THE EFFECTIVENESS OF ONLINE QUIZZING AS A REPETITIVE LEARNING TOOL IN A MARKETING CLASS: A FIELD STUDY OF THE TESTING EFFECT

William Wellington
University of Windsor
r87@uwindsor.ca

David Hutchinson
University of Windsor
dhutch@uwindsor.ca

A. J. Faria
University of Windsor
ad9@uwindsor.ca

ABSTRACT

A field study of the testing effect (Roediger and Karpicke, 2006) using a low grade weighted online learning management system applied to learning in a large lecture introductory marketing class was undertaken. The testing effect predicts that students who practice more with online quizzes will remember more of what they learned. The subjects were 454 students who completed a twelve week semester in which they were offered nine marketing topic area quizzes, each of which presented seven total attempts broken into three different time frames. The grade weight of the quizzes was very low (1.11% each) with the best attempt score being counted. At the conclusion of the course the marketing students were doubly classified into high, medium and low repetition attempt groups and high, medium and low examination performance groups. An ANOVA analysis of repetition grouping versus final and midterm examination percentage performance and then examination grouping versus total quiz attempts was undertaken. The findings showed significant differences in midterm and final examination performances for the repetition groups when controlling for inclass quiz performance. The ANOVA of the high, medium and low examination performance groupings indicated that for the final examination, the low performance group had statistically significant fewer quiz attempts than the medium performance group which also had statistically significant fewer quiz attempts than the high exam performance group. The findings for the midterm examination were not as strong as there was only a marginally significant difference (.065) between the low exam performance group and medium exam performance group in terms of total quiz attempts while there was a statistically significant difference in total quiz attempts between the high performance group and the medium performance group. The conclusions drawn from these findings are that the testing effect was present and that low stakes quizzing was beneficial to student learning.

INTRODUCTION

An often quoted maxim is that "practice makes perfect" which suggests that repetition can lead to better performance. In the sporting world the maxim is altered to "practice makes habit, perfect practice makes perfect" which suggests that repetition with good quality leads to better performance. Regardless, the notion being presented is that repetition can lead to better performance in one's endeavours. Roediger and Karpicke (2006) report on the "testing effect" and provide strong evidence that indicates that repeated testing leads to greater memory retention than repeated study. Psychological research on memory is done with strict controls but course instructors are interested as to whether the theories on memory retention will produce similar results when they are applied to a practical field setting in education.

In this study, the authors investigated the implementation of the testing effect in a marketing class using a series of short and time spaced (Thalheimer 2006) online multiple choice quizzes to encourage students to engage more frequently with the learning material. The authors sought to observe whether this more frequent engagement would lead to better midterm and final examination performances as would be suggested by the findings of Roediger and Karpicke (2006). Understanding the right amount of frequency of practice is also an issue of interest. The authors felt that at some point repetition might also become tedious and boring. The authors do not directly study the impact of wear-out associated with repetition in this paper but do acknowledge it's potential and thus report participation rates over time to identify when wear-out might be occurring. The frequency of practice was measured through the number of quiz attempts the students undertook while the quality of the practice was assessed according to the performance of the students on these quizzes.

STUDY BACKGROUND AND PURPOSE

The use of online quizzing and testing to assess student learning is a common practice in education (Gikandi, Morrow and Davis, 2011). However, the use of online quiz testing as a learning management system tool is relatively new (Angus and Watson, 2009) while the concept of using testing to enhance learning retention (known as the testing effect) has a very long history in psychological research (Roediger and Karpicke, 2006, 184-189; Tulving 1967). Using repetitive testing to enhance memory involves a number of key considerations including the frequency of testing, time spacing between tests, the lag-time between learning/testing and the learning assessment (examination) (Kupper-Tetzel, Erdfelder and Dickhauser, 2014).

