The Impact of Frequent Quizzing on Student Performance

Andrea Negron

Department of Education, Alliance University

EDG 500: Educational Research and Statistics

Dr. Bennet Schepens

April 30, 2023

Abstract

The use of tests and quizzes in the classroom have long been incorporated as tools for assessment. Most teachers would admit to utilizing quizzes in this fashion, but do quizzes serve any other purpose in the classroom? Could they be used as an instructional tool and even improve student performance? This study had seven students in a fourth-grade classroom take part in two different kinds of history instruction: their regular day-to-day instruction as normally presented by their teacher and instruction involving low stakes quizzes taken three times a week. During the experimental instruction period, it was found that students performed better on summative assessments after having taken several formative assessments throughout the course of the lesson. These results have important implications for educators across all grade levels to put into action.

Keywords: assessments, quizzes, testing, instruction, elementary, effect, student performance

The Impact of Frequent Quizzing on Student Performance

There have been several research studies conducted on the use of frequent quizzing, also termed retrieval practice. Despite the research that has been conducted, there is yet to be seen any widespread incorporation of this practice into common instructional practices today. Therefore, it is deemed that further study is necessary in order to support the already existing body of research. The problem that this research project seeks to address is the low academic performance of students on assessments.

The purpose of this study is to identify whether there is a relationship between consistent low stakes quizzing and academic performance. If a relationship is found, this study will also seek to determine the direction of that relationship meaning whether frequent quizzing has a positive or negative influence on student test scores. The independent variable for this study is the frequency with which quizzes are administered. Quizzes will be administered three times weekly. The dependent variable is students' scores on summative assessments. Based on the existing literature, the research hypothesis guiding this study is the conjecture that increasing the frequency of low stakes, formative assessments within the classroom will increase students' academic performance.

Literature Review

The Importance and Effectiveness of Retrieval Practice

Karpicke and Grimaldi (2012) detail the reason why retrieval may be the key to understanding learning and how the retrieval process can influence that learning. The authors state that rather than being only a tool for assessment, retrieval contributes to student learning both directly and indirectly. The indirect effects of retrieval are demonstrated when retrieval practice only enhances learning by some other process. The feedback given after quizzes can

help to guide later studying tactics, making it an indirect effect. The direct effect is as a result of the alteration of knowledge upon retrieval and the enhancement of the ability to reconstruct that knowledge in the future. The main point made by Karpicke and Grimaldi (2012) is that all situations in which knowledge is being expressed or demonstrated require some form of retrieval. No matter the academic task at hand, all students will need to retrieve information that has been restored prior to engagement with the task.

This necessity for retrieval is likely why Sotola and Crede (2021) found that frequent low stakes quizzing is being used as a popular teaching tool. In this review of the literature, Sotola and Crede (2021) defined two terms important to the development of the current study. The first is the testing effect, which is the idea that testing leads to improved memory of the material. The reason for this is because active retrieval has a greater effect on memory as opposed to passive absorption of the material such as studying. The second term is test-potentiated new learning (TPNL), which is the finding that ensuing learning is improved by testing.

The authors also outlined five possible variables which make quizzes more effective in the enhancement of student learning, three of which will be a guide for this study. Quizzes were found to be most effective when administered with high frequency, when they require a generation of an answer rather than the recognition of one, and they are more effective for struggling students. However, despite the benefits found for struggling students, Gordon (2020) warns that those administering the tests should be cautious that participants do not get discouraged by repeated testing. Gordon also warns not to be overly cautious and that testing results in better retention when it is done relatively early in the learning process. Learning is also enhanced when students are given multiple retrieval attempts despite having already given correct answers.

