Bulletin of Education and Research
December 2024, Vol. 46, No.3 pp. 43-55

Effect of Frequent Testing on Students' Academic Achievement at Undergraduate Level

Bushra Haleem[*] and Muhammad Saeed[**]

Abstract

The study was proposed to find out the effect of frequent testing on students' academic achievement at undergraduate level. In this study, quasi-experimental pretest post-test control group design was used. Two intact groups, control (28) and experimental (51) were selected that were undergone a Research Methods course at undergraduate level. Two instruments were used for data collection; achievement test and weekly tests. Achievement test was developed as a research tool for both control group and experimental group. The achievement test and weekly tests were developed by the researchers and comprised of multiple-choice questions and descriptive questions. The validity of instrument was ensured by opinions of experts who were teaching the course. In the intervention, the experimental group was assessed by frequent testing and the control group was assessed by traditional testing. Experimental group was tested on weekly basis. Including midterm and final term total of 14 tests were administered on weekly basis only to the experimental group and for the control group, only midterm and final term tests were administered. Both groups were compared on the scores of achievement test. The result revealed the statistically significant difference between the achievement scores of control group and experimental group with Cohen's large effect size (d=1.67) in the post-test mean scores. The results showed that the frequent testing affect the students' academic achievement. Hence it is recommended for teachers and higher education institutions that they may include frequent testing in educational teaching courses for their prospective teachers.

Keywords: Test, Frequent Testing, Academic achievement, Scores, Undergraduate level

[*] Lecturer, Syed Babar Ali Department of Education, Government College University Lahore.
Email: bushra.haleem@gcu.edu.pk
[**] Professor, Department of Education, Minhaj University Lahore, Pakistan.
Email: drsaeed1961@hotmail.com

Introduction

Teachers at all levels of education have access to a variety of instructional strategies and resources, but testing is a common and widely utilized one. Tests are an essential tool of educational instruction and evaluation, because they assess students' knowledge of classroom materials, allowing teachers to understand students' performance and progress in the course. The frequency of testing seems beneficial for the retention of materials and student attitudes in the classroom. When testing occurs frequently, students can receive feedback on a regular basis, which increase their motivation. As the frequency of the test increases, class attendance will also increase because students may be afraid of missing the test (Kling, McCorkle, Miller, Reardon, 2005; Leeming, 2002). Additionally, higher attendance rates help improve student learning (Chen & Lin, 2008). Testing can help students to improve their academic achievement as long as it has valid content and reliable ways to measure this content area. It brings confidence and a positive attitude toward instruction and course.

The researchers provided many possible explanations to explain why frequent testing should be beneficial to learning and teaching. Firstly, the frequent testing provides extrinsic motivation. During study students work hard because they wanted to get good scores in test (Curo, 1963; Dustin, 1971; Khalaf, 1989; Standlee & Popham, 1960). Secondly, frequent testing allows students to give feedback or understand their marks, provide opportunity to know their strengths and weaknesses, and also give them time to work more to eliminate weaknesses (Standlee & Popham, 1960; McDaris, 1984; Bangert Drowns, 1986). According to Standlee and Popham (1960), call the third explanation "forced activity" on the issue. The exam process forces the students to learn information at a deeper level than other means. Selakovich (1962) provided a fourth explanation that frequent testing can improve classroom discussion. Finally, Dustin (1971) explained that the frequent testing can decrease stress, because each test only represents a small part of the overall score.

Alade and Kuku's (2017) stated that when exposed to different test frequencies, the average math scores of the students on the achievement test varied significantly. Comparing the study's results with those of other experimental groups (such as schools which administer tests every three or four weeks), schools that test once every two weeks and then test every week have a greater impact on improving average academic achievement. Control group participants had the lowest average scores. These results are in line with Deck's (2008) study, which indicated a significant difference among students who evaluated weekly and those who evaluated once a month. According to Shirvani (2009) and (De Paola & Scoppa, 2010), as well as other comparable research, students in the treatment group who took the intermediate test scored higher than those in the control group who did not get intervention.

However, the experiments reported by Casem (2006) show positive correlation between the frequent testing and student achievement. In his research, students who took the biology course nine exams per semester performed better than students who took only three exams. Scoppa and De Paola (2011) identified effect related to the increase in testing frequency. He conducted his research on two exams per semester in two courses of introductory economics, in which he compared the two exam to final exam methods. The findings were that the students who took tests frequently scored higher than those students who only took one test.