Although not identified as the purpose in their study, DeSousa and Fleming (2003) actually illustrated the impact of the "testing effect" when they compared the examination performance of students who undertook quiz related learning using online quiz testing versus in-class quiz testing in a Theories of Personality course. The students were divided into two different course administration groups where they were examined four times during the course and were required to complete a total of 16 quizzes (four quizzes per exam) during the course. The in-class tested students were given paper and pencil quizzes, while the online tested students were allowed to take their quizzes anytime they wished and were offered up to five repetitions of their quizzes for grading purposes and more practice attempts which were ungraded. As such, the online students were actually a "testing effect" group. DeSousa and Fleming (2003) essentially offered the online students unlimited opportunities for repetition while the in-class students were limited to a maximum of 16 quiz attempts. In designing their study, DeSousa and Fleming (2003) acknowledge the impact of the spacing effect and were careful to offer the quizzes at different time intervals. The results of their study demonstrated that the examination performance of the class that undertook online quizzing was better than the group who undertook paper and pencil quizzing and this difference was significant. In essence, the "testing effect" was likely occurring and led to better examination performance. However, the focus of the study was a simple comparison of the performance of the two groups who were subjected to essentially two different treatments. The authors did not analyze or report on the impact of the number of potential repetitions for the online students versus the in-class students in this paper.

Angus and Watson (2009) report on the use of online quizzing in a business mathematics course as a tool to enhance learning and performance as measured on examinations. Once again, their study was in support of the "testing effect" and the impact of repetition. They stated the following: "Our main finding is that exposure to regular (low-mark) online testing significantly improves student learning as measured by a final proctored examination. Importantly, this result is independent of a student's actual performance on each online quiz" (Angus and Watson, 2009, p. 256). In contrast, Andergassen, Modritscher and Neumann (2014) undertook a study of the effects of repetition on final examination performance on a paper and pencil MCQ exam with a sample of 1850 students who were enrolled in three different online courses. They report weak correlations between repetition of learning exercises and

TABLE 1
MEANS AND STANDARD DEVIATIONS OF COURSE PERFORMANCE MEASURES

Course Performance Measure	N	Mean	Standard Deviation
Midterm Examination % of 100	452	68.34	12.55
Final Examination % of 100	451	64.48	12.68
In-class Quiz Grades (Max 10)	454	7.84	2.32
Online Quiz Grades (Max 10)	454	7.10	2.30
Online Quiz Attempts (Max 63)	454	26.93	12.52

FIGURE 1

Time Spaced Repetitive Learning Design for Each of the Nine Topic Quizzes

	Prequiz	Postquiz	Exam Quiz
Available Attempts	2	2	3
Time Spacing	1 week prior to lecture	1 week after-lecture	Few days prior to exam
Number of Questions	10	10	10
Quiz Duration	10 minutes	8 minutes	7 minutes

Note: For the very first quiz (Pre-Quiz 1) the time spacing was 1 week prior to lecture but owing to academic rules at the University, this quiz could not be opened or closed until after the first class meeting and thus departs from the planned design.

examination performances in their studies which call into question whether repetition does lead to better learning and the impact of the "testing effect".

Thalheimer (2006) reports on the impact of repetition of learning experiences with spacing. Thalheimer (2006, p. 6) presents the following definition: "So what is the spacing effect? It is the finding that spaced repetitions produce more learning—better long-term retention—than repetitions that are not spaced. It is also the finding that longer spacings tend to produce more long-term retention than shorter spacings (up to a point where even longer spacings are sometimes counterproductive)." Thalheimer discusses how cramming, which involves a lot of repetition for short-term memory retention, can be effective for remembering knowledge for a short period but this knowledge is not retained over time.

