

## Gamifying Online Tests to Promote Retrieval-Based Learning

Maristela Petrovic-Dzerdz

Volume 20, Number 2, April 2019

URI: <https://id.erudit.org/iderudit/1061329ar>  
DOI: <https://doi.org/10.19173/irrodl.v20i2.3812>

[See table of contents](#)

Publisher(s)

Athabasca University Press (AU Press)

ISSN

1492-3831 (digital)

[Explore this journal](#)

Cite this article

Petrovic-Dzerdz, M. (2019). Gamifying Online Tests to Promote Retrieval-Based Learning. *International Review of Research in Open and Distributed Learning*, 20(2). <https://doi.org/10.19173/irrodl.v20i2.3812>

Article abstract

Recent findings have provided strong evidence that retrieval-based learning is an effective strategy for enhancing knowledge retention and long-term meaningful learning, but it is not a preferred learning strategy for the majority of students. The present research analyzes the application of learning gamification principles in online, open-book, multiple-choice tests in order to motivate students to engage in repeated retrieval-based learning activities. The results reveal a strong positive correlation between the number of successful retrieval attempts in these tests that cover content from the course textbook, and long-term knowledge retention as demonstrated in a live, final, closed-book, cumulative exam consisting of multiple-choice, labeling, definitions, and open-ended questions covering the content of both textbook readings and lectures. The presented results suggest that online, open-book tests designed using gamification principles, even when covering partial course content and one type of questions, are an effective strategy for using educational technology to motivate students to repeatedly engage in retrieval-based learning activities and improve long-term knowledge retention, regardless of the course delivery mode.

Copyright (c) Maristela Petrovic-Dzerdz, 2019



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

**Érudit**

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

April – 2019

# Gamifying Online Tests to Promote Retrieval-Based Learning

Maristela Petrovic-Dziedz  
*Carleton University, Ottawa*

## Abstract

Recent findings have provided strong evidence that retrieval-based learning is an effective strategy for enhancing knowledge retention and long-term meaningful learning, but it is not a preferred learning strategy for the majority of students. The present research analyzes the application of learning gamification principles in online, open-book, multiple-choice tests in order to motivate students to engage in repeated retrieval-based learning activities. The results reveal a strong positive correlation between the number of successful retrieval attempts in these tests that cover content from the course textbook, and long-term knowledge retention as demonstrated in a live, final, closed-book, cumulative exam consisting of multiple-choice, labeling, definitions, and open-ended questions covering the content of both textbook readings and lectures. The presented results suggest that online, open-book tests designed using gamification principles, even when covering partial course content and one type of questions, are an effective strategy for using educational technology to motivate students to repeatedly engage in retrieval-based learning activities and improve long-term knowledge retention, regardless of the course delivery mode.

*Keywords:* gamification, retrieval-based learning, multiple-choice tests, online, learning management system, learning analytics

## Introduction

Regardless of course delivery mode (face-to-face, blended, or online), the challenge often encountered by the author as an instructional designer is to devise instructional strategies that motivate students to study frequently and not procrastinate. This is particularly important for student success in content-heavy courses such as first- or second-year science courses, which usually cover a broad range of declarative knowledge and numerous concepts as a foundation for further studies. Additionally, these courses are typically offered in a traditional face-to-face format and experience large enrollment numbers, the combination of which can pose a challenge for effective and efficient formative assessment and feedback, which are essential to supporting learning success. Despite the aforementioned challenges, students in these courses certainly benefit from any learning activity that can help them integrate and retain the knowledge they need to master, and they deserve the effort invested by course designers to devise such activities.

### Retrieval-Based Learning

Recent findings have provided strong evidence that practicing active retrieval (recall) enhances not only long-term memory but also long-term meaningful learning, supporting the claim that these types of learning strategies could be more effective than many currently popular “active learning” strategies (Blunt & Karpicke, 2011). This confirms that what has traditionally been considered as learning—the “importing” of new information and its integration with existing knowledge—is only one aspect of the learning process, and that another equally important aspect of learning consists of the retrieval processes; specifically, those “involved in using available cues to actively reconstruct knowledge” (Karpicke, 2012, p. 158). According to Nunes and Karpicke (2015), although the idea that practicing active recall improves learning has existed for centuries, it has undergone a significant revival with increased interest owing to the integration of cognitive science research and educational practice. Nunes and Karpicke use the term “retrieval-based learning” to encompass both the instructional strategies that promote this type of learning and the fact that the process of retrieval itself enhances learning. Although there is strong evidence supporting its effectiveness, research also shows that retrieval is still not a learning strategy of choice for the majority of students, nor are they aware of its positive effects (Karpicke, 2012).

One of the main tasks of instructional designers is to identify strategies to make learning experiences effective and efficient, and to improve knowledge retention. Furthermore, they need to find ways to both extrinsically and intrinsically motivate learners to engage in learning activities that normally require significant effort and include the experience of failure. Black and Wiliam (2010) note that if they have a choice, students will avoid difficult tasks; they also point to a “fear of failure” that can be detrimental for learning success. Although making mistakes and experiencing failure are essential experiences in every learning process, “pupils who encounter difficulties are led to believe that they lack ability” (Black & Wiliam, 2010, p. 6). Motivating students to persist with repeated engagement in activities that incorporate the experience of both difficulty and failure is a real instructional design challenge, clearly articulated by Karpicke (2012) in the conclusion of his article: “The central challenge for future research will be to continue identifying the most effective ways to use retrieval as a tool to enhance meaningful learning” (p. 162).

