a great deal of authority.

The development of reliable assessments is a long process that can take years. Questions must be written, analyzed for their appropriateness in measuring the content, run through pilot tests, and statistically analyzed. The researcher was challenged to find a validated assessment that was both content and age appropriate for the participants in the study. The consequence of using an unreliable instrument is the inability to draw strong conclusions from the data produced.

Summative tests must also be a length that will accurately demonstrate student knowledge, in addition to their demonstrated reliability. The researcher's assessment was 15 multiple-choice questions and one graphing/analysis question (with several subsections). Roediger, et al.'s study (2011) found highly variable results on their end-of-the-year exam due to the small number of items on the exam, leading to non-reportable results. Significant results were obtained in an experiment with more test items (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011), suggesting that future research would need to ensure that the end assessment included a larger number of test items to measure each construct, which might provide researchers a more accurate measure of student learning.

Finally, the current study's design used an active control group, providing students with a review sheet on which to work for the same amount of time as that which was provided to the control group when completing the quiz. An active control group could conceivably be considered a separate type of intervention. The treatment group and control group would both have learning gains, and the comparison of the study would more closely represent the comparison of two interventions, instead of the efficacy of the retrieval quizzes. Including an additional group of students who would not be provided with any type of intervention might provide a more accurate analysis of the influence of retrieval practice on student learning.

It is possible that the treatment group did have higher learning gains than the control group, but because of the confluence of the low effect of the treatment, small number of participants, unreliability of the testing instrument, and the use of an active control group, statistically significant differences in scores were not realized. While the criteria for statistical significance were not met with this study, raw test scores demonstrated that learning gains were somewhat greater for the treatment group when compared to the control group. Anecdotally, two students approached the researcher, both in the treatment group, and inquired about the results from the study. At that point the researcher had not completed the statistical analysis of the data, and was unable to speak about conclusions from the study. Instead, she spent some time explaining to the students the two study questions that were being investigated. Both students suggested that the quizzes given over the course of the study's instructional unit aided them in both understanding of the material and recalling the information on the final test. Their comments lead to the consideration of the value of meaningful differences and practical significance as opposed to purely statistical significance. In practice, the p-value should be considered as one of multiple pieces of evidence in statistical decision making (McShane, Gal, Gelman, Robert, & Tackett, 2019). A researcher might consider a more holistic approach to analysis of data from educational studies because of the varying environments in which they are conducted. The broad contextual factors involved in arriving at scientific inferences might be considered in concert with p values, as opposed to treating the p value as a threshold providing the value of the information gained from a study (McShane, et al., 2019). Factors such as the design of the study, the validity and reliability of the testing instrument, and external influences on the population play significant roles in a study's results. Statistically insignificant results do not necessarily mean that there is no value in the information gained from the study.

Additional Considerations for Nonsignificant Results

The time frame of the study fell at the end of the school's first semester. Thanksgiving break was in the middle of the instructional unit, causing an interruption of three school days. Students also had one day off due to a professional development day for teachers. The natural rhythm of a school schedule, and then an interruption of that schedule might have had an impact on student learning. Weiss, Brown, & Weiss (2013) stated that the traditional school calendar is associated with disparate test scores among students with varying backgrounds. The break possibly negatively impacted student learning in a non-recoverable way prior to the post-test. In addition, the timeframe of the study (winter holidays) is when students can experience increased stress due to a looming finals schedule, an upcoming long break in the school routine, and a myriad of interruptions that are an inherent part of the holiday season. The heightened emotions in some students might affect their learning, as emotions often take precedence during cerebral processing and can either impede or assist cognitive learning (Sousa, 2017). While these things are a part of the normal course of a school schedule, they could have had an impact on a student's learning cycle, thus overriding the gains that might have occurred due to the treatment.

The possibility of diffusion (cross-talking) occurring between students in the treatment group and students in the control group is another potential reason that the treatment group did not show statistically significant gains. The school in which the study took place operated on an alternating block schedule, and therefore students who were given the quiz on one day might have talked with students who were in the control group on the opposite day, thus offering some of the same benefits as those who actually took the quiz. Students in the control group would have higher scores because of a benefit similar to the treatment effect if this occurred. The length of the study (approximately three weeks) is such that some cross talking is likely. Anecdotally, the researcher heard from two students (both in the treatment group) who admitted talking with their friends from a different block (in the control group), as to why one group was given quizzes and the other group was not. They did not specifically say they conversed about the content, but it does support the hypothesis that some cross talking occurred.

Lastly, it is possible that statistically significant learning gains between the two groups were not realized because the research was conducted in the researcher's classroom. The researcher sought to provide effective instruction to all of her students, and it is possible that methodologies used over the course of the unit provided enough learning gains for all students that the quizzes did not boost the treatment group's scores enough to be realized. The researcher was a veteran teacher with teaching skills honed over the years. While it is true that experience does not always provide for highly effective pedagogy, it is also true that experience provides opportunities for an instructor to develop multiple strategies that will meet the variety of student learning needs found in a high school classroom. The current study demonstrated significant learning gains within each group (treatment and control), suggesting that all students gained content knowledge. Future research could include students who are not a part of the researcher's assigned classes, and who are under the tutelage of instructors with varying degrees of experience.

Summary

In summary, the researcher's study did not show a statistically significant difference in learning gains in the treatment group of students compared to students who were not given quizzes (control group). These findings are surprising, as the body of research referenced in this dissertation provides strong evidence for learning gains students experience from the testing effect. The researcher, who was also a biology instructor, was interested in investigating teaching and learning strategies which would enhance instruction for high school students. Prior research outlined in Chapter 2 of this document provided evidence that intermittent quizzes given over the course of instruction could improve the learning gains in students that educators often seek.