Retrieval Practice on the Primary Level

Jones et al. (2015) conducted three experiments to determine the effectiveness of retrieval practice in comparison to rainbow writing, a commonly used method of spelling instruction. The first two experiments were done with second graders and the third with first graders, each presenting only minor changes to the methods used. All three variations of the experiment produced the same results in which students scored better on summative exams after the use of retrieval practice as the method of instruction. Roediger et al. (2011) examine the effect of testing on middle schoolers in the sixth grade. In experiment one, students were given three tests (pretest, posttest, and review test) and they performed better on chapter exams and the end-of-term exams than the control group. In the second, a rereading control was added to test if simply re-exposure to the material was the cause of the benefit rather than the testing itself. This was not the case and the group that was tested continued to score better. In the third experiment, students were given one pretest and then encouraged to continue quizzing themselves at home using online quizzes. This group also did better than the control group, even on short answer tests.

Retrieval Practice on the College Level

While testing was proven to be effective with some of the youngest students, Gokcora and DePaulo (2018) examined its effectiveness with college students. Despite the recognition of formative assessments as powerful instructional tools, colleges continue to use mainly summative assessments such as exams, midterms, and finals. This experiment found that not only are quizzes a powerful tool for assisting college students, but they also help to increase students' preparedness for class. Marcell (2008) also found this to be true. The author conducted an experiment to study the effectiveness of frequent online quizzing on increasing student preparedness and participation in class. It was found that due to the format of the quizzes,

students needed to read the chapter to find the answers. This resulted in higher levels of preparedness and higher levels of participation, evidenced by more questions asked and deeper discussions during class.

Opposing Research

Another study done with college students argued that perhaps frequent testing is not the best method of instruction. Ward (1984) used three different instructional methods with 301 undergraduate students to determine which produced the best results. The usual method of instruction included regular homework, two exams, and a final. The quiz method included 30-minute weekly quizzes, and a final. In the multiple final method, all course material was taught within five weeks. Upon completion, students were given multiple opportunities to take the final every two weeks until they get a score of 95 or above or until regularly scheduled finals week. The multiple final method proved to be most effective in enhancing the students' learning and mastery of the course material. Davis and Chan (2015) also found that retrieval practice might impair learning. The authors introduced the borrowed time hypothesis in which it was found that as a result of repeated testing, students might try to relearn what they are getting tested on at the expense of the new material they are learning. This information will be used to inform the contents of the current study.

Methods

Setting and Participants

The experiment described in this study took place in a private Christian school, located in the Middle Country Central School District (MCCSD) on Long Island, New York. MCCSD is a large district that covers approximately 16 square miles and has a population of about 10,840 students based on the 2015-16 school year enrollment (*About the district*). Although it is located

within MCCSD, the school does not receive any federal funding. It offers pre-kindergarten through twelfth grade and has an enrollment of just over 100 students. The participants included in this study are a total of seven fourth-grade students, two males and five females. Three students are white, three are black, and one is Hispanic. All students have both parents in the home who are actively involved in their education. One female participant had been previously retained in the third grade and, at eleven years of age, is one year older than her ten-year-old classmates. One boy gets pulled out for special education services twice a week. While the classroom is culturally diverse, all students have English as their first language.

Instrumentation

The instrumentation that will be used to measure students' academic achievement will be a summative assessment on the facts surrounding the specific historical event for that week. The assessment will be made up of only 10 short answer questions and will focus solely on recalling the facts. Each question is worth the same amount, ending with a total score of 10. These tests are created by curriculum developers Veritas Press. Veritas Press is a Christian curriculum that aligns itself with the classical education method, focusing on memorization and repetition as the key to retention.

Procedure

The design for this study will be an action research design. The participants will first be exposed to one kind of treatment and then a second to determine which yields the better exam scores. The first treatment administered will be the control. Students will be taught normally according to the curriculum design. On Tuesday, students will read about the history event as a class and then complete a worksheet containing short response questions, referring back to the reading for help. The students will watch educational videos used within the curriculum to

develop their knowledge of the historical event and then complete a teacher developed, fill-in-the-blank study guide in preparation for the exam. The students will take the exam which is made up of only short response questions the following Monday, seven days later. This will occur for four weeks yielding a total of four exam scores, covering four different historical events.