The present research attempts to tackle this challenge and to examine ways to motivate students to engage in difficult learning activities that can result in meaningful learning and knowledge retention. The approach examined in this research is the implementation of distributed, open-book, online tests covering relevant content from the adopted textbook, in a foundational, second-year, high-enrollment, core neuroscience course. To motivate students to *repeatedly* engage in these activities, some gamification principles were used. Data analytics from Moodle, a learning management system (LMS), were used to examine student engagement patterns and retrieval success in online tests, while statistical analysis was applied to determine their correlation with long-term knowledge retention as demonstrated in a live, final, cumulative, closed-book summative assessment.

### **Multiple-Choice Tests for Retrieval-Based Learning**

There are many ways to implement retrieval-based learning, with tests being the most researched. Smith and Karpicke (2014) investigated the effectiveness of retrieval practice with different question types (short-answer, multiple-choice, and hybrid) and concluded that retrieval practice with each of these question forms can enhance knowledge retention when compared to a study-only condition. However, short-answer questions must be graded manually, which requires more time. The learning effects are better if students receive feedback in the form of correct answers. This is also not easy to administer with short-answer questions, but it is possible with multiple-choice questions, making them a better solution if we are not able to provide efficient feedback for other question forms. Smith and Karpicke (2014) provide evidence from multiple pieces of published research (Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007; Pyc & Rawson, 2009) that seem to indicate that the need for corrective feedback during the process of learning, along with the balance between retrieval difficulty (questions and problems that require more cognitive effort to answer, such as short-answer questions) and retrieval success (questions and problems that result in more correct answers, such as multiple-choice questions), leads to a consideration for hybrid tests (e.g., a mix of short-answer and multiple-choice questions) as likely the most effective retrieval-practice solution.

Several other studies have provided evidence for the effectiveness of multiple-choice tests as tools to promote learning (Little, Bjork, Bjork, & Angello, 2012; Smith & Karpicke, 2014; Cantor, Eslick, March, Bjork, & Bjork, 2015; Little & Bjork, 2015). According to Little and Bjork (2015), if multiple-choice tests are optimized by properly constructing competitive and plausible alternatives and developing items that assess beyond the knowledge level, they are effective for learning even non-tested information. There is a belief, though, in some parts of the education community that the “testing effect” does not apply to complex materials, but that view has been challenged by Karpicke and Aue (2015), who provide evidence from previous research that this assumption is not correct. Nunes and Karpicke (2015) remind us that the “testing effect” *is* the effect of active retrieval, which, Karpicke and Aue (2015) emphasize, has been repeatedly proven to have positive effects on meaningful learning of complex materials.

Despite previously discussed research that has provided evidence that practicing active retrieval promotes meaningful learning, according to Grimaldi and Karpicke (2014), three major application problems present challenges to the implementation of this learning strategy: 1) a lack of student awareness about the effectiveness of a study method, 2) a lack of student willingness to repeatedly retrieve material, and 3)

student inability to correctly evaluate the success of their retrieval attempts. Based on research by Grimaldi and Karpicke (2014), students struggle with all three of these components necessary for success of retrieval-based learning activities. For this reason, they support the use of computer-based approaches with automated scoring and feedback to aid students in getting the most out of the available learning strategies. In the present research, a novel approach of applying gamification principles to motivate students to engage in repeated retrieval by taking online tests, while receiving automated feedback, was analyzed as a potential solution to two out of three problems outlined by Grimaldi and Karpicke (2014): a lack of student willingness to repeatedly retrieve material and student inability to correctly evaluate the success of their retrieval attempts.

## Learning Gamification

When applied in the field of education, the term “gamification” typically does not refer to playing games but is “broadly defined as the application of game features and game mechanics in a nongame context” (Becker & Nicholson, 2016, p. 62). Some of the core game mechanics include the use of task levels with progressively increasing difficulty, “unlocking” subsequent levels when you reach a certain mastery skill requirement, and an experience of failure not as a deterrent, but as a natural component of the skill-building and learning process. By carefully manipulating our internal mental reward system (assisted by brain transmitters often called “pleasure chemicals,” such as dopamine), game mechanics keep players “hooked” in a continuous task-failure-success upward spiral, contrary to our natural inclination to give up when the task seems unachievable or after repeated failure.

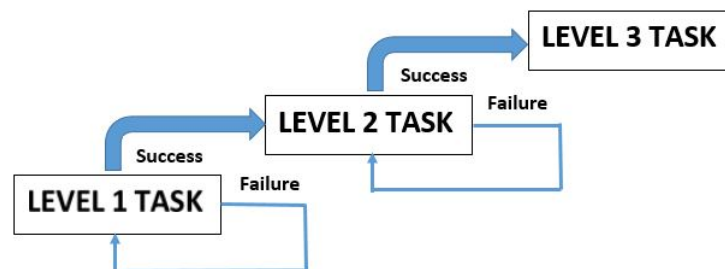


Figure 1. Typical game mechanics.

Even more surprisingly (if we are not familiar with the effects of dopamine), after achieving a goal, a player is typically looking for a bigger challenge, unconsciously hoping for the next level of intrinsic reward in a form of mental pleasure if the challenge is successfully tackled. This typical game activity structure is very conducive to learning, which naturally requires taking on challenges at a progressively increased cognitive level, so it comes as no surprise that gamification strategies are becoming more popular in education.