Kupper-Tetzel, Erdfelder and Dickhauser (2014) report on the impact of repetition, time spacing and the lag-time between learning and repetition on memory testing. The lag effect is related to the spacing effect in that it seeks to determine the optimal time between learning repetitions to provide the maximal memory retention. In their paper, these researchers reviewed the literature on the concepts of the spacing effect and the lag effect which they characterize as follows: "For example, laboratory studies have demonstrated that long-term retention of a wide range of to-be-learned materials can be enhanced when multiple restudying units are not massed together, but rather distributed over time. . . . It has also been established that long-term memory benefits more from multiple relearning units that are separated by long lags instead of short lags" (Kupper-Tetzel, Erdfelder and Dickhauser, 2014, p. 374). In their study they report that if a test is going to occur a week later, a one-day lag between initial learning and relearning is best. In contrast, a 35 day-lag between initial learning and relearning produced the best results for maintaining long-term memory. These kinds of time-frames suggest that the pace of learning needs to be slower if long-term retention of the learned material is going to be maximized. The findings regarding the lag-effect for memory retention run counter to the educational design of most college and university courses which tend to compress the amount of time for learning while expanding the quantity of material to be learned. Although beyond the scope of this paper, it does cause one to think that based on the findings surrounding lag-effects for learning retention, it should come as no surprise that post-secondary students have trouble demonstrating retained knowledge.

The purpose of this study was to explore the following research question: Based on the testing effect (Roediger and Karpicke, 2006) and using spacing (Thalheimer, 2006) would students who participated more frequently in the offering of repeated multiple choice quizzes have better examination performances on midterm and final examinations than students who participated less frequently?

In order to test this research question, the following hypotheses were formulated:

H1: Students grouped into high, medium and low categories based on their quiz attempts will have significantly different midterm and final examination performances with the high attempt group having higher midterm and final exam scores than the medium attempt group which, in turn, will have higher midterm and final exam scores than the low attempt group.

H2: Students grouped into high, medium and low categories based on their respective midterm and final examperformances will have significantly different levels of online quiz attempts with the high examination performance group having more online quiz attempts than the medium examination performance group which, in turn, will have higher online quiz attempts than the low examination performance group.

METHODOLOGY

The subjects for this study were 454 first year students enrolled in a Principles of Marketing course which is open to all students at the university. The students attended one of two sections of a one semester course of 12 weeks duration taught by the same instructor who used the same textbook, syllabus and evaluation scheme in each section. The two sections were taught during the same winter semester. The course was designed as a basic mass lecture style course delivered to class sizes of up to 250 students using multiple instructional and evaluation approaches. The course had two major examination evaluations: a multiple choice midterm examination worth 40% of the course grade with a 70 minute time limit which was composed of 65 questions and a multiple choice final examination worth 40% of the course grade with a 120 minute time limit which was composed of 100 questions. The focus of these examinations was to measure content learning. As part of the class instruction, students were asked to take in-class quizzes worth 10% of the course grade. The in-class quizzes were used as teaching tools and were composed of pop-up multiple choice questions contained within powerpoint presentations. These quizzes were graded for completion only

TABLE 2
Correlations of Course Performance Measures

	Final Exam	Inclass Quiz	Online Quiz Grades	Total Quiz Attempts
Midterm Exam	.738 (.000)	.304 (.000)	.309 (.000)	.262 (.000)
Final Exam	1	.303 (.000)	.373 (.000)	.252 (.000)
In-class Quiz Grades		1	.511 (.000)	.486 (.000)
Online Quiz Grades			1	.618 (.000)

and represent a measure of class participation rather than competency knowledge. The "lag effect" was not studied or controlled for in this study due to the compressed nature of the course semester (12 weeks) and the constraints of the course design which involved learning new material right up to as short as one day prior to the administration of the midterm examination. As such, one day and thirty-five day lag periods for review of new materials as suggested by the research of Kupper-Tetzel, Erdfelder and Dickhauser (2014) could not be fit into the course design and ergo, the research design.