Implications for Future Research

Future research in an authentic K-12 environment is needed to continue the academic conversations around this topic. Authentic school settings are often difficult to study for a myriad of reasons: a researcher typically cannot use random assignment, student schedules are oftentimes interrupted for reasons beyond the experimenter's control, students or parents sometimes do not sign consent forms limiting the size of the sample, and students have varying levels of interest influencing their efforts in testing or quizzing. However, a K-12 educator lives and works within this very context, therefore, additional studies could offer valuable and practical information for educators and students as they search for the most effective teaching and learning strategies.

Developing a Research Line

Finding ways to conduct educational research in a broader setting can be challenging. Many of the limitations found in this study result from those challenges and would need to be overcome to provide better data to investigate the testing effect in a high school biology classroom. Specific limitations included having a small sample size, using an unreliable assessment, and the short time frame of the study. However, most high school biology curricula have commonalities, and a researcher could capitalize on those commonalities, and weave them into the research design to address some of the challenges.

A large recruitment campaign could be developed to increase the number of the participants in the study, harnessing the power of social media, science list-serves, national professional science associations, state professional associations, and science lobbyists. Johnson & Christensen (2014) state that the larger the sample size of a study the better, because larger samples results in smaller sampling errors. In addition, larger groups of students might allow an effect, if there is one, to be realized.

Increasing the size of the study to include students from various geographical areas, various sizes of schools, and various socio-economic backgrounds could strengthen the value of the research, because the data collected would be less restricted to a specific demographic. Randomly selecting participants for inclusion in the study from broader areas could produce representative samples, strengthening the ability of the researcher to generalize from the sample to the population (Johnson & Christensen, 2014).

The participants in a broader study could be randomized into treatment and control groups once recruited, as a way to strengthen the design. Random assignment is the best technique to build a study in which extraneous variables are equal in both groups prior to the initiation of the independent variable. Random assignment might be considered the gold standard for research design because it equates groups in experimental research, increasing the study's internal validity. High internal validity allows for strong conclusions about cause and effect (Johnson & Christensen, 2014).

A series of assessments could be developed around key biological concepts that are likely taught in most biology courses. The summative assessments would need to be appropriately validated and tested for reliability as a way to provide more accurate information to the researcher(s). Common short, low-stakes quizzes could be given electronically to the treatment group, using testing safeguards such as providing feedback after the testing window has closed for right and wrong answers, placing time limits on the quiz, and locking a quiz after the first trial so that students cannot go back and change their answers. Electronic quizzes can be delivered with very little instruction from an on-site educator, increasing the feasibility of using common assessments throughout the study.

The time frame of the study could be increased from a single instructional unit, to multiple units, or semesters. Increasing the time frame increases the value of a broader study, because teachers would not necessarily need to adjust their pacing. The sequence of the instruction would matter less than the delivery of the quizzes during the appropriate unit. For example, most biology courses cover photosynthesis, but the time of year this unit falls varies between schools. An instructor could provide participating students (in the treatment group) access to an on-line quiz bank that is specific to the content. A teacher could open the quiz for the students once they have completed the unit.

A research line of this magnitude would require the collaboration of multiple researchers, including those with expertise in test development, test administration, data gathering, data analysis, and data dissemination. A similar, large scale design for educational research was conducted by Yeager, Hanselman, Walton, Murray, Crosnoe, Muller, Tipton, Schneider, Hulleman, Hinojosa, Paunesku, Romero, Flint, Roberts, Trott, Iachan, Buontempo, Yang, Carvalho, Hahn, Gopalan, Mhatre, Ferguson, Duckworth, & Dweck (2019) studying the effect of a short, growth mindset intervention. While the scope of the Yeager et al. (2019) study was impressive, a meaningful research project could be developed in a similar way, but on a smaller scale. The idea of expanding the size of the project from a single teacher in a single classroom to a broader population is worth pursuing.

Final Thoughts

Manyika, Chui, Madgavkar, & Lund (2017) prepared an executive briefing for the McKinsey Global Institute stating that educational systems have not kept pace with the changing nature of work. In a McKinsey survey of young people and employers in nine countries, almost 40 percent of employers said lack of skills was the main reason for entry-level job vacancies, while 60 percent said that new graduates were not adequately prepared for the world of work. Informing effective instruction for teachers is more critical than ever. Teaching and learning strategies must keep up with the rapid changes found in our current global society as a way to better prepare students to meet the swiftly evolving world of science and technology. It is important for researchers to continue to investigate the effectiveness of teaching and learning approaches. However, research will be more relevant when it is conducted within the context of authentic classrooms, and the broad band of variability found within those settings. Research done in this context can provide valuable information to enhance teaching and learning.

References

ACT. Using Your ACT Results. https://www.act.org/content/dam/act/unsecured/documents/Using-Your-ACT-Results-19-20.pdf

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The Value of Applied Research: Retrieval Practice Improves Classroom Learning and Recommendations from a Teacher, a Principal, and a Scientist. *Educational Psychology Review*, *24*(3), 437–448. https://doi.org/10.1007/s10648-012-9210-2

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. http://dx.doi.org.er.lib.k-state.edu/10.1037/0033-2909.93.2.203

Albouy, G., Fogel, S., King, B. R., Laventure, S., Benali, H., Karni, A., Carrier, J., Robertson, E. M., & Doyon, J. (2015). Maintaining vs. enhancing motor sequence memories: Respective roles of striatal and hippocampal systems. *NeuroImage*, *108*, 423–434. https://doi.org/10.1016/j.neuroimage.2014.12.049

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. https://doi.org/10.1016/j.jml.2003.08.006

Anderson, M.C., & Bjork, R.A. (1994). Mechanisms of inhibition in long-term memory: A new taxonomy. In D. Dagenbach, & T. Carr (Eds.), Inhibitory processes in attention, memory and Language (pp. 265-326). San Diego: Academic Press.