As suggested by Zichermann and Cunningham (2011), designers who plan a gamification system first must identify behaviours they wish to encourage. The behaviours to be encouraged in this research study were frequent and repeated engagement with (studying from) a textbook and pausing to test comprehension and knowledge retention. A typical student usually does not close a textbook after studying, voluntarily answering the questions posed at the end of each chapter, then checking for correct answers. On the

contrary, research shows that the “majority of students indicated that they repeatedly read their notes or textbook while studying,” and if they engage in any self-testing activity, “they do it to generate feedback or knowledge about the status of their own learning, not because they believe practicing recall itself enhances learning” (Karpicke, Butler, & Roediger, 2009, p. 477). Therefore, to engage students in learning activities that are not their first choice, some innovative strategies in course design must be employed. The usual assessment approach, particularly in online and blended courses, is to design graded online tests and distribute them at a certain frequency in the course schedule. From this author’s perspective, if a face-to-face course can use a LMS, the course can be enriched by taking advantage of technological affordances in a similar way. However, when conducting online tests purely as an assessment activity, as students feel that they have no room for mistakes that can “cost” them a grade percentage, the temptation rises to look for “alternative ways” to answer questions correctly and earn a good grade. Students might try to pair up with another student to take the test, ask an expert for help, or extensively consult available resources. The desired behavior in this study was that students *repeatedly* engage with the content, on their own, with minimal consultation with outside resources. This required providing opportunities for several test attempts, allowance for mistakes without penalty, and rewards for those who follow suggestions for using the available activities in the most effective way. In essence, the goal was to transform a typical formative or summative assessment activity, a multiple-choice test, into a learning activity, and to accomplish this, some game mechanics were borrowed and implemented in online, bi-weekly tests. The goal of the present research was to observe a variety of student-activity engagement behavioral patterns and retrieval success in bi-weekly tests, and to look for possible correlations between them and long-term knowledge retention as demonstrated by a live, final, cumulative exam.

## Method

### Design

This was a correlational study in which the FE test score was the Final Exam score for the student and MTL1A was the total number of mastery test Level 1 attempts. Specifically, the study examined the relationship between live Final Exam test scores, and the number of online Level 1 test attempts that students completed at the mastery level. In this context, mastery means achieving a grade of 80% or higher.

### Participants

Over 200 students in a second-year core neuroscience course in an Ontario post-secondary institution were participants in this retrospective study. Ethics approval was granted to retroactively collect and examine student online activity engagement and success data from Moodle (upon completion of the term), which means that students were not aware of the study while taking the course. The class was a combination of an “in-class” cohort, and a “distance” cohort; the “in-class” cohort of students attended live lectures that were video recorded, while students in a “distance” cohort opted to watch recorded lectures (“video on demand”) at their discretion, instead of attending a live class. However, if “distance” students wanted to attend live lectures, they were encouraged to do so. Similarly, “in-class” students had the option to purchase a recording of live lectures and re-watch them at their discretion. Data were complete for 200 of 204 students

who finished all required components of the course, and whose final exam test score indicated that they took the final exam. Data for the four students whose final exam test score was entered in the Moodle gradebook as “0” were removed from the research data pool because it is highly unlikely that a student who had not withdrawn from the course and who took a final exam would receive a score of “0”. The more likely explanation is that a student did not take the final exam before the grades were collected from Moodle.

## **Course Description**

The content for the course was available to students from two main sources: the textbook and lectures. While the textbook remained as an important learning resource, live lectures supplemented the textbook in order to assist students grasp more difficult concepts. Lecture slides and required weekly textbook chapter readings were posted on the LMS, and online communication was housed in Moodle as ungraded discussion forums. In terms of graded assessment activities, all students, regardless of how they decided to attend lectures, engaged in bi-weekly online tests, and took part in live midterm and final exams. Midterm and final exams were based on the content of both the lectures and the assigned textbook chapter readings, while online tests were based solely on the content of the textbook. Moodle records student activity (e.g., when a student took a test, the answers he or she selected in multiple-choice questions, the test score earned on each test taken, the time it took the student to finish the test, etc.) and allows users with higher role privileges, such as teachers or instructional designers, to look into this information and analyze the success of instructional strategies they have implemented in the online portion of the course.

## **The Frequency and Structure of Online Tests**

Six bi-weekly online tests were designed in Moodle. The textbook publisher provided multiple-choice, short answer, and labeling textbook review questions in a digital format, out of which more than 1,500 questions from the assigned chapter readings were selected to build a question bank for the course. The study adopted a categorization of question bank items as provided by the publisher, where all questions were assigned a cognitive difficulty level of either “easy,” “medium,” or “hard.” Although labeling questions were common in the live midterm and final exams, the version of Moodle used in this course did not allow for this type of question. Before they were entered into the question bank, they were transformed into “matching” questions (match the image with a correct label). Multiple-choice, short answer, and matching questions were used to populate a Moodle question bank, from which the bi-weekly test questions were pulled (Figure A1). Questions were then sorted based on the chapters that were covered in a two-week study period and question type in order to design two test levels: Level 1 and Level 2.

Level 1 tests consisted of 10 multiple-choice questions, each with four possible answers (one correct and three distractors) taken from designated chapter readings, where the questions were pulled randomly from a question bank in a defined pattern: four “easy,” four “medium,” and two “hard” questions per Level 1 test (Figure A2). Students would log onto the course at their preferred time during the week when the test was open and had 60 minutes to complete the Level 1 test. Access to correct answers was not provided during the test or immediately after taking the test; it was made available after the test had been closed for a week. Immediately after taking the Level 1 test, students saw their test score as a percentage out of 100%. Students could earn a maximum of 2% of the course grade by completing each Level 1 test with a 100% success rate.

