7-2020

EVALUATING THE EFFECTIVENESS OF PROBABILISTIC REINFORCEMENT, AS A GAMIFICATION ELEMENT, IN A COLLEGE CLASS SETTING

Tiffanie Weeden
tweeden@nmu.edu

EVALUATING THE EFFECTIVENESS OF PROBABILISTIC REINFORCEMENT, AS A GAMIFICATION ELEMENT, IN A COLLEGE CLASS SETTING

By

Tiffanie R. Weeden

THESIS

Submitted to

Northern Michigan University

In partial fulfillments of the requirements

For the degree of

MASTER OF SCIENCE IN APPLIED BEHAVIOR ANALYSIS

Office of Graduate Education and Research

July 2020

SIGNATURE APPROVAL FORM

Evaluating The Effectiveness Of Probabilistic Reinforcement, As A Gamification Element, In A College Class Setting

This thesis by Tiffanie Weeden is recommended for approval by the student's Thesis Committee and Department Head in the Department of Psychology and by the Dean of Graduate Education and Research.

Jacob Daar — 07/06/2020

Committee Chair: Dr. Jacob Daar — Date

7/7/2020

First Reader: Ashley Shayter, M.S. — Date

Joe Lubig — 7-6-2020

Second Reader: Dr. Joseph Lubig — Date

7/9/20

Department Head: Dr. Adam Prus — Date

Lisa S. Eckert — 8/10/20

Dr. Lisa S. Eckert — Date

Dean of Graduate Education

ABSTRACT

EVALUATING THE EFFECTIVENESS OF PROBABILISTIC REINFORCEMENT, AS A GAMIFICATION ELEMENT, IN A COLLEGE CLASS SETTING

By

Tiffanie R. Weeden

Gamification, or the use of game mechanics in non-game activities, has potential utility in enhancing course materials and help to motivate students. The present study sought to determine whether a gamified quiz application utilizing a probabilistic reinforcement schedule, a common game mechanic in which magnitude of reinforcement is randomized, would increase interaction with course material and subsequently increase exam scores when compared to a traditional fixed ratio point scoring system in a college class setting. An undergraduate class of 40 students were randomly split into two groups (green = 18, gold = 22). After baseline data was probed, the groups underwent a series of two phases, either A or B. During phase A the green group started the experiment in the gamification quiz condition while the gold group was in the control condition. In phase B the green group started in the control condition and the gold group started in the gamification condition. Each phase consisted of three quizzes, followed by an exam. After an exam, the groups switch conditions, conducting three full phases (A-B-A). After an adjustment taking into account Exam 1 scores it was determined that the participants did not show a statistically significant difference in increasing interaction in course material between the interventions, $F(1,37)=.280$, $p=.600$ ($\alpha < 0.05$) or an increase in exam scores $F(1,37)=2.231$, $p=.144$ ($\alpha < 0.05$).

ACKNOWLEDGEMENTS

The author would like to thank her thesis director and advisor, Dr. Jacob Daar, PhD, BCBA-D, for his support and guidance throughout her studies, without his leadership and knowledge, this thesis would not have been possible; Professor Ashley Shayter, MS, LBA, BCBA, CBIS, and Dr. Joseph Lubig, EdD, for sitting on her thesis committee and for providing assistance in making this thesis; Kimber Thompson, MS, BCaBA, for enlightening her with her knowledge of the Learning Management System used for the delivery of the gamified quizzes. Additionally, the author would like to thank her fellow cohort members for their constant words of encouragement throughout this process.

This thesis follows the format prescribed by the Publication Manual of the American Psychological Association, 7th edition and the Department of Psychological Science.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

It is common for some first-year students not to succeed in college. First year students are unaware and unprepared for the different academic challenges that are expected to be met by universities (Fleming, 2002). It is required by most universities to have a C average to pass a course. Universities recommend that a student should spend two hours of out of class time studying for every credit of enrollment to achieve a passing grade. This means that if a student takes 16 credits, they should spend a minimum of 32 hours studying. With that, it is not shocking that first-year college students do not have the skills to study, to that magnitude, in their repertoire (Michael, 1991). If students learn study strategies early on, they can avoid the burden of trial and error. This will help students capitalize on their chance for achievement and continuing their college career (Fleming, 2002).

It has been shown that student performance is enhanced when teachers create learning opportunities through active engagement with class material such as required homework assignments. Only a few studies have been conducted examining homework assignments, specifically with college students, and with most of those studies using self-reported measures, more research is needed on the topic (Ryan & Hemmes, 2005). Ryan and Hemmes (2005) studied the effect of two different homework submission contingencies, points awarded, and no points awarded, on the probability of homework submission and performance on related quizzes. The effects were analyzed for individual participants and evaluated