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences*, *21*(8), 573–576. https://doi.org/10.1016/j.tics.2017.05.001

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin; Washington*, *128*(4), 612–637. http://dx.doi.org.er.lib.k-state.edu/10.1037/0033-2909.128.4.612

Bjork, E. L., Bjork, R. A., & MacLeod, M. D. (2013). *Types and consequences of forgetting: Intended and unintended*. Taylor and Francis.

Bjork, R. A., & Bjork, E. L. (2019). Forgetting as the friend of learning: Implications for teaching and self-regulated learning. *Advances in Physiology Education*, *43*(2), 164–167. https://doi.org/10.1152/advan.00001.2019

BloomsTaxonomy.pdf. (n.d.). Retrieved February 12, 2020, from https://www.d41.org/cms/lib/IL01904672/Centricity/Domain/422/BloomsTaxonomy.pdf

Boise State University. Guidelines for Investigators Using Own Students as Research Subjects. https://www.boisestate.edu/research-compliance/irb/guidance/own-students-research-subjects/

Brown, P., Roediger III, H., & McDaniel, M. (2014). *Make it stick: The science of successful learning*. The Belknap Press of Harvard University Press.

Bower, G. H., Donaldson, W., & Tulving, E. (1972). *Organization of memory*. Academic Press,. http://hdl.handle.net/2027/wu.89033911280

Burke. Johnson. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (Fifth edition.). Thousand Oaks, Calif: Sage Publications.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133. http://dx.doi.org.er.lib.k-state.edu/10.1037/a0019902

Carpenter, S., Rahman, Rahman, S., Lund, T., Armstrong, P., Lamm, M., Reason, R., & Coffman, C. (2017) Students' Use of Optional Online Reviews and Its Relationship to Summative Assessment Outcomes in Introductory Biology. CBE Life Sciences Education, 16(2), Ar23.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*(2), 153–170. https://doi.org/10.1016/j.jml.2009.04.004

Cho, K.W, Neely, J.H.,Crocco, S., & Vitrano, D. (2017) Testing enhances both encoding and retrieval for both tested and untested items, The Quarterly Journal of Experimental Psychology, 70:7, 1211-1235, DOI: 10.1080/17470218.2016.1175485

Conroy, R., & Salmon, K. (2005). Selective postevent review and children's memory for nonreviewed materials. *Journal of Experimental Child Psychology*, *90*(3), 185–207. https://doi.org/10.1016/j.jecp.2004.11.004

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 599–607. https://doi.org/10.1016/S0022-5371(73)80039-8

Crockett, L., Jukes, I., & Churches, A. (2011). *Literacy is not enough: 21st-century fluencies for the digital age*. (Vol. 44). 21st Century Fluency Project.

Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press,. http://hdl.handle.net/2027/uc1.32106006742131

Dong, T., He, J., Wang, S., Wang, L., Cheng, Y., & Zhong, Y. (2016). Inability to activate Rac1-dependent forgetting contributes to behavioral inflexibility in mutants of multiple autism-risk genes. *Proceedings of the National Academy of Sciences*, *113*(27), 7644–7649. https://doi.org/10.1073/pnas.1602152113

Dudai, Y., Karni, A., & Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, *88*(1), 20–32. https://doi.org/10.1016/j.neuron.2015.09.004

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Dunlosky, J., Rawson, K. A., & McDonald, S. L. (2002). Influence of practice test on the accuracy of predicting memory performance for paired associates, sentences, and text material. In T. J. Perfect & B. L. Schwartz (Eds.), Applied metacognition (pp.68_92). Cambridge, UK: Cambridge University Press.

Eisenkraemer, R. E., Jaeger, A., & Stein, L. M. (2013). A Systematic Review of the Testing Effect in Learning. *Paidéia (Ribeirão Preto)*, *23*(56), 397–406. https://doi.org/10.1590/1982-43272356201314

Einstein, G. O., Mullet, H. G., & Harrison, T. L. (2012). The Testing Effect: Illustrating a Fundamental Concept and Changing Study Strategies. *Teaching of Psychology*, *39*(3), 190–193. https://doi.org/10.1177/0098628312450432

Field, A. (2017). *Discovering Statistics Using IBM SPSS Statistics: North American Edition*. SAGE.

File:Brain-layers.gif. (2018, October 17). *Wikimedia Commons, the free media repository*. Retrieved 19:03, May 13, 2020 from https://commons.wikimedia.org/w/index.php?title=File:Brain-layers.gif&oldid=324350238.

File:1511 The Limbic Lobe.jpg. (2020, May 7). *Wikimedia Commons, the free media repository*. Retrieved 18:57, May 13, 2020 from https://commons.wikimedia.org/wiki/File:1511_The_Limbic_Lobe.jpg

Gao, C., Rosburg, T., Hou, M., Li, B., Xiao, X., & Guo, C. (2016). The role of retrieval mode and retrieval orientation in retrieval practice: Insights from comparing recognition memory testing formats and restudying. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(6), 977–990. https://doi.org/10.3758/s13415-016-0446-z

Garcia-Bajos, E., Migueles, M., & Anderson, MichaelC. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory*, *17*(1), 92–103. https://doi.org/10.1080/09658210802572454

Glen, S. (2019, June 04). Phi Coefficient (Mean Square Contingency Coefficient). Retrieved October 10, 2020, from https://www.statisticshowto.com/phi-coefficient-mean-square-contingency-coefficient/