Frequent tests in all research areas are described in different means. Some researchers suggest testing as once every two weeks (Khalaf & Hanna, 1992), some recommend as weekly (Kika, McLaughlin & Dixon, 1992), other recommend as daily (Dineen, Taylor & Stephens, 1989). The frequent testing is also varying in control group and experimental group (Martin & Srikameswaran, 1974), and also testing per chapter in a month (Grover, Becker, & Davis, 1989), or once in every quarter (Kling, Miller, & Reardon, 2005).

Testing can be an important topic in education. Too many tests are not good for students, which means they will hinder their education. Due to frequent testing, teachers focus excessively on the exam and teach just what is necessary to help students do well on the test. This may lead to a learning environment that is "unsuitable for independent learning" at the school (Marshall, 2007, p. 34). However, the frequency of testing also needs to be balanced. Too few tests are of no use to students, but too frequent tests will have the same effect and are not good for students. It was revealed by Dineen, Taylor, and Stephens (1989) that when students were tested a day by day and then on a weekly basis, there was no significant difference in their findings. This indicates that the daily test is detrimental to students' educational and academic learning.

In the literature frequencies of testing varies from research to research. There is methodological difference among research, in addition to differences in frequency. Tests, without a doubt, motivate students to study while also providing feedback to teachers and administrators. They are indicators of instructional efficacy as well as student achievement levels. For decades, they've been a necessary component of instructional practice. Tests are essential because they examine students' responses to specific questions or problems, regardless of how frequently they are administered. The majority of studies indicated that regular testing improves academic achievement.

While some research focuses on the impact of frequent testing on achievement, others look into the impacts of frequent testing on anxiety, attitude, and retention. Frequent testing was frequently cited as an increasing factor of academic achievement in studies of frequent testing. However, a few studies have concluded that frequent testing

had little or only a minor impact on students' academic achievement. These findings are confusing, and after seventy years of research on the topic, we're still not sure what repeated testing does. Many questions remain unclear about the use of frequent testing. Whether frequent testing is more effective when the tests are graded, whether they are administered to university students rather than school students, or when they are given in a different format. Frequent tests have been studied for years; however, the research has not yielded convincing results. In fact, many researchers have presented contradictory results. Therefore, it is not clear that the expected results of frequent testing are true or not. Therefore, this proposed research is designed to determine the effective test frequency of evaluating students in the teaching process in respect to improve the students' academic achievement in undergraduate education course.

Objective of the Study

1. To investigate the effect of frequent testing on students' academic achievement at undergraduate level.

Hypotheses

H_o.1: There is no significant difference between the achievement scores of control group in the pretest and post-test.

H_o.2: There is no significant difference between the achievement scores of experimental group in the pretest and posttest.

H_o.3: There is no significant difference between the achievement scores of control group and experimental group in the pretest.

H_o.4: There is no significant difference between the achievement scores of control group and experimental group in the post test.

H_o.5: There is no significant difference between mid-term scores of control group and experimental group.

H_o.6: There is no significant difference between final-term scores of control group and experimental group.

Research Methodology

The experimental research was proposed to identify the effect of frequent testing on the achievement scores of students at undergraduate level. A quasi-experimental control group design was used in this study. Two intact groups were selected at Institute of Education and Research, University of the Punjab, Lahore. Two groups, one control (28) and experimental (51) were selected that were undergone a Research Methods course at undergraduate level. The experimental group was assessed by frequent testing and the control group was assessed through the traditional method. The comparison was measured on their achievement tests scores as posttest of both groups.

Pretest –Posttest Control Group Design

Group I	Pretest	Treatment	Posttest
Group II	Pretest	Control	Posttest

Instrumentation

For data collection two instruments i.e. achievement test and weekly tests were used. Achievement and weekly tests were consisted of MCQs and restricted questions. The steps of test construction are given in figure 1:

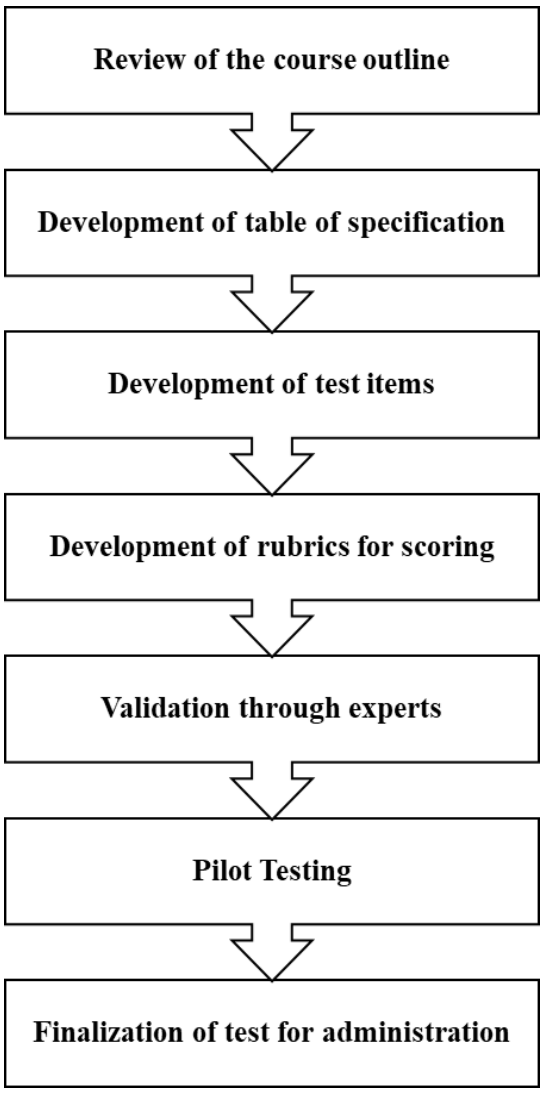


Figure 1. *Steps of Test Construction (Source: Author)*

In achievement test there were 30 MCQs and six restricted questions. For the scoring of restricted questions rubrics were developed. For experimental group total 14 weekly tests were developed. The mid-term and final term tests were developed according to the criteria given by the Institute for control and experimental groups. Weekly tests were comprised of multiple-choice questions and descriptive questions. Tests were developed according to the basics rules of test construction. To maintain the difficulty level, tests were made according to the table of specification. Items in the tests were measured the three levels of cognitive domain presented by Bloom (1956). The validity of the instruments was ensured through the experts' opinion from five relevant teachers having expertise in test development and research. The experts who were teaching this course were selected. The suggestions given by the experts were incorporated in the instruments.

Piloting of Achievement Test and Weekly Tests

Achievement test and weekly tests were administered on the 50 students enrolled in this course at Institute of education and research**,** not where the actual study was conducted. The item characteristics were checked through the pilot testing of the instruments. Items having difficulty index ranging from 0.30 – 0.70 and discrimination index above 0.30-01 were retained in the final test as suggested by Munir, Sabir, Shah and Tipu (2013). However, items having higher value of discrimination index and moderate difficulty were preferred.

Intervention Procedures

The intervention period of the study was a full semester i.e.16 weeks in the spring semester, three hours per week to meet three credits hours course requirement. The course research methods in education was selected for intervention. A pretest post-test control group design was used to investigate the testing effect of the intervention in a classroom setting. Experimental group was taught by researcher herself and control group was taught by another teacher.

The experiment took place over a four-month period. After selecting two intact groups from the science department, the selected groups were randomly assigned to experimental (51) and control (28) groups. There were two classes in a week, the time period of each class was 1:30 hours. Both classes were on the same day but the timings were different for both groups. Before intervention, pretest was taken from both control and experimental groups for comparison of data after intervention. After the administration of pretest, the treatment was given only to experimental group. The treatment group was assessed by frequent testing and the control group was assessed by the traditional method. Tests were administered weekly for the experimental group, but

mid-term and final term tests were conducted for both experimental group and control group. Including midterm and final term, a total 14 tests were taken weekly to the experimental group and only midterm and final term tests were administered for the control group. Both groups were compared on their scores on achievement test as posttest. The intervention period was 16 weeks, three hours per week.

Findings

For the comparison of scores between pretest and post-test paired sample t-test was applied. For the comparison of scores of academic achievement of both control and experimental groups independent samples t-test was applied.

Table 1
Paired Sample t-test for Comparison of Mean Scores of Pretest and Post Test of Control Group in Achievement Test

Scores	*N*	*M*	*SD*	*Df*	*t-value*	*Sig*
Pretest	28	7.89	2.67	27	-11.43	.000
Post test	28	27.14	9.39			

Table 1 shows that paired sample t-test was applied to compare the pretest and posttest mean scores of control group. The result revealed that (t=-11.43, p<.001) is statistically significant difference between the achievement scores of pretest (M = 7.89, SD = 2.67) and posttest (M = 27.14, SD = 9.39) in the mean scores of control group. So the null hypothesis "there is no significant difference between the achievement scores of control group in the pretest and posttest" is rejected. The comparison between pretest scores and post test scores revealed a significant difference. So it is concluded that achievement scores of posttest were higher than pretest scores of control group assessed through traditional method.