The format of the exam questions impacts the level of success. Consistent with the work of Smith and Karpicke (2014), short-answer exam questions, which typically require answer construction, are usually considered more difficult in terms of retrieval, while multiple-choice questions, which require answer selection, yield more retrieval success. Hence, to enrich student learning, a hybrid model that combines multiple-choice and short-answer questions may be optimal. As Smith and Karpicke (2014) point out, “hybrid formats can be used to balance retrieval difficulty and retrieval success” (p. 799). Therefore, Level 2 tests were designed to consist of two questions of higher retrieval difficulty—typically one short answer and one labeling question—pulled randomly from a question bank for bi-weekly chapter readings. However, students had to “earn” the right to “play” the next test “level.”

### Gamifying Online Tests

Even when students use retrieval practice as a study strategy, according to Grimaldi and Karpicke (2014), “they do not tend to practice repeated retrieval (additional retrieval beyond recalling items once),” and they “have great difficulty evaluating the accuracy of their own responses” (p. 2). To motivate students to engage with the textbook and repeatedly practice retrieval, bi-weekly tests were designed using the following gamification principles: multiple test attempts without penalty, test levels with progressively increasing difficulty, and advancement to the next test level based on “mastery” of the previous one.

Students were only allowed to take a Level 2 test for the week if they reached a “mastery level,” set as 80% or more success, on the Level 1 test. This is an example of test levels with progressively increasing difficulty and an advancement to the next test level based on the “mastery” of the previous one. To motivate students to keep trying, they were given five opportunities to achieve mastery on the Level 1 test, while only the highest Level 1 test score out of all attempts was recorded in the Moodle gradebook. This is an example of multiple test attempts without penalty. By taking a Level 1 test up to five times, students were, potentially, exposed to new questions, although some questions would repeat since they were randomly pulled from the question bank. There was no minimal time delay set in the test design, which means that students could take a Level 1 test again immediately after the previous attempt.

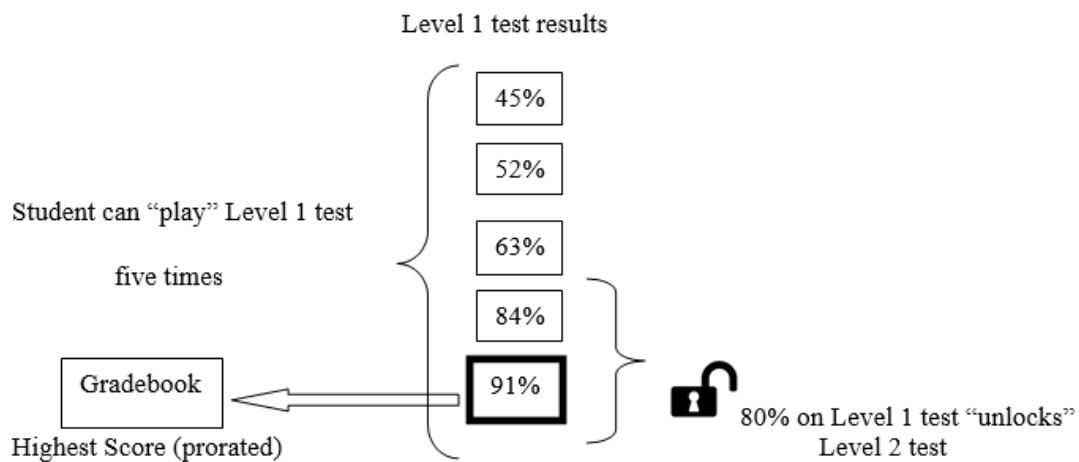


Figure 2. Online test mechanics with applied gamification principles.



Overall, to encourage students to attempt Level 1 tests more than once, several motivational and gamification strategies were used:

- Students could attempt a Level 1 test up to five times each week when the test was open.
- Out of the five available Level 1 test attempts, only the best test score was recorded in a Moodle gradebook (prorated, based on success); therefore, students could not lose a test grade if they took the Level 1 test more than once; they could only improve their grade, and get exposed to more questions, if they “played” the test more times. This approach, borrowed from games, allows for mistakes during repeated attempts, which are essential learning events.
- Upon completion of each Level 1 test, a motivational message was displayed for the students (“overall feedback” in Moodle), congratulating them if they scored high, and encouraging them to keep “playing” if they could still improve their test grade.
- If a student scored 80% or higher on a Level 1 test (“mastery level”), the Level 2 test would “unlock.” Students could only take Level 2 tests once and earn up to an additional 2% of the final grade per test if they managed to score 100%, or they received a prorated grade.
- Out of six bi-weekly tests, the lowest combined score (Level 1 test best grade + Level 2 test grade) would get “dropped” by the Moodle gradebook, so that only the “best five” combined weekly test grades would be used for the final course grade calculations.
- Students were told (in class and in the course syllabus) that a selection of the test questions would be repeated in the live midterm and final exam, as an incentive to get exposed to as many questions as possible by taking Level 1 tests more than once, and “unlocking” Level 2 tests.