Greenberg, D. L., Keane, M. M., Ryan, L., & Verfaellie, M. (2009). Impaired category fluency in medial temperol lobe amnesia: the role of episodic memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(35), 10900–10908. https://doi.org/10.1523/JNEUROSCI.1202-09.2009

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: Evidence from neuropsychology. *Journal of the International Neuropsychological Society*, *16*(5), 748–753. https://doi.org/10.1017/S1355617710000676

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic Bulletin & Review*, *17*(6), 797–801. https://doi.org/10.3758/PBR.17.6.797

Hostetter, A. B., Penix, E. A., Norman, M. Z., Batsell, W. R., & Carr, T. H. (2019). The role of retrieval practice in memory and analogical problem solving. *Quarterly Journal of Experimental Psychology*, *72*(4), 858–871. https://doi.org/10.1177/1747021818771928

Hutchinson, J., Pak, S., & Turk-Browne, N. (2016). Biased Competition during Long-term Memory Formation. *Journal of Cognitive Neuroscience*, *28*(1), 187–18797. https://doi.org/10.1162/jocn_a_00889

Johnson, B., & Christensen, Larry. (2014). *Educational Research: Quantitative, Qualitative, and Mixed Approaches, 5th Edition* (5th ed., Vol. 1). Sage Publications.

Josselyn, S. A., Kohler, S., & Frankland, P. W. (2015, September 1). *Finding the engram*. Nature Reviews Neuroscience. https://doi.org/10.1038/nrn4000

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Karpicke, J. D., & Aue, W. R. (2015). The Testing Effect Is Alive and Well with Complex Materials. *Educational Psychology Review*, *27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3

Karpicke, J.D., Butler, A.C., & Roediger III, H.L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Karpicke, J.D., & Roediger III, H.L. (2008). The Critical Importance of Retrieval for Learning. *Science, New Series*, *319*(5865), 966–968. http://www.jstor.org/stable/20053384

Kansas State Department of Education. https://www.kneat.org/SNP/SNP_Menus/SNP_Home.htm

Kansas State Department of Education. www.ksde.org/Agency/Fiscal-and-Administrative-Services/Communications-and-Recognition-Programs/Vision-Kansans-Can, retrieved 6/24/19).

Kansas State Department of Education. DCI Arrangements of the Next Generation Science Standards. https://community.ksde.org/LinkClick.aspx?fileticket=r_hKuRCOzkY%3d&tabid=5785&mid=14106

Kansas State Department of Education (04/2013). The Next Generation Science Standards Executive Summary. https://community.ksde.org/LinkClick.aspx?fileticket=wZ4EnOxG1vY%3d&tabid=5785&mid=14106

Kivunja, C. (2014). Do You Want Your Students to Be Job-Ready with 21st Century Skills? Change Pedagogies: A Pedagogical Paradigm Shift from Vygotskyian Social Constructivism to Critical Thinking, Problem-solving and Siemens' Digital Connectivism *| Kivunja | International Journal of Higher Education*. http://www.sciedu.ca/journal/index.php/ijhe/article/view/5156/3176

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology*, *32*(3), 401–410. https://doi.org/10.1002/acp.3410

Larsen, D. (2018). Planning Education for Long-Term Retention: The Cognitive Science and Implementation of Retrieval Practice. *Seminars in Neurology*, *38*(04), 449–456. https://doi.org/10.1055/s-0038-1666983

Larsen, D.P., Butler, A. C., & Roediger III, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education*, *47*(7), 674–682. https://doi.org/10.1111/medu.12141

Leahy, W., Hanham, J., & Sweller, J. (2015). High Element Interactivity Information During Problem Solving may Lead to Failure to Obtain the Testing Effect. *Educational Psychology Review*, *27*(2), 291–304. https://doi.org/10.1007/s10648-015-9296-4

Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology*, *110*(2), 203–217. http://dx.doi.org.er.lib.k-state.edu/10.1037/edu0000211

Liu, Y., Rosburg, T., Gao, C., Weber, C., & Guo, C. (2017). Differentiation of subsequent memory effects between retrieval practice and elaborative study. *Biological Psychology*, *127*, 134–147. https://doi.org/10.1016/j.biopsycho.2017.05.010

Lombardi, M. M. (2007). *Authentic Learning for the 21st Century: An Overview*. 13.

Manyika, J. Chui, M., Madgavkar, A., & Lund, S. (2017). Technology, jobs, and the future of work. *McKinsey Global Institute*. (n.d.). *Technology, jobs, and the future of work*. 13.

McClelland, J. L., McNaughton, B. L., O'Reilly, R. C., (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. http://dx.doi.org.er.lib.k-state.edu/10.1037/0033-295X.102.3.419

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in Middle-School Science: Successful Transfer Performance on Classroom Exams. *Applied Cognitive Psychology*, *27*(3), 360–372. https://doi.org/10.1002/acp.2914

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*(1), 3–21. http://dx.doi.org.er.lib.k-state.edu/10.1037/xap0000004

McShane, B., Gal, D., Gelman, A., Robert, C., & Tackett, J. (2019). Abandon Statistical Significance, *The American Statistician*, 73:sup1, 235-245, DOI: 10.1080/00031305.2018.1527253

Mellanby, J., & Theobald, K. (2014). *Education and Learning: An Evidence-Based Approach*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/ksu/detail.action?docID=1666484

Meyer, A.N.D. & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*(1), 142–147. http://dx.doi.org.er.lib.k-state.edu/10.1037/a0030890

Migueles, M., & García-Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, *21*(9), 1157–1172. https://doi.org/10.1002/acp.1323