Table 2
Paired Sample t-test for Comparison of Mean Scores of Pretest and Post-test of Experimental Group in Achievement Test

Scores	*N*	*M*	*SD*	*Df*	*t-value*	*Sig*
Pretest	28	7.29	2.02	50	-47.47	.000
Posttest	28	39.47	4.48			

Table 2 shows that paired sample t-test was applied to compare the pretest and posttest mean scores of experimental groups. The result revealed that (t=-47.47, p<.001) is statistically significant difference between the achievement scores of pretest (M = 7.29, SD = 2.02) and posttest (M = 39.47, SD = 4.48) in the mean scores of experimental groups. So the null hypothesis "there is no significant difference between the achievement scores of experimental group in the pretest and posttest" is rejected. The

comparison between pretest scores and post test scores revealed a significant difference. So it is concluded that achievement scores of posttest were higher than pretest scores of experimental group assessed through frequent testing.

Table 3
Independent Sample t-test for Comparison of Achievement Scores *of Control and Experimental Group in the Pretest Mean Scores*

Groups	*N*	*M*	*SD*	*Df*	*t-value*	*Sig*
Control	28	7.89	2.67	77	1.03	.30
Experiment	51	7.29	2.02			

Table 3 shows that independent sample t-test was applied to compare the achievement scores of control and experimental group in the pretest mean scores. The result revealed that (t=1.03, p>.30) is statistically not significant difference between the achievement scores of control group (M = 7.89, SD = 2.67) and experimental group (M = 7.29, SD = 2.02) in the pretest. So the null hypothesis "there is no significant difference between the achievement scores of control group and experimental group in the pretest" is accepted. The comparison between control and experimental group did not reveal a significant difference. So it is concluded that achievement scores of control group assessed through traditional method and experimental group assessed through frequent testing were same in pretest.

Table 4
Independent Sample t-test for Comparison of Achievement Scores *of Control and Experimental Group in the Post-Test Mean Scores*

Groups	*N*	*M*	*SD*	*Df*	*t-value*	*Sig*	*Cohen's d*
Control	28	27.14	9.39	77	-6.54	.000	1.67
Experiment	51	39.47	4.48				

Table 4 shows that independent sample t-test was applied to compare the achievement scores of control and experimental group in the post test mean scores. The result revealed that (t=-6.54, p<.001) is statistically significant difference between the achievement scores of control group (M = 27.14, SD = 9.39) and experimental group (M = 39.47, SD = 4.48) with Cohen's large effect size (d=1.67) in the post-test mean scores. So the null hypothesis "there is no significant difference between the achievement scores of control group and experimental group in the post test" is rejected. The comparison between control and experimental group revealed a significant difference. So it is concluded that achievement scores of experimental group assessed through frequent testing were higher in posttest than control group assessed through traditional method.

Table 5
Independent Sample t-test for Comparison of Midterm Mean Scores of Control and Experimental Group

Groups	*N*	*Mean*	*SD*	*Df*	*t-value*	*Sig*	*Cohen's d*
Control	28	25.53	2.13	77	-6.60	.000	1.56
Experiment	51	28.94	2.22				

Table 5 shows that independent sample t-test was applied to compare the midterm scores of control and experimental group. The result revealed that (t=-6.60, p<.000) is statistically significant difference between the achievement scores of control group (M = 25.53, SD = 2.13) and experimental group (M = 28.94, SD = 2.22) with Cohen's large effect size (d=1.56) in the post-test mean scores. So the null hypothesis "there is no significant difference between mid-term scores of control group and experimental group" is rejected. The comparison between control and experimental group revealed a significant difference. So it is concluded that scores in midterm of experimental group that was tested through frequent testing were higher in post-test than control group that was tested through traditional testing.