Participants will then be exposed to the second treatment, the experimental treatment. This treatment will appear very similar to the first, reading about a historical event and completing a worksheet on a Tuesday, watching educational videos, and completing a study guide. However, students will also be quizzed multiple times a week using a fill-in-the-blank method. According to the curriculum, one lesson on a historical event lasts for a week, running from Tuesday to Monday. Therefore, students will be quizzed three times a week on Wednesday, Thursday, and Friday. The quizzes will be low-stakes and students will be informed that their quiz scores will not impact the overall grade for the class. They will also be extremely similar to the study guide which students will be given in order to prepare for the Monday exam. This treatment will continue for another four weeks after the initial control treatment, therefore totaling four more exam scores (a combined total of eight) and covering four more unique historical events.

Statistical Analysis

In this study the null hypothesis (H_0) will be that there is no significant change in the students' test scores between the two treatment periods. The alternative hypothesis (H_a) will be that students' exam scores during the second half of the study, after the four-week experimental treatment period, will be significantly better than their original exam scores during the first four weeks. Using the SPSS statistics software, a paired samples *t* test will be conducted, in order to

determine the statistical significance and reject the null hypothesis. The alpha level will be 0.05 to determine statistical significance.

Results

Four trials of regular instruction and four trials of experimental instruction were conducted incorporating the use of frequent quizzing. At a rate of one trial per week, the entire study lasted eight weeks, yielding a total of eight quiz scores per participant. Quizzes were scored on a 10-point scale. Each quiz was made up of 10 short answer response questions, each worth one point. The table and histograms below demonstrate the average scores and the variations for both types of instruction.

Table 1

Means and Standard Deviations of Quiz Scores for Two Types of Instruction

Instruction Type	Mean	Standard Deviation
Regular Instruction	8.00	1.66
Instruction with Frequent Quizzing	8.46	1.43

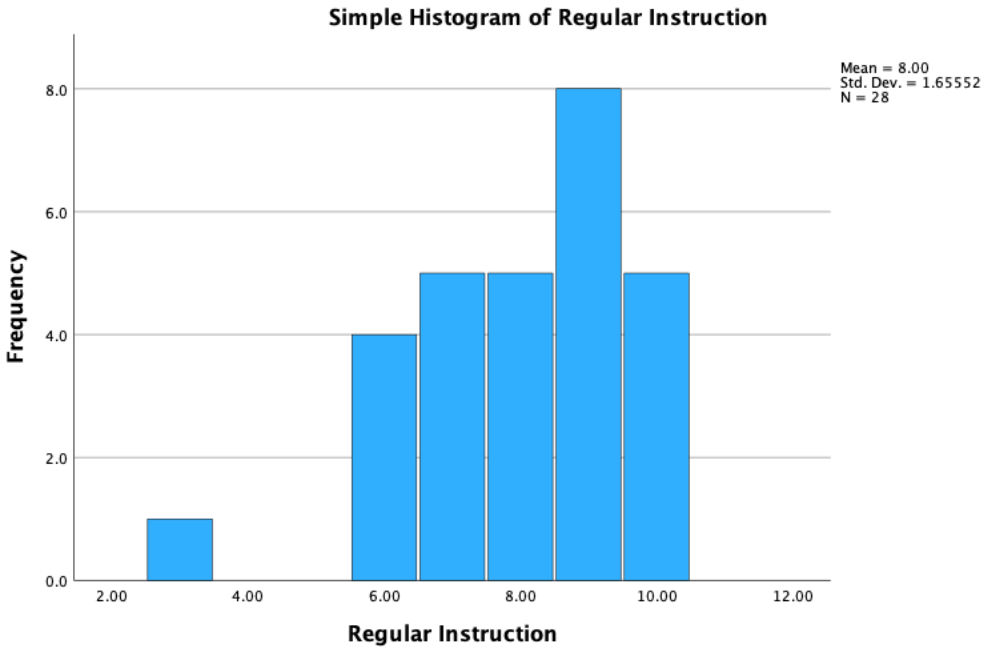

Figure 1. Simple histogram of regular instruction.