In order to implement the testing effect which was designed to support the content learning and examination preparation, the students were required to complete nine online topic quizzes associated with nine main topics in marketing. Each of the topic quiz assessments was offered with multiple assessments to encourage repetitive completion and time-spaced learning efforts. The multiple assessments involved two distinct times for credit and one time for practice but no credit. The basic design (See Figure 1) involved a time-spaced learning approach to encourage students to undertake multiple engagements with the material with the intent being that time-spacing and multiple engagements would lead to more learning and better learning retention prior to the examinations. The students were offered three opportunities to engage with each of the nine learning topics by taking quizzes which were time-spaced as follows: two attempts available up to one week before the material was lectured upon with a 10 minute time limit for each quiz (identified as a pre-quiz and graded for course credit); two attempts up to one week after the material was lectured upon with an 8 minute time limit for each quiz (identified as a post-quiz and graded for course credit); and three attempts a few days prior to the delivery of either their midterm or final examination but closing the day before the actual examination with a 7 minute time limit for each quiz (identified as an exam quiz which was graded but not for course credit).

There was one important exception to this design. The very first pre-quiz could not be made available to the students prior to the first class meeting because they needed to be briefed on how to use the quizzing system and under the institutional rules of the University, they could not be assessed prior to the first class. The students could know their quiz performance scores but could not actually review the specific questions and answers until the examination quizzes had closed which was the day before the respective exams. The purpose of this delay was to encourage students to re-engage with the basic material for learning after each quiz performance rather than simply learning the answers to the test bank questions and memorizing them which could happen if they could access the answers to the quiz questions immediately after having taken them.

Each quiz was made up of 10 multiple choice questions selected at random from a large test bank which provided coverage on virtually all of the key concepts of marketing presented in the course. The test-banks were constructed as follows: Quiz 1) the definition, philosophies and role of marketing - 34 questions; Quiz 2) strategic market planning – 27 questions; Quiz 3) marketing environments and marketing ethics – 31 questions; Quiz 4) buyer behavior – 55 questions; Quiz 5) market segmentation and information for marketing decision making – 39 questions; Quiz 6) product management – 45 questions; Quiz 7) distribution management – 38 questions; Quiz 8) promotion management – 40 questions; and Quiz 9) pricing management – 34 questions. Each of the quizzes was worth 1.11% (collectively 10%) and the quizzes were offered in an online format associated with a University based website known as CLEW (Collaboration and Learning Environment) which is based on the Sakai open source software. The students were given a preset schedule of when the quizzes would be available and were informed that deadlines would be strictly enforced and that technical difficulties would not be accepted as an excuse for incomplete quizzes or failure to complete quizzes. They were also instructed to complete a connection test quiz of meaningless computer operation questions prior to taking their actual quizzes to ensure that they had a proper and working internet connection.

Measures of performance on class learning activities involved the online quiz scores which were worth 10% of the

TABLE 3
ANOVA Comparison of Midterm Exam % Grades, Final Exam % Grades By High, Medium and Low Total Quiz Attempt Groupings

	Total Quiz Attempt Groupings			ANOVA	
	High >35 (n=109)	Med 19-35 (n=230)	Low <19 (n=113)	F	Sig.
Mean Midterm Exam Performance %	72.69	68.44	63.95	14.25	.000*
Bonferonni Contrast Tests					
High vs Med					.009*
High vs Low					.000*
Med Vs Low					.004*
Mean Final Exam Performance %	68.35	64.60	60.46	11.22	.000*
Bonferonni Contrast Tests					
High vs Med					.029*
High vs Low					.000*
Med Vs Low					.012*

course grade, in-class completion quizzes worth 10% of the course grade and then a midterm and final examination each worth 40% of the course grade. The class performance on the learning measures is reported on in Table 1.

The in-class completion quiz activity represents a covariate of diligent student behaviour which could be controlled for to determine if repetition was really beneficial or if it was simply a measure of student diligence. At the conclusion of the course the marketing students were doubly classified into three repetition attempt groups and three examination performance groups based on the number of online quiz attempts available to them for learning preparation and examination performance on the midterm and final examinations. The groupings were based on an approximation of a normal curve with a lower 25%, a middle 50% and an upper 25% for number of quiz attempts to form the repetition attempt groups and then the lower 25% examination performances, middle 50% examination performances and upper 25% examination performances for the midterm and final examination groups. These groups were then used to undertake two ANOVA analyses of repetition grouping versus final and midterm examination percentage performance using in-class quizzes as a covariate and then examination grouping versus total quiz attempts using in-class quizzes as a covariate to test the hypotheses that were presented.