Table 6
Independent Sample t-test for Comparison of Final Term Mean Scores of Control and Experimental Group

Groups	*N*	*Mean*	*SD*	*Df*	*t-value*	*Sig*
Control	28	31.10	2.72	77	1.35	.17
Experiment	51	30.21	2.83			

Table 6 shows that independent sample t-test was applied to compare the final term scores of control and experimental group. The result revealed that (t=1.35, p>.17) is statistically insignificant difference between the achievement scores of control group (M =31.10, SD = 2.72) and experimental group (M = 30.21, SD = 2.83) in the final term scores. So the null hypothesis "there is no significant difference between final-term scores of control group and experimental group" is accepted. The comparison between control and experimental group did not reveal a significant difference. So it is concluded that scores of final term of control group that was tested through traditional testing and experimental group that was tested through frequent testing were same.

Discussion

The purpose of this research was to measure the effect of frequent testing on students' academic achievement scores at undergraduate level. In this study the frequent testing has a positive effect on students' academic achievement in the subject of research methods in education at undergraduate level. The experiment group was tested through frequent testing on weekly basis. When the experimental and control groups were compared, the experimental group outperformed the control group. The findings of the current study are

similar with those of (Martin & Srikameswaran, 1974), who stated that students who got weekly quizzes outscored those who did not get any quizzes during the course which they were enrolled. Another study by Gholami and Moghaddam (2013) showed that the performance of the weekly quiz group was much better than that of the control group. Students' performance was shown to be improved when they took weekly quizzes. The findings of present research are also consistent with the previous research findings. The research results of Alade and Kuku (2017) show that when exposed to different test frequencies, between students, there is a statistically significant difference in their average math exam scores. The findings reveal that, when compared to the other experimental groups, schools that test every three and four weeks, schools that test once every two weeks, and then schools that test weekly, have a greater impact on improving average academic achievement. Participants in the control group had the lowest average grades.

Another study by Casem (2006) reported a positive correlation between the frequency of tests and the achievement of students in his experiment. In his study, he revealed that students in a biology class who took nine exams per semester performed much better than those who took just three examinations. There were additional experiments conducted in which no overlapping material was used. They found that the final examination in two basic economics courses was much more difficult than a series of two exams given per semester, as reported by De Paola and Scoppa (2011). Those who were tested more often got better marks, according to the results, compared to those who were only tested once.

This result is similar with the findings of Deck (2008), when evaluations were conducted on a weekly or monthly basis, there was a statistically significant difference in student achievement between the two groups. In previous research of a similar kind, it was discovered that students in the treatment group who took the intermediate test scored much better than students in the control group (De Paola & Scoppa, 2011; Shirvani, 2009). These results are also consistent with Zgraggen's (2009) research findings that the students who took final exam every two weeks scored higher on the final exam than students who took the final exam every week.

In contrast to the findings of the current study, Haberyan (2003) revealed that there was no statistically significant difference in students' achievement between the experimental group who took a weekly quiz and the control group who did not take a quiz throughout the study. Mines (2014) studied the relationships between test frequency and final grades in an environmental engineering course. The researcher looked at data from ten courses that took place between 2001 and 2012. Statistics show that testing frequency has little effect on students' final grades. According to the researchers (Geist &

Soehren, 1997) and (Ballard & Johnson, 2004), who conducted a similar study in which they compared weekly quizzes versus no quizzes. There was insignificant relationship between test frequency and student achievement on a range of frequencies. The findings of these studies, on the other hand, have been unclear.

In this study it was found that the scores of experimental group taught through frequent testing were higher in mid-term than control group taught through traditional method. It was also found that the final term scores were same of both experimental group taught through frequent testing and control group taught through traditional method. Standlee (1962) conducted another early study to examine if the use of quizzes was associated with enhanced subject matter learning. Indiana University researchers selected 104 undergraduates from four different sections of an introductory educational psychology course. The researchers found no significant differences across quiz technique groups or sections when using the final examination as a measure of achievement. The research also revealed that by the mid-term test, quizzes increased student success in the first half of the course, but the effect decreased towards the end. Therefore, in many universities classrooms, there are fewer exams, which may be mid-term and final-term, which is very common. Some people believe that when more tests are taken, students will learn less because each test has a lower overall weight in the overall class performance (Mines, 2014). Another meta-analysis showed that the most frequent tests are not linearly related to student achievement (Kuo & Simon, 2009). Kling, McCorkle, Miller, and Reardon (2005) conducted an investigation on the achievement of students on the final test for the marketing class. It is important to note that although both studies examine the impact of weekly and monthly testing, the contents of the two frequency tests are totally different from one another. No study has been able to arrive to the conclusion that the often tested group outperformed the control group in significant way.