Mourshed, M., Farrell, D., & Barton, D. (n.d.). Mckinseyonsociety.com/education-to-employment *NAEP 2008 Trends in Academic Progress*. (n.d.-a). 56.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686. http://dx.doi.org.er.lib.k-state.edu/10.1037/0278-7393.14.4.676

Nguyen, K., & McDaniel, M. A. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, *22*(1), 59–71. http://dx.doi.org.er.lib.k-state.edu/10.1037/xap0000066

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology-Learning Memory and Cognition*, *31*(1), 3–8. https://escholarship.org/uc/item/7fg8h6zq#main

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00286

Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the Novelty Effect: When Familiarity, Not Novelty, Enhances Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition; Washington*, *36*(5), 1321. http://search.proquest.com/docview/753815331?rfr_id=info%3Axri%2Fsid%3Aprimo

Postman L. (1971). Transfer, interference and forgetting. In Woodworth & Schlosberg's Experimental Psychology, Volume H: Learning, Motivation, and Memory, ed. JW Kling, LA Riggs, pp. 1019-32. New York: Holt, Reinhart & Winston. 3rd ed.

Postle, B. R. (2016). *Chapter 21—The Hippocampus, Memory, and Consciousness*. Elsevier Ltd.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*(4), 437–447. https://doi.org/10.1016/j.jml.2009.01.004

Radvansky, G. A., & Copeland, D. E. (2006). Memory retrieval and interference: Working memory issues. *Journal of Memory and Language*, *55*(1), 33–46. https://doi.org/10.1016/j.jml.2006.02.001

Random number generator. https://www.google.com/search?q=random+number+generator&rlz=1C1CHZL_enUS712US713&oq=rand&aqs=chrome.0.69i59j69i57j0l6.2416j0j7&sourceid=chrome&ie=UTF-8).

Richards, B. A., & Frankland, P. W. (2017). The Persistence and Transience of Memory. *Neuron*, *94*(6), 1071–1084. https://doi.org/10.1016/j.neuron.2017.04.037

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382–395. http://dx.doi.org.er.lib.k-state.edu/10.1037/a0026252

Roediger, H. L., & Karpicke, J. D. (2006a). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Karpicke, J. D. (2006b). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 233–239. http://dx.doi.org.er.lib.k-state.edu/10.1037/a0017678

Schacter, D. L., & Tulving, E. (1994). *Memory systems 1994*. MIT Press. http://er.lib.ksu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1744

Shadish, W., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 716–727. http://dx.doi.org.er.lib.k-state.edu/10.1037/0278-7393.14.4.716

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, *73*, 99–115. https://doi.org/10.1016/j.jml.2014.03.003

Sotola, Lukas K, & Crede, Marcus. (2020). Regarding Class Quizzes: A Meta-analytic Synthesis of Studies on the Relationship Between Frequent Low-Stakes Testing and Class Performance. Educational Psychology Review, Educational psychology review, 2020-08-14.

Sousa, D.A. (2017). *How the brain learns* (5th edition.). Corwin Press.

Squire, L. R., & Dede, A. J. O. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*(3). https://doi.org/10.1101/cshperspect.a021667

Squire, L. R., & Zola-Morgan, S. (1988). Memory: Brain systems and behavior. *Trends in Neurosciences*, *11*(4), 170–175. https://doi.org/10.1016/0166-2236(88)90144-0

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380-. Gale Academic OneFile. http://link.gale.com/apps/doc/A11335027/AONE?u=ksu&sid=zotero&xid=5a5d6d71

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(16), 6313–6317. JSTOR. https://www.jstor.org/stable/42590416

Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times* (1st edition..). Jossey-Bass.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology; Palo Alto*, *53*, 1–25. http://search.proquest.com/docview/205797903/abstract/AEE51D4F340A41ACPQ/1

Unsworth, N. (2016). Working Memory Capacity and Recall from Long-Term Memory: Examining the Influences of Encoding Strategies, Study Time Allocation, Search Efficiency, and Monitoring Abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 50–61. http://dx.doi.org.er.lib.k-state.edu/10.1037/xlm0000148

Urry, L. A., Cain, M. L., Wasserman, S. A., & Minorsky, P. V. (2016). *Campbell biology in focus*. Hoboken, NJ: Pearson.

van Gog, T., & Sweller, J. (2015). Not New, but Nearly Forgotten: The Testing Effect Decreases or even Disappears as the Complexity of Learning Materials Increases. *Educational Psychology Review*, *27*(2), 247–264. https://doi.org/10.1007/s10648-015-9310-x

Walck-Shannon, Elise M, Cahill, Michael J, McDaniel, Mark A, & Frey, Regina F. (2019). Participation in Voluntary Re-quizzing Is Predictive of Increased Performance on Cumulative Assessments in Introductory Biology. CBE Life Sciences Education, 18(2), Ar15.

Weiss, J., Brown, R. S., Jr., & Weiss, J. (2013). *Telling Tales over Time: Calendars, Clocks, and School Effectiveness*. Sense Publishers. http://ebookcentral.proquest.com/lib/ksu/detail.action?docID=3034847

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571. https://doi.org/10.1080/09658210244000414

Wimber, M., Rutschmann, R. M., Greenlee, M. W., & Bäuml, K.-H. (2008). Retrieval from Episodic Memory: Neural Mechanisms of Interference Resolution. *Journal of Cognitive Neuroscience*, *21*(3), 538–549. https://doi.org/10.1162/jocn.2009.21043

Wiseman, A. W. (2012). The Impact of Student Poverty on Science Teaching and Learning: A Cross-National Comparison of the South African Case. *American Behavioral Scientist*, *56*(7), 941–960. https://doi.org/10.1177/0002764211408861

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied*, *23*(3), 263–277. http://dx.doi.org.er.lib.k-state.edu/10.1037/xap0000122