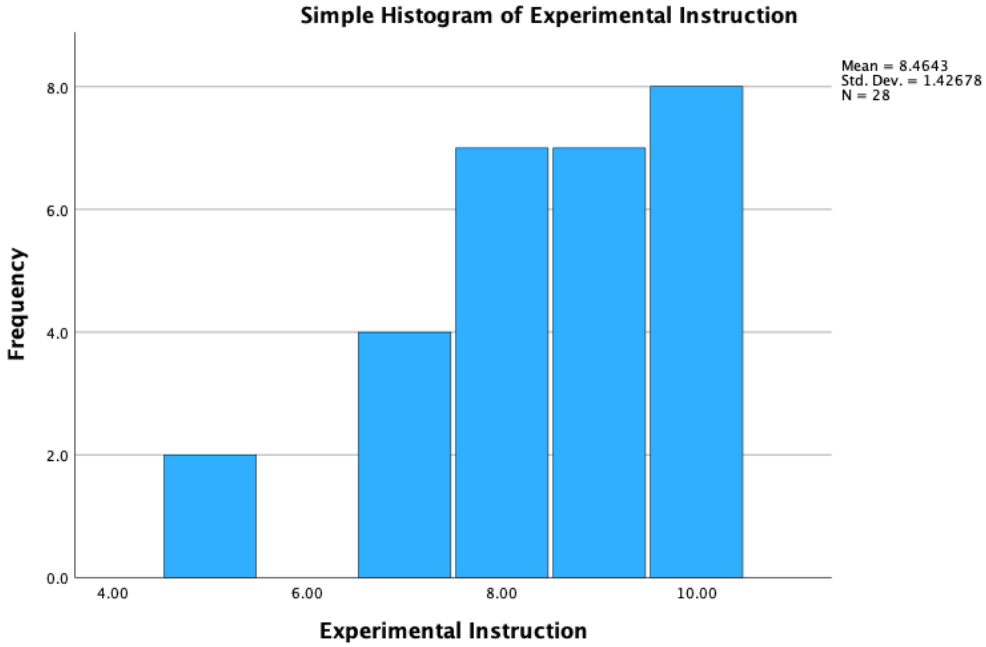

Figure 2. Simple histogram of experimental instruction.

A paired samples *t* test was performed to analyze the data collected in this study and to determine if the null hypothesis could be rejected. The results of this test are presented in the table below. The null hypothesis (H_0) states there is no significant change in the students' test scores between the two treatment periods. According to the data, the mean quiz score increased from 8.00 (*sd* = 1.66) during regular instruction to 8.46 (*sd* = 1.43) with the use of frequent quizzes as seen in Table 1. The difference between the two means is statistically significant at the 0.05 level ($t = -2.56$, $p = .017$). Based on the *t* test, the null hypothesis can be rejected, meaning that there is in fact a difference in students' scores between the two treatment periods.

Table 2

Paired Samples t Test for Regular and Experimental Instruction

Paired Samples Test

		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			One-Sided p	Two-Sided p
Pair 1	Regular Instruction - Experimental Instruction	-.46429	.96156	.18172	-.83714	-.09143	-2.555	27	.008	.017

Discussion

The findings from this study illustrate the relationship between consistent retrieval practice and fourth grade students' academic performance on history tests. The findings demonstrate the positive effect that utilizing frequent low stakes quizzing as a tool for instruction has on summative assessment scores. According to the data, the students performed significantly

better on assessments when provided with instruction that included the use of frequent quizzing in comparison to their regular method of instruction. These results are consistent with the findings from other published studies.