FINDINGS

The direct measures of relationships between examination performance and number of quiz attempts via correlation are reported on in Table 2. These findings indicate that the strength of relationship between examination grades and number of quiz attempts are significant but weak because their correlations are less than .30 (Cohen and Cohen, 1978, p. 61, reports that correlations greater than .50 are considered strong, those between .3 and .5 are considered medium and those between .1 and .30 are considered weak). These correlational findings are greater than those reported by Andergassen, Modritscher and Neumann (2014) who reported values ranging from .08 to .20 between the frequency of repetition and examination performances. Online quiz grades and online quiz attempts did have a strong correlation value of .61 which might be expected given students were encouraged to retake quizzes to improve performances. However, online quiz grades and examination grades had medium-low correlations of .309 between quiz scores and midterm exam performance and a correlation of .373 between online quiz grades and final examination performance. In class quiz grades and online quiz grades had a strong correlation of .511 indicating that class attendance and online quiz performance exhibited some relationship. However, class attendance and examination performance had only medium correlations of .304 and .303 with the midterm and final examination grades, indicating that class attendance is not a strong predictor of examination performance. As might be expected, midterm examination performance was strongly related to final examination performance with a correlation of .73.

H1 was tested using an ANOVA analysis of repetition attempt group versus examination percentage performance. The findings of this analysis showed there were significant differences in midterm and final examination performances for the repetition groups when controlling for in-class quiz performance (see Table 3). On the basis of Bonferroni contrast tests the low repetition attempt group (<19 quiz attempts) had statistically significant lower midterm and final examination performances than the medium attempt group (19-35 quiz attempts) and the medium attempt group had statistically significant lower midterm and final examination scores than the high attempt group (>35 quiz attempts). The comparison of the exam performance of the low attempt group and the high attempt group was also statistically significant. As such, H1 was accepted.

Similarly, H2 was tested using an ANOVA analysis of the high, medium and low examination performance groupings (see Table 4). The findings of this analysis indicated that for the

TABLE 4
ANOVA Comparison of Total Quiz Attempts By High, Medium and Low Exam Performance Groupings

	Exam Performance Groups			ANOVA	
	High (n=112)	Med (n=223)	Low (n=116)	F	Sig.
Mean Quiz Attempts vs Final Exam Group	31.35	27.17	22.82	14.25	.000*
Bonferonni Contrast Tests					
High vs Med					.009*
High vs Low					.000*
Med Vs Low					.005*
	(n=107)	(n=222)	(n=123)	F	Sig.
Mean Quiz Attempts vs Midterm Exam Group	31.97	26.64	23.50	14.30	.000*
Bonferonni Contrast Tests					
High vs Med					.001*
High vs Low					.000*
Med Vs Low					.065**

final examination, the low performance group had statistically significant fewer quiz attempts than the medium performance group, which also had statistically significant fewer quiz attempts than the high final exam performance group. As expected, the high final exam performance group had more attempts than the low performance groups and the difference was statistically significant. The findings for the midterm examination were not as clear as there was a marginally significant difference between the low midterm exam performance group and medium midterm exam performance group (.065) in terms of total quiz attempts. There was, though, a statistically significant difference in total quiz attempts between the high performance group and the medium performance group on the midterm exam. As was the case with the final exam, the high performance midterm exam group had significantly more quiz attempts than the low performance midterm exam group. Given the overwhelming levels of significance for most of the comparisons on the examinations, H2 is accepted with some caution.