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(3), 485–492. http://dx.doi.org.er.lib.k-state.edu/10.1037/xlm0000449

Zhang, D., Zhao, H., Bai, W., & Tian, X. (2016). Functional connectivity among multi-channel EEGs when working memory load reaches the capacity. *Brain Research*, *1631*, 101–112. https://doi.org/10.1016/j.brainres.2015.11.036

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Appendix A - Multiple-Choice Exam Data

Student (T)	Treatment pre	Treatment post	Student (C)	Control pre	Control post
1	7	10	1	11	11
2	10	12	2	5	10
3	6	13	3	9	13
4	4	11	4	5	9
5	7	14	5	8	12
6	5	10	6	8	12
7	4	7	7	4	6
8	4	13	8	10	14
9	6	8	9	5	11
10	8	10	10	4	11
11	4	11	11	8	9
12	5	10	12	6	11
13	5	11	13	10	12
14	6	9	14	8	12
15	5	11	15	6	5
16	2	8	16	6	9
17	5	10	17	6	11
18	7	8	18	5	11
19	4	11	19	6	7
20	4	9	20	8	9
21	9	9	21	7	6
22	3	9	22	6	12
23	8	8	23	4	5
24	7	9	24	4	12
25	3	14	25	6	6
26	6	6	26	7	10
27	6	12	27	2	9
28	8	13	28	4	11
29	8	10	29	9	10
30	8	9			
31	6	6			
32	10	13			
33	6	13			

Appendix B - Graphing Question Data

Student (T)	T PreGraph	T Postgraph	Student (C)	C Pre Graph	C post-graph
1	2	4	1	4	5
2	2	5	2	4	7
3	6	6	3	6	7
4	1	5	4	6	5
5	7	7	5	5	7
6	2	4	6	6	6
7	0	4	7	1	2
8	1	5	8	6	6
9	1	1	9	2	4
10	4	6	10	2	6
11	7	5	11	4	5
12	6	6	12	1	7
13	6	6	13	5	6
14	5	6	14	5	7
15	1	2	15	4	3
16	2	4	16	3	6
17	2	3	17	2	3
18	6	6	18	5	7
19	3	5	19	6	3
20	5	6	20	3	5
21	7	7	21	7	6
22	2	4	22	4	6
23	2	2	23	2	7
24	3	3	24	7	7
25	5	5	25	2	1
26	2	2	26	2	3
27	5	6	27	1	5
28	6	6	28	3	5
29	1	5	29	3	5
30	1	3			
31	3	4			
32	5	7			
33	2	5			

Appendix C - Informed Consent Form

Informed Consent for Participation in Educational Research

Educational research of often undertaken as a means to provide empirical evidence for the most effective teaching strategies. Shannon Ralph will be undertaking such a study.

In this research, students assigned the experimental group will be using a learning strategy known as retrieval practice for the upcoming unit over photosynthesis. Students in the control group will proceed through the unit without using this particular learning strategy.

Each student will take a pre-test and a post-test covering the information learned in the unit. At the conclusion of the study, post-test scores will be compared between the two groups to see if there is a statistical difference between them.

Note: data used in this study will be disassociated from any identifying student information; scores on the pre-test and post-test will be entered into a spreadsheet with assigned numbers, no names or identifying numbers.

While each student will be tested over photosynthesis and the scores will be used in Skyward as a regular biology grade, the participation for the scores to be used in research is completely optional.

By signing the document below, you are providing your approval of your student's scores to be used in this research (again, with no identifying information).

If you have questions or concerns you may contact:

Shannon Ralph, GEHS Biology instructor, PhD candidate: 913-856-2600

Dr. Thomas Vontz, Professor, Kansas State University: 785-532-5927

KSU Research Compliance Office: 785-532-3224

Student: ____________________________________(print name)

Student signature___________________________________ Date: ________

Parent signature____________________________________ Date: ________

Appendix D - Pre-Test and Post-Test Questions

Does Atmospheric CO_2 Concentration Affect the Productivity of Agricultural Crops?

Atmospheric concentration of CO_2 has been rising globally, and scientists wondered whether this would affect plants differently. We will compare corn and velvetleaf, which is a weed found in cornfields.

How the Experiment Was Done: Researchers grew corn and velvetleaf plants under controlled conditions for 45 days, where all plants received the same amounts of water and light. The plants were divided into three groups, and each was exposed to a different concentration of CO_2 in the air: 350, 600, or 1,000 (parts per million).

Data from the Experiment: The table shows the dry mass (in grams) of corn and velvetleaf plants grown in the three concentrations of CO_2. The dry mass values are averages of the leaves, stems, and roots of eight plants.

Table D.1.

Dry Mass Values for Corn and Velvetleaf Plant

	350 ppm CO_2	600 ppm CO_2	1,000 ppm CO_2
Average dry mass of one corn plant (g)	91	89	80
Average dry mass of one velvetleaf plant (g)	35	48	54

Interpret the data

1. a) Place labels for the dependent and independent variables on the appropriate axes.

b) Plot the data points for corn and velvetleaf using different symbols for each set of data.

2. Draw a “best fit” line for each set of data points. A best fit line does not necessarily pass though all or even most points. It is a straight line that passes as close as possible to all data points from that set.

3. Describe the trends shown by the lines.

a) Compare the relationship between increasing concentration of CO2 and the dry mass of corn with that of velvetleaf.

b) Since velvetleaf is a weed invasive to cornfields, predict how increased CO2 concentration may affect interactions between the two species.

4. Based on the data in the scatter plot, do these results support the conclusion from other experiments that corn grow better than velvetleaf under increased CO2 concentration? Why or why not?