Table 1

*Comparison Between Major Assessments in the Course*

<b>Assessment characteristic</b>	<b>Midterm exam and final exam</b>	<b>Level 1 tests</b>	<b>Level 2 tests</b>
<b>Content tested</b>	lectures and textbook	textbook	textbook
<b>Testing environment</b>	live	online	online
<b>Testing condition</b>	closed-book	open-book	open-book
<b>Types of questions</b>	multiple-choice short answer definitions labeling	multiple-choice	short answer matching
<b>% of final course grade</b>	35% (Midterm Exam) 45% (Final Exam)	10% (6 Tests, 2% each, lowest grade “dropped”)	10% (6 Tests, 2% each, lowest grade “dropped”)

Distributed bi-weekly, open-book, unsupervised, graded online activities in the format of tests were worth a maximum of 20% of the course grade, while the closed-book, live midterm exam was worth a maximum of 35% of the course grade, and the closed-book, live, final cumulative exam was worth a maximum of 45% of the course grade.

## Results and Discussion

Smith and Karpicke (2014) suggest that it is not only the number of times students practice retrieval that influences long-term knowledge retention but also how successful they are during the retrieval process. Therefore, we decided to examine the following research variables:

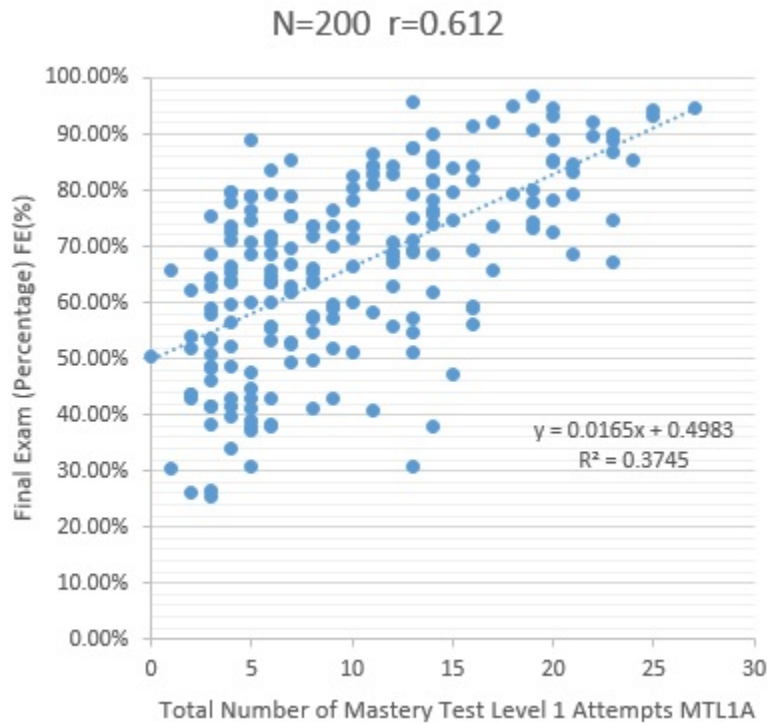
**TL1A** (Test Level 1 Attempts) – total number of times a student attempted Level 1 online tests for the duration of the course (maximum is 30; six tests with five available attempts each)

**MTL1A** (Mastery Test Level 1 Attempts) - total number of times a student reached mastery level (80% or more correct) on Level 1 online test attempts (maximum is 30; six tests with five available attempts each)

**ME** (Midterm Exam test score) – live exam (percentage, maximum is 100%)

**FE** (Final Exam test score) – live exam (percentage, maximum is 100%)

We first looked at the correlation between the number of Mastery Test Level 1 Attempts (MTL1A, max=30) and the Final Exam test score (FE, max=100%), for N=200.



*Figure 3.* Scattergram showing the correlation between MTL1A and FE for N=200.

As Figure 3 indicates, there is a strong positive correlation ( $r = 0.612$ ) between the number of Mastery Test Level 1 Attempts (MTL1A, max=30) and the Final Exam test score (FE, max=100%). The independent variable MTL1A explains 37.5% of the variability in the FE variable (Table A1).

For comparison, we looked at the correlation between the Midterm Exam test score (ME) and the Final Exam test score (FE):

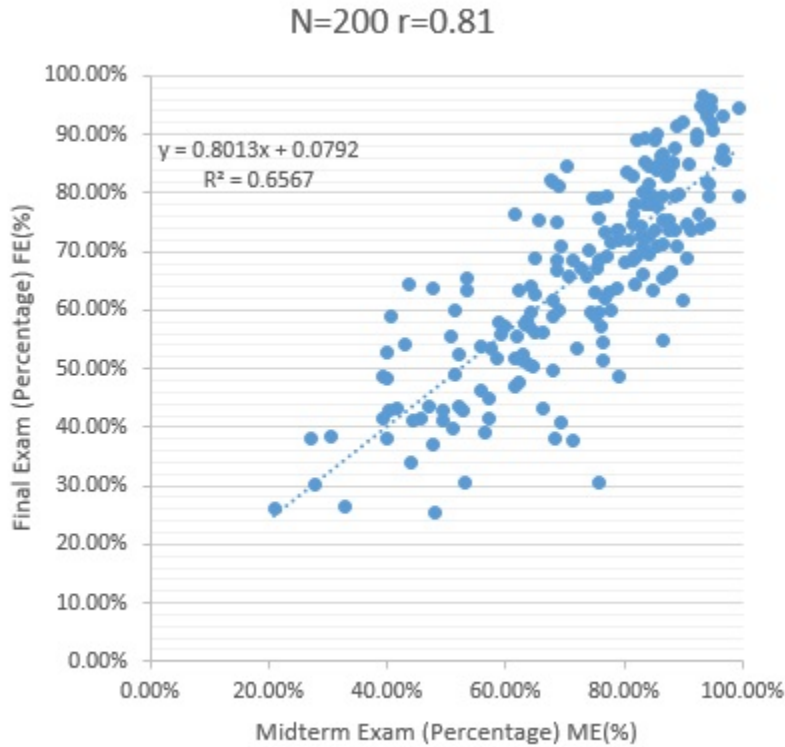


Figure 4. Scattergram showing the correlation between ME and FE for N=200.