Jones et al. (2015) found in each of their three experiments that primary grade students performed better on exams when consistent quizzing was embedded into the instruction. Variations in the experiments allowed Jones et al. to rule out any other factors that could have been contributing to the students' increased performance levels. Gokcora and DePaulo (2018) concluded that frequent quizzing is beneficial on the college level as well. While this current study seems to support the theories of the testing effect and test-potentiated new learning defined by Sotola and Crede (2021), more research would need to be completed to determine the certainty of this theory.

Limitations

There are some limitations to this study. First, the sample size was very small. Because the study was conducted in a private school, the class sizes are very limited, resulting in a sample size of only seven students. Future research should seek to include more participants in order to allow for more generalizability. Furthermore, the sample included only two boys. In the future, researchers should attempt to use a random sample that more closely resembles the population in regard to culture and gender.

Implications for Practice

This study centered on the use of frequent quizzing in the classroom. The implications of this experiment offer significant changes to instruction that are beneficial to students and educators. First, frequent quizzing improves academic performance. This is beneficial to students, especially in the upper grades, who have difficulty preparing for exams on their own

but want good grades and a good grade point average for eventual college and university admission. This can also help teachers to improve their class scores on high stakes standardized tests. Second, frequent quizzing helps students to better retain information learned. Consistent retrieval practice will over time improve retention of the material being retrieved. Upon retention of the basic foundational knowledge, students can start to interact with the material on a higher level such as discussions, debates, and experiments.

Conclusion

Grade school students often experience many difficulties in regard to taking exams. Just one of these difficulties lies in preparing for the test and achieving the scores they so desire. They often study ineffectively or even fail to study at all. Their consistent poor grades might even become a sense of insecurity, leading to yet another problem, testing anxiety. Despite the evidence shown by previous studies that frequent, formative, low stakes assessments aid in the improvement of scores on summative and high stakes assessments, this practice is not yet a widely accepted method of instruction. This current study adds to the body of literature already in existence illustrating the effectiveness of frequent quizzing. Quizzes are not only a tool for assessment but are also a vital instructional tool to supplement regular instruction.

References

About the district. About: Middle Country Central School District. (n.d.). Retrieved March 23, 2023, from https://mymiddlecountryschools.net/about

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, *85*(2), 89–99. https://doi.org/10.1080/00220671.1991.10702818

Davis, S. D., & Chan, J. C. (2015). Studying on borrowed time: How does testing impair new learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1741–1754. https://doi.org/10.1037/xlm0000126

Gokcora, D., & DePaulo, D. (2018). Frequent Quizzes and Student Improvement of Reading: A Pilot Study in a Community College Setting. *SAGE Open*, *8*(2).

Gordon, K. R. (2020). The advantages of retrieval-based and spaced practice: Implications for word learning in clinical and educational contexts. *Language, Speech, and Hearing Services in Schools*, *51*(4), 955–965. https://doi.org/10.1044/2020_lshss-19i-00001

Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2015). Beyond the rainbow: Retrieval practice leads to better spelling than does rainbow writing. *Educational Psychology Review*, *28*(2), 385–400. https://doi.org/10.1007/s10648-015-9330-6

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24*(3), 401–418. https://doi.org/10.1007/s10648-012-9202-2

Marcell, M. (2008). Effectiveness of Regular Online Quizzing in Increasing Class Participation and Preparation. *International Journal for the Scholarship of Teaching and Learning*, *2*(1).

Roediger, H. L., III, 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.

Sotola, L. K., & Crede, M. (2021). Regarding Class Quizzes: A Meta-Analytic Synthesis of Studies on the Relationship between Frequent Low-Stakes Testing and Class Performance. *Educational Psychology Review*, *33*(2), 407–426.

Ward, E. F. (1984). Statistics mastery: A novel approach. *Teaching of Psychology*, *11*(4), 223–225. https://doi.org/10.1177/009862838401100409