Multiple-choice Questions:

1. Based on photosynthesis, one can say that the ultimate source of energy on earth is

a. Sugar

b. The sun

c. Oxygen

d. ATP

2. All organic molecules contain carbon atoms that can be traced back to

a. The bodies of heterotrophs

b. Carbon dioxide (CO2)

c. Water absorbed by plants

d. Carbon from the sun

3. Where does the oxygen gas produced during photosynthesis come from?
a. Splitting of carbon dioxide
b. Splitting of water
c. Splitting of ATP
d. Splitting of glucose

4. Light dependent photosynthetic reactions produce
a. ATP, NADPH, oxygen gas
b. ATP, NADPH, carbon dioxide gas
c. Glucose, ATP, oxygen gas
d. Glucose, ATP, carbon dioxide gas

5. Carbon fixation requires which of the following?
a. Sunlight
b. Products of light dependent reactions
c. High levels of oxygen gas and low levels of carbon dioxide gas
d. Water, ADP, and NADP

6. _____________ are organisms that produce their own organic material using inorganic materials as their energy source, while organisms that use other living organisms as a source of energy are called ___________. .
a. Autotrophs – heterotrophs
b. Heterotrophs – autotrophs
c. Carnivores – Omnivores
d. Herbivores – Carnivores

7. What are the products of photosynthesis?
a. CO2 and H2O
b. CO2 and glucose

c.	O_2 and glucose

d.	O_2 and CO_2

8.	How do living things remove carbon in the atmosphere in the carbon cycle?

a.	Through the process of photosynthesis

b.	Through the process of cellular respiration

c.	Through both photosynthesis and cellular respiration

d.	Neither photosynthesis or cellular respiration remove carbon from the atmosphere

9.	As energy flows through food chains, the amount of available energy

a.	Always decreases

b.	Always increases

c.	Neither increases or decreases – it stays the same

d.	Energy does not move through food chains.

10.	Energy that is lost at each trophic level of an ecosystem is replaced by

a.	Heat

b.	Nutrients from decomposers

c.	Sunlight that is captured by photo synthesizers

d.	Sunlight that is captured by heterotrophs.

11.	About _____ of the energy available at a particular trophic level is incorporated into the tissues at the next trophic level.

a.	1%

b.	10%

c.	50%

d.	100%

12.	When light strikes chlorophyll molecules in the reaction center complex, they lose electrons, which are ultimately replaced by

a.	Splitting water

b. Breaking down ATP

c. Fixing carbon

d. Breaking down glucose

13. The reactions of the Calvin cycle are not directly dependent on light, by they usually do not occur at night. Why? (Explain.)

a. It is often too cold for these reactions to take place.

b. Carbon dioxide concentrations decrease at night.

c. The Calvin cycle depends on the products of the light reactions.

d. Plants usually close their stomata at night.

14. Which of the following groups is absolutely essential to the functioning of an ecosystem?

a. Producers

b. Producers and herbivores

c. Producers, herbivores, and carnivores

d. Carnivores

15. Deforestation is thought to contribute to climate change because:

a. Fewer trees means less photosynthesis, thus decreasing the amount of CO2 being removed from the atmosphere.

b. Fewer trees means less photosynthesis, thus increasing the amount of CO2 being removed from the atmosphere.

c. Fewer trees means fewer consumers, thus decreasing the amount of CO2 they breathe out.

d. Fewer trees does not really have an effect on the CO2 in the atmosphere.

Appendix E - Biology Quiz 1

Biology Quiz- Photosynthesis Name____________________ Block________

Write the equation for photosynthesis:

_________ + _____________ -------------------→ __________ + _____________

Put a circle around the reactants of photosynthesis.

Explain this statement: No process is more important than photosynthesis to the welfare of life on Earth.

Which of the following are produced by reactions that take place in the thylakoids (light dependent reaction) and consumed by reactions in the stroma (Calvin Cycle)?

CO_2 and H_2O

$NADP^+$ and ADP

ATP and NADPH

O_2 and ATP

What is the main job of pigments in the light dependent reaction?

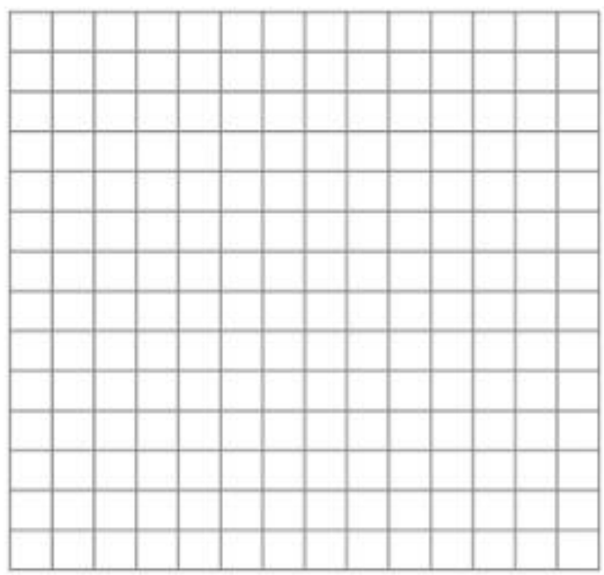
What is the <u>main job</u> of the Calvin Cycle?

Use the following data to create graphs of the effect of light on the rate of photosynthesis.

Table E.1.

Light Intensity vs Rate of Photosynthesis

Light Intensity (angstroms)	Rate of Photosynthesis (O_2 ppt/ min)
0	0
500	58
1000	89
1500	99
2000	106



Based on the graph above, write a complete sentence that summarizes the relationship between light and rate of photosynthesis. (Be specific in describing the SHAPE of your graph.)