As Figure 4 indicates, there is a strong, positive correlation ( $r = 0.81$ ) between the live Midterm Exam test score (ME) and the live Final Exam test score (FE). The independent variable ME explains 65.67% of the variability in the FE variable (Table A2).

Table 2 summarizes the results of this research for the entire class (combined in-class and distance students), as well as separately for in-class and distance cohorts. The average number of times students took Level 1 test was similar across the student population, and about 20 out of 30 available attempts. Likewise, the percentage of Level 1 test attempts resulting in mastery achievement (80% or higher) was similar for both cohorts, about 50%.

Table 2

*Summary of Research Findings for In-class and Distance Cohorts (Combined and Separate)*

Variable\Student cohort	In-class and distance N=200	In-class only N=91	Distance only N=109
-------------------------	--------------------------------	-----------------------	------------------------

<b>Average number of times students took Level 1 test</b>	19.8	20.68	19.04
<b>Average number of times students finished Level 1 tests at mastery level</b>	10.04	10.56	9.61
<b>Correlation between TL1A and MTL1A</b>	r = 0.63 R <sup>2</sup> = 0.40	r = 0.59 R <sup>2</sup> = 0.35	r = 0.65 R <sup>2</sup> = 0.43
<b>Correlation between MTL1A and FE</b>	r = 0.61 R <sup>2</sup> = 0.375	r = 0.58 R <sup>2</sup> = 0.34	r = 0.63 R <sup>2</sup> = 0.40
<b>Correlation between ME and FE</b>	r = 0.81 R <sup>2</sup> = 0.66	r = 0.84 R <sup>2</sup> = 0.71	r = 0.79 R <sup>2</sup> = 0.62

The LMS Moodle allows for monitoring of student online activity and success patterns, and these patterns varied in this course. Some students reached either a mastery level or 100% during early Level 1 test attempts and stopped taking tests before they used up all five available tries. These were the students who seemed concerned only with a good course grade and “unlocking” the Level 2 test; however, they did not use all available opportunities to “play,” get exposed to more questions, and test their knowledge. Other students reached Level 1 test mastery level or 100% during early test attempts but kept “playing” until they had exhausted all five attempts. It seems that these students engaged in activities not only to earn a good grade but also for practicing and learning. On the other hand, there were students who did not manage to reach mastery level during early Level 1 test attempts, but gave up early, did not use all five tries, and never managed to unlock the Level 2 test, even though the Level 1 test was open-book and they had nothing to lose (except time and some mental energy). These students did not seem willing, or able, to invest sufficient effort and time in learning activities, even if they could only gain advantages by doing so. Therefore, by randomly observing student behavior when taking Level 1 tests, we found that both the most and least successful students (based on the final course grade) engaged in a similar way with the online activities. The patterns only started to emerge when we statistically analyzed the data collected from the LMS. It was expected that success on midterm and final exams, which used the same assessment tool and testing environment, and covered the same course content sources, would highly correlate (results reveal  $r = 0.81$ ,  $R^2 = 0.66$  for the whole class). In addition, the results of this study indicate that the total number of times students reached mastery in Level 1 test attempts had a strong positive correlation with success on the final exam ( $r = 0.61$ ,  $R^2 = 0.375$ ). This finding provides support for previously found results which suggest that “actively attempting to retrieve and reconstruct one’s knowledge is a simple yet powerful way to enhance long-term, meaningful learning” (Karpicke, 2012, p. 162). Furthermore, Level 1 tests and the final exam assessment environments were different (open vs. closed book, online vs. live), and so was the scope of content covered in these assessment events (textbook vs. textbook and lectures), as well as the type of assessment questions (multiple-choice vs. multiple-choice, labeling, and short-answer). This finding, consistent with findings of Smith and Karpicke (2014), Karpicke and Aue (2015), and Little and Bjork (2015), suggest that benefits of retrieval-based learning activity may transfer to long-term learning even when the retrieval and final assessment events are in different formats, and of different content scope. The results reveal that for the entire class, 37.5% of the variation in Final Exam test scores (FE) could be explained with a single variable, the number of Mastery Test Level 1 Attempts (MTL1A).

The logistics of this retrospective study did not allow for the collection of subjective student feedback; however, students in comparable courses who are exposed to similar online strategies that promote frequent and continuous engagement with course content often report on the positive effects those strategies have on their motivation to study regularly, as well as on their success in the course. It is important to note that students' comments are anecdotal in nature, and it will be necessary to develop more empirical methods to confirm these observations.

In the present study, there was no delay required between repeated Level 1 test attempts. Research supporting spaced learning or "repeated stimuli spaced by periods without stimuli" (Kelley & Watson, 2013, p. 1), suggests positive effects of test spacing on long-term memory. Further research should examine if introducing a delay between two test attempts (which could be done in Moodle) would result in improved learning effects of repeatedly taking the tests as a retrieval-based learning strategy.

## Conclusion

Online, open-book testing is not usually a preferred assessment option for instructors, who fear abundant opportunities for cheating, ranging from consulting notes and textbooks, and compromising the questions, to an inability to identify the person taking the test. Nevertheless, if we shift the focus from *assessment* to *learning*, then online, open-book tests have promising applications. For retrieval-based learning in the form of testing to be effective, students need to get corrective feedback in a timely manner, and online activities with automated scoring, such as multiple-choice tests, are one possible solution to this challenge. Presented findings indicate that applying gamification principles to motivate students to repeatedly engage with online tests, even if they are open-book and unsupervised, could be an effective option for supporting student learning, especially in content-heavy courses.