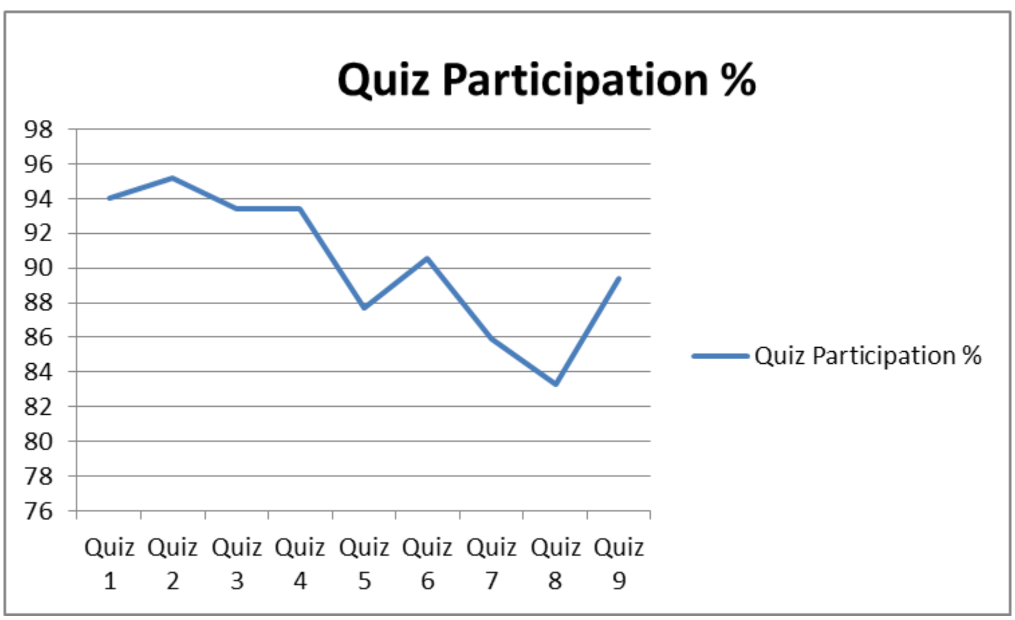
DISCUSSION AND CONCLUSIONS

The research reported here sought to explore whether the testing effect involving online quiz repetition on low stakes quizzes using spacing would lead to better examination performances. The findings indicate there is a weak direct relationship between the total number of quiz attempts and examination grade performance (Table 2 - correlation coefficients of .262 between midterm examination score and the total number of quiz attempts and .252 between the final examination score and the total number of quiz attempts with both significant at the .000 level). In addition, the correlation between performance scores on the quizzes and the final examination and midterm examination scores were in the .30 range indicating a medium-low relationship.

The authors were a bit surprised by this medium relationship in light of the strong relationship between the midterm and final exam scores. The quizzes were essentially the same types of questions as the midterm and final exam questions so the lack of relationship was not due to the difficulty of the questioning. There are a couple of possible explanations for the lack of correlation. In many cases the students were using knowledge aids while undertaking the quizzes. It is likely that they could not repeat the same performance on the examinations because they could not use knowledge aids and thus performed better on the quizzes and worse on the examinations. It is important to state that the online quizzes were designed as learning exercises and not assessment exercises so the use of learning aids was not discouraged in the least. Even collaborative learning could be expected although it was discouraged when it came to actually taking the online quizzes. Students were asked to work on their own while taking the online quizzes, but there was no way to monitor for sure that they did so. Conversely, given the low stakes nature of the online quizzes, some students may have attempted them when they were not prepared at all and thus used the quizzes as their learning aids and tools. As such, their performance could well be worse than that of the examinations as they were gaining their first exposure to the material rather than testing their competency knowledge with the quizzes which was the intention. Given both these scenarios, it could well be expected that quiz performance and examination performance might not be correlated. In addition, these findings are not dissimilar from those of Angus and Watson (2009, p. 256) who reported that they found improved examination performance related to online testing was independent of quiz performance results.