Conclusions and Recommendations

The results revealed that after intervention achievement scores of experimental group were increased as compared to control group. Before doing intervention the achievement scores of both groups were same in the pretest. This shows that the treatment improves the achievement of students in posttest. Students who taught with frequent testing got higher scores in posttest as compared to those students who taught with traditional method. The scores of experimental group tested through frequent testing were higher in mid-term than control group taught through traditional method. But the final term scores were same of both groups. The findings of this study also conclude that the achievement scores of experimental group tested through frequent testing are higher in post-test than control group who is not exposed to frequent testing. Hence, it is concluded that frequent testing is helpful for the students to improve the academic achievement.

On the basis of results frequent testing is recommended. Frequent testing has effect on student academic achievement. Moreover, frequent testing can encourage a positive attitude in classroom and about instructor as well. During semester frequent testing may involve the students in the study. In the semester teacher may give a short quiz early that can encourage students to participate more actively in the class. Furthermore, frequent testing can provide helpful feedback on student performance in the classroom to the school, teachers, parents, and students.

References

Alade, O. M., & Igbinosa, V. O. (2014). Table of specification and its relevance in educational development assessment. *European Journal of Educational & Development Psychology, 2*, 1–17.

Bangert-Drowns, R. L., Kulik, C.C., Kulik, J. A., & Morgan, M.T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 85*(2)*,* 89.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives.* David McKay Company Inc.

Casem, M. L. (2006). Active learning is not enough. *Journal of College Science Teaching*, *35*(6), 52.

Chen, J., & Lin, T.F. (2008). Class attendance and exam performance a randomized experiment. *The Journal of Economic Education*, *39*, 213–227.

De Paola, M., & Scoppa, V. (2011). Frequency of examinations and student achievement in a randomized experiment. *Economics of Education Review*, *30*, 1416–1429.

Deck, D. W. (2008). *The effects of frequency of testing on college students in principles of marketing course* (Doctoral dissertation). Virginia Polytechnic Institute and State University, Virginia.

Dempster, F.N. (1991). Synthesis of research on reviews and tests. *Educational Leadership*, *48*, 71–76.

Dochy, F. (2008). The Edumetric Quality of New Modes of Assessment: Some Issues and Prospects. Assessment, Learning and Judgement in Higher Education. Dordrecht: Springer Netherlands

Geist, J. R., & Soehren, S. E. (1997). The effect of frequent quizzes on short-and long-term academic performance. *Journal of Dental Education*, *61*(4), 339-345.

Gholami, V., & Moghaddam, M. M. (2013). The effect of weekly quizzes on students' final achievement score. *Modern Education and Computer Science, 1*, 36-41.

Haberyan, K. A. (2003). Do weekly quizzes improve student performance on general biology exams? *The American Biology Teacher*, *65*(2), 110-114.

Kika, F. M., McLaughlin, T. F., & Dixon, J. (1992). Effects of frequent testing of secondary algebra students. *The Journal of Educational Research, 85*(3), 159-162.

Kling, N., McCorkle, D., Miller, C., Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business*, *8(1)*, 67–72.

Kuku, O. O., & Alade, O. M. (2017). Impact of frequency of testing on study habit and achievement in mathematics among secondary school students in Ogun State, *Nigeria. Journal of Educational Research and Practice, 7*(1), 1.

Kuo, T., & Simon, A. (2009). How many tests do we really need? *College Teaching*, *5(7)*, 156–160.

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

Martin, R. R. & Srikameswaran, K. (1974). Correlation between frequent testing and student performance. *Journal of Chemical Education*, *51*(7), 485-486.

Marshall, B. (2007). A crisis for efficacy? *Education Review, 20*(1), 29-35.

Mines, R. O. (2014). *The impact of testing frequency and final exams on student performance.* American Society for Engineering Education Southeast Section Conference. randomized experiment. *Economics of Education Review*, *30*(6), 1416-1429.

Shah, S. S. H., Munir, T. A., Sabir, M., &Tipu, S. A. (2013). Psychometric Analysis of MCQs Used in Assessing the Students at Entrance to a Medical College. *Annals of King Edward Medical University*, *18*(3).

Shirvani, H. (2009). Examining an assessment strategy on high school mathematics achievement. *American Secondary Education, 38*(1), 34-45.

Zgraggen, F. D. (2009). *The effect of frequent testing in the mathematics classroom* (Masters dissertation). University of Wisconsin–Stout, Menomonie, Wisconsin.