There are other positive consequences of taking multiple and frequent tests. Soderstrom and Bjork (2014) provide evidence that testing improves students' subsequent self-regulated study habits by making them more aware of the state of their knowledge, so they can make better decisions when regulating their further study behavior. This could have been a factor which, apart from the effects of repeated retrieval, positively affected student success on the final cumulative exam in this study. Furthermore, Agarwal, Karpicke, Kang, Roediger, and McDermott (2008) provided evidence that both open-book and closed-book tests with feedback can produce similar final performances, while students report less anxiety when preparing for open-book tests. This is another argument why open-book, online tests can be used as a powerful tool to support student learning and long-term knowledge retention.

Although there are positive effects of using multiple-choice tests for variety of purposes in education, there is an inherent challenge in the task of designing a large number of good-quality multiple-choice test items. Little et al. (2012) point out that multiple-choice exams have long been used in high-stake professional examinations and certifications, including different branches of medicine, and as a tool whose results are used to determine acceptance in highly competitive academic institutions and programs. However, the process of creating a question bank for those purposes is rather different from a typical process in an educational setting. In the world of high-stake examinations based on multiple-choice tests, significant

effort by subject matter experts is put into a peer-reviewed process of constructing multiple-choice questions that have a carefully crafted stem and plausible incorrect alternatives, thus minimizing the possibility of simply “recognizing” the correct answer. The same amount of effort is rarely invested in constructing a question bank in an educational environment, even for the purpose of high-stake assessment, as it normally requires an assembly of a “writing cell” (a group of subject matter experts, usually the instructor and one or more graduate students, who are willing to invest significant time and effort in writing a required number of good quality test items after being trained on this assessment tool design and utilization). This “luxury” is rarely present in post-secondary environments, which, more often than is ideal, results in multiple-choice test items of inconsistent quality, written by course instructors who have not been adequately trained to develop effective items and conduct item analysis. This can, unfortunately, result in the poor reputation of the assessment tool. The present research examined the option of using a textbook publisher-provided question bank (even if the questions are already available to students or have been used by many institutions) by creating a version of a low-stakes formative assessment/learning tool, which gives questions a “second life,” and, based on this research, a potentially very valuable one.

Despite the promising effects of the learning approaches implemented in this study, challenges for improving awareness about the effectiveness of these strategies within the teaching community remain, which brings into focus an apparent and growing need for teachers in post-secondary education to either work closely with professionals, such as instructional designers and educational developers, or to engage in professional development programs to gain necessary insight into the latest research on effective learning, and to improve skills needed to design activities that can positively affect students’ learning habits and aid the processes of knowledge acquisition and retention.

Regarding the implementation of ubiquitous educational technology such as learning management systems, instead of using such systems primarily as content repositories and platforms for online discussions and assessment, this research suggests that there is a real benefit to using them in the full sense of their name: as technological tools to help students manage learning and succeed in a world with growing demands for mastering complex knowledge and skills. Another useful application of a LMS that can be derived from this research is the effective use of data analytics. Our knowledge about the learning strategies students choose when studying on their own has always been based on self-reported data, which repeatedly confirmed that students often use less efficient study strategies, such as re-reading. In this research, and for the first time, LMS enabled data analytics enabled us to gain a glimpse into student learning behaviors and habits in a “non-invasive” way by retrieving the data for students who had previously completed the course.

The challenge posed by Karpicke (2012), to identify the most effective ways to promote retrieval as a learning strategy, continues. The method described in this study is one example of concrete activities instructors can implement in their course to enhance student learning through the implementation of online retrieval-based learning activities, whether the course mode of delivery is face-to-face, blended, or online. Media might be, as Clark (1983) positioned, “mere vehicles that deliver instruction,” (p. 445) but we can strive to invent new and effective ways to “drive” them, benefitting students and their learning success.

## **Acknowledgments**

The author would like to acknowledge Dr. Kim Hellemans for implementing the examined strategies and providing valuable insight into her course; Dr. Cynthia Blodgett-Griffin for advising on graduate research topics; Dr. Anthony Marini for mentoring, guidance, and raising appreciation of multiple-choice tests capabilities; and Dr. Kevin Cheung for devising the original gamified test method in his courses and conversations on the topic. The author thanks Marko Dziedz for assistance in data processing.



## References

- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861–876. doi:10.1002/acp.1391
- Becker, K., & Nicholson, S. (2016). Gamification in the classroom: Old wine in new badges (Online document). Retrieved from [https://mruir.mtroyal.ca/xmlui/bitstream/handle/11205/268/LEGBOOK\\_TWO\\_FINALMANUSCRIPT\\_112414-CH3-edited.pdf?sequence=1](https://mruir.mtroyal.ca/xmlui/bitstream/handle/11205/268/LEGBOOK_TWO_FINALMANUSCRIPT_112414-CH3-edited.pdf?sequence=1)
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81–90. doi:10.1177/003172171009200119
- Blunt, J. R., & Karpicke, J. D. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*, 772. doi:10.1126/science.1199327
- Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2014). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition, XX*, 1–13. doi:10.3758/s13421-014-0462-6
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459. doi:10.3102/00346543053004445
- Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology, 106*, 58–68. doi:10.1037/a0033208
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effects of testing on long-term retention. *European Journal of Cognitive Psychology, 19*, 528–558. doi:10.1080/09541440601056620
- Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21*, 157–163. doi:10.1177/0963721412443552
- Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27*, 317–326. doi:10.1007/s10648-015-9309-3
- Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17*, 471–479. doi:10.1080/09658210802647009
- Kelley, P., & Watson, T. (2013). Making long-term memories in minutes: A spaced learning pattern from memory research in education. *Frontiers in Human Neuroscience, 7*, 589. doi:10.3389/fnhum.2013.00589

- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science, 23*, 1337–1344. doi:10.1177/0956797612443370
- Little, J., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition, 43*, 14–26. doi:10.3758/s13421-014-0452-8
- McDaniel, M.A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*(2), 200–206. doi:10.3758/BF03194052
- Nunes, L. D., & Karpicke, J. D. (2015). Retrieval-based learning: Research at the interface between cognitive science and education. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1-16). New York: John Wiley & Sons, Inc. doi:10.1002/9781118900772.etrds0289
- 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*, 437–447. doi:10.1016/j.jml.2009.01.004
- Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory, 22*(7), 784–802. doi:10.1080/09658211.2013.831454
- Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99–115. doi:10.1016/j.jml.2014.03.003
- Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. Sebastopol, CA: O'Reilly Media.

## Appendix A

### Moodle Test Structure

#### CHAPTER 7

CHAPTER 7 EASY - MCQ - How Do We Study the Brain's Structure and Functions? (10)

CHAPTER 7 MEDIUM - MCQ - How Do We Study the Brain's Structure and Functions? (68)

CHAPTER 7 HARD - MCQ - How Do We Study the Brain's Structure and Functions? (10)

CHAPTERS 7,8 - SHORT ANSWER (37)

#### CHAPTER 8

CHAPTER 8 EASY - MCQ - How Does the Nervous System Develop and Adapt? (10)

CHAPTER 8 MEDIUM - MCQ - How Does the Nervous System Develop and Adapt? (80)

CHAPTER 8 HARD - MCQ - How Does the Nervous System Develop and Adapt? (7)

CHAPTER 8 - Matching questions with images (3)

#### CHAPTER 9

CHAPTER 9 EASY - MCQ - How Do We Sense, Perceive, and See the World? (13)

CHAPTER 9 MEDIUM - MCQ - How Do We Sense, Perceive, and See the World? (73)

CHAPTER 9 HARD - MCQ - How Do We Sense, Perceive, and See the World? (11)

*Figure A1.* Moodle question bank with questions sorted based on cognitive difficulty level.

Page	Question ID	Difficulty	Question Text	Action
Page 1	1	Easy	Random (CHAPTER 3 EASY - MCQ - What Are the Functional Units of the Nervous System?)	(See questions)
	2	Easy	Random (CHAPTER 3 EASY - MCQ - What Are the Functional Units of the Nervous System?)	(See questions)
	3	Easy	Random (CHAPTER 4 EASY - MCQ - How Do Neurons Use Electrical Signals to Transmit Information?)	(See questions)
	4	Easy	Random (CHAPTER 4 EASY - MCQ - How Do Neurons Use Electrical Signals to Transmit Information?)	(See questions)
Page 2	5	Medium	Random (CHAPTER 3 MEDIUM - MCQ - What Are the Functional Units of the Nervous System?)	(See questions)
	6	Medium	Random (CHAPTER 3 MEDIUM - MCQ - What Are the Functional Units of the Nervous System?)	(See questions)
	7	Medium	Random (CHAPTER 4 MEDIUM - MCQ - How Do Neurons Use Electrical Signals to Transmit Information? )	(See questions)
	8	Medium	Random (CHAPTER 4 MEDIUM - MCQ - How Do Neurons Use Electrical Signals to Transmit Information? )	(See questions)
Page 3	9	Hard	Random (CHAPTER 3 HARD - MCQ)	(See questions)
	10	Hard	Random (CHAPTER 4 HARD - MCQ - How Do Neurons Use Electrical Signals to Transmit Information?)	(See questions)

*Figure A2.* Typical structure of a level 1 test in Moodle consisting of 10 questions.

## Appendix B

### Microsoft Excel Regression Outputs

Table A1

*Regression Output for Independent Variable MTL1A and Dependent Variable FE (%) for Sample Size N=200 and Confidence Level of 95%*

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.611981537							
R Square	0.374521402							
Adjusted R Square	0.371362419							
Standard Error	0.133254647							
Observations	200							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	2.105203581	2.105204	118.5576	6.13478E-22			
Residual	198	3.515846563	0.017757					
Total	199	5.621050143						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.498300776	0.017904963	27.83032	2.34E-70	0.462991876	0.533609676	0.462991876	0.533609676
X Variable 1	0.016511676	0.001516444	10.88842	6.13E-22	0.013521221	0.01950213	0.013521221	0.01950213

Table A2

*Regression Output for Independent Variable ME(%) and Dependent Variable FE(%) for Sample Size N=200 and Confidence Level of 95%*

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.810400479							
R Square	0.656748936							
Adjusted R Square	0.655015345							
Standard Error	0.098714757							
Observations	200							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	3.6916187	3.6916187	378.8372503	7.3928E-48			
Residual	198	1.929431443	0.009744603					
Total	199	5.621050143						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	0.07919749	0.030849802	2.567196075	0.010990327	0.018361142	0.140033839	0.018361142	0.140033839
X Variable 1	0.801340649	0.041170945	19.46374194	7.3928E-48	0.720150826	0.882530472	0.720150826	0.882530472