The ANOVA findings of the high, medium and low attempt groupings reveal that students who undertook a high number of quiz attempts outperformed students who undertook

FIGURE 2
Quiz Participation Percentage of all Class Members by Quiz



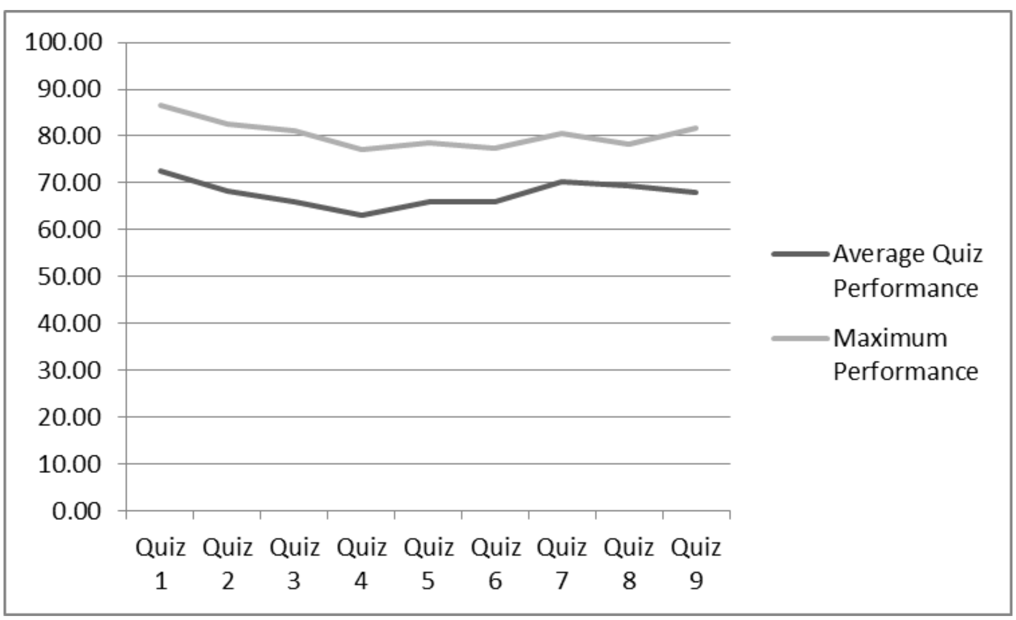
a medium or low number of attempts on both the midterm and final examination when controlling for class attendance as a measure of student diligence to studies. The ANOVA findings of the high, medium and low examination groupings indicate that better performers also attempted more online quizzes when controlling for class attendance. These findings are asserted to be evidence that the testing effect seems to be present in this field setting.

The authors wondered if a fatigue factor or boredom factor would set in given the large number of tests for the students (e.g., as many as 63 distinct tests taken over a 12 week period). Participation rates dropped off amongst the general class population from as high as 95% in the early quizzes into the mid-to-high 80% levels in the later graded quizzes even though the material was new and the later quizzes had the same grade weight as earlier quizzes (see Figure 2). In addition, the ungraded examination quizzes had extremely low participation rates for both the midterm (ranging from a maximum of 50.6% for a first attempt to a low of 10.4% for a third attempt on one of the quizzes) and final examinations (ranging from a maximum of 16.7% for a first quiz attempt to a low of 5.29% for the third quiz attempt). The lower participation rates on these ungraded quizzes suggest that graded performance was an important motivation for taking the quizzes and not just learning. Although the dramatic decline in maximum participation rates from the midterm examination to the final examination quizzes might be indicative of fatigue or boredom with the number of quizzes this may not be the only explanation.

Although participation rates fell off over time, the level of performance on the quizzes remained fairly steady from start to finish (see Figure 3) with maximum performance score averages in the 80% range for the class. Given students were typically achieving high quiz scores in the 80% range amongst those who completed them they may not have felt the need to repeat the process in preparation for their midterm and final examinations. It was a bit surprising to discover that the very first quiz had the highest performance score. One would think that this would come later after students had taken a few quizzes and developed some experience with them. However, the explanation for this is likely related to the one day lag effect as described by Kupper-Tetzel, Erdfelder and Dickhauser (2014). Although made available one week before the first lecture the academic assessment rules at the University where the study occurred required that no graded assessment could be given prior to the first class. As such, students had the advantage of receiving a lecture prior to the Quiz 1 pre-quiz assessment and thus it would represent a one day lag assessment situation which could result in higher retention and thus higher performance. The first quiz did have the highest grades associated with it.