Appendix F - Biology Quiz 2

Biology Quiz 2- Carbon cycle and food chains

Name ______________________________ Blk __________

1. In the carbon cycle, through which process does carbon move from an abiotic resource into biotic resources?

a. Deforestation

b. Combustion

c. Respiration

d. Photosynthesis

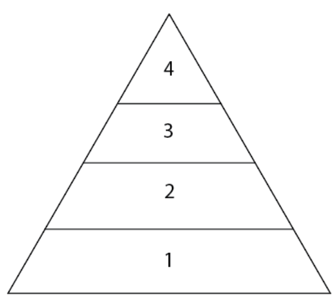
2. In which direction does energy flow through the energy pyramid shown?

a. 4,3,2,1

b. 1,2,3,4

c. 2,1,3,4

d. 3,4,2,1



Consider this food chain: grass → mouse → cat → coyote

3. Suppose 10,000 units of energy are available at the level of the grasses. What is the total number of energy units that is available to the coyote?

a. 1000 units

b. 100 units

c. 10 units

d. 1 unit

4. What is the difference between an autotroph and a heterotroph?

5. Diagram a general sketch of the carbon cycle. Start with CO_2 in the atmosphere and show how it moves through an ecosystem.

Appendix G - Photosynthesis Review Sheet

Biology Review Sheet

"Photosynthesis"

1. Write the equation for photosynthesis.

   (Make sure it is balanced.)

2. What types of organisms can photosynthesize? (There are 3)

3. Diagram a chloroplast. Label the following: thylakoid, stroma, granum

4. What is the function of pigments? When we "see" a pigment as green, which color of light is it absorbing? Reflecting?

5. Draw a diagram of the light dependent reaction of photosynthesis. Be sure to label: reaction centers, photosystems I and II, electron acceptors, electron transport system, where water is split, and where NADP+→ NADPH

- Where are electrons replaced for each PS?
- -Where is this process located?
- -Where is ATP produced?
- -What is the function of accessory pigments?
- -Is sunlight necessary for this to take place?
- -What two products are produced that are necessary for the Calvin Cycle

6. Diagram the Calvin Cycle. Label: RuBP, rubisco, CO2, 6 carbon compound, PGA, PGAL.

- What is the difference between PGA and PGAL?
- Where are ATP and NADPH necessary?
- Why does the 6-carbon compound split?
- Which product is used to build organic compounds?
- What is the name of the enzyme that fixes carbon?

7. Write a food chain with at least four critters that would be found in Kansas.

   Describe what happens to the amount of energy as it moves through the food chain.

   Why are photosynthesizers so important to ecosystems?

8. Diagram the carbon cycle.

9. Review how to graph data. Practice specifically: how to label each axis and how to do an appropriate key. (See attached practice graphs.)

Appendix H - IRB Approval Letter

KANSAS STATE UNIVERSITY | University Research Compliance Office

TO: Dr. Thomas Vontz
Curriculum and Instruction
203 Bluemont Hall

FROM: Rick Scheidt, Chair
Committee on Research Involving Human Subjects

DATE: 12/03/2019

RE: Proposal #9078.1, entitled "Studying the Effects of Retrieval Practice on Long Term Memory of Biological Concepts."

MODIFICATION OF IRB PROTOCOL #9078, ENTITLED, "Studying the Effects of Retrieval Practice on Long Term Memory of Biological Concepts"

EXPIRATION DATE: 01/04/2020

The Committee on Research Involving Human Subjects (IRB) has reviewed and approved the request identified above as a modification of a previously approved protocol. **Please note that the original expiration remains the same.**

All approved IRB protocols are subject to continuing review at least annually, which may include the examination of records connected with the project. Announced in-progress reviews may also be performed during the course of this approval period by a member of the University Research Compliance Office staff. Unanticipated adverse events involving risk to subjects or to others must be reported immediately to the Chair of the IRB, and / or the URCO

It is important that your human subjects activity is consistent with submissions to funding / contract entities. It is your responsibility to initiate notification procedures to any funding / contract entity of any changes in your activity that affects the use of human subjects.

203 Fairchild Hall, Lower Mezzanine, 1601 Vattier St., Manhattan, KS 66506-1103 | 785-532-3224 | fax: 785-532-3278
comply@k-state.edu | k-state.edu/comply

Appendix I - Phi Coefficient Data for Multiple-Choice Post-Test Data

Table I.1.

NGSS Standard HS-LS-1-5: Use a model to illustrate how photosynthesis transforms light energy into stored chemical energy

Standard 1						
Question	3	4	5	7	12	13
3	-					
4	0.255*	-				
5	0.149	0.217*	-			
7	0.314*	0.207*	0.285*	-		
12	0.463*	0.33*	0.214*	0.174*	-	
13	0.148	0.274*	0.373*	0.1	0.239*	-

*$p < .05$

Table I.2.

NGSS Standard HS-LS-2-4: Use mathematical representations to support claims for the cycling of matter and flow of energy among organisms in an ecosystem.

Standard 2		
Question	9	11
9	-	
11	0.291*	-

*$p < .05$

Table I.3.

NGSS Standard HS-LS-2-5: Develop a model to illustrate the role of photosynthesis and cellular respiration in the cycling of carbon among the biosphere, atmosphere, hydrosphere, and geosphere.

Standard 3							
Question	1	2	6	8	10	14	15
1	-						
2	-0.11	-					
6	0.121	0.193*	-				
8	-0.177*	0.029	0.198*	-			
10	0.061	-0.126	-0.056	-0.179	-		
14	-0.12	0.086	0.178*	-0.049	0.04	-	
15	0.119	0.061	0.052	0.143	-0.147	-0.089	-

*$p < .05$