The study has a number of very important limitations. Firstly, this report does not include a control group who took the same course, from the same instructor, and had the same examinations. As such, a direct comparison of those students who were learning with and without repeated testing cannot be made. Secondly, the study design did not randomly assign students to treatment groups with different levels of repetition required. This would have produced a more rigorous measure of the impact of different levels of repetition on examination performance. This kind of treatment is constrained by ethical considerations because within class and between classes

FIGURE 3
Average and Maximum Quiz Performance Percentage by Quiz



Note: For the very first quiz (Quiz 1) the time spacing was 1 week prior to lecture but owing to academic rules at the University, this quiz could not be opened or closed until after the first class meeting and as such performance results for this quiz were anomalous in comparison to all of the other quizzes evaluation methods have to be the same during any particular semester of instruction. Therefore, to develop a control group the course would have to be changed between semesters. Alternatively, a simultaneous study involving multiple institutions and multiple classes could be designed to address this but this design would also introduce a host of confounds unless the same textbook and course syllabus was adopted at all institutions and they all had similar lengths of semesters, etc. The authors would need to investigate the feasibility of this approach.

Another key limitation is that there were no controls or attempts to study the lag-effect (Kupper-Tetzel, Erdfelder and Dickhauser 2014) which may have had an impact on the learning performance of all of the students in the class as the time-frame between quiz repetitions may have been less than optimal. The impact of the lag effect may have been exhibited by the aforementioned high level of performance on the first quiz.

Despite these limitations, the findings of this study were not dissimilar to those of past studies (DeSousa and Fleming, 2003; Angus and Watson, 2009; and Andergassen, Modritscher and Neumann, 2014). Based on these findings, the authors believe that there is evidence that the testing effect (Roediger and Karpicke, 2006) was occurring in the course and that it was beneficial to the learning of the students taking the marketing class. As such, the use of online testing to encourage learning and retention of learning in marketing courses is justified.

However, the authors did note that although the frequency of testing seemed to be related to better examination performance, there appeared to be an upper threshold on the amount of testing that students were willing to engage in because participation rates began to fall for later quizzes - even ones that were graded. Based on the results of this study and also the findings of DeSousa and Fleming (2003), it would appear that graded testing is necessary to motivate students. The participation rate fall off for graded quizzes and the lower participation rates for ungraded quizzes suggest that repeated testing levels beyond four or five are not accepted by the majority of students.

The following suggestions for future research are made: 1) to implement online quizzing that takes into account short-term lag effects which would be practical for university and college level semester time frames; 2) to examine the impact of eliminating non-graded quizzes by replacing them with graded quizzes to increase participation rates; and 3) to try and determine what the optimal number of quiz attempt offerings would be appropriate for maximum learning retention.

REFERENCES

Andergassen, M., Modritscher, F., and Neumann, G. (2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations With Learning Results. *Journal of Learning Analystics,* 1 (1), 48-74.

Angus, S.D. and Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology*, 40 (2), 255–272.

Cohen, J. & Cohen, P (1983). *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences*, Hillsdale, New Jersey: Erlbaum Press, 25-78.

DeSousa, E. and Fleming, M. (2003). A Comparison of In-Class and Online Quizzes on Student Exam Performance. *Journal of Computing in Higher Education*, 14 (2), 121-134.

Gikandi, J.W., Morrow, D., and Davis, N.E. (2011), Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57, 2333–2351.

Kupper-Tetzel, C.E., Erdfelder, E., and Dickhauser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science*, 42:373–388.

Roediger III, H.L. and Karpick, J.D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science,* 1 (3), 181-210.

Thalheimer, W. (2006, February). *Spacing Learning Events Over Time: What the Research Says.* Retrieved November 12, 2014, from http://www.work-learning.com/catalog/

Tulving, E. (1967). The Effects of Presentation and Recall of Material in Free-Recall Learning. *Journal of Verbal Learning and Verbal Behavior*, 6, 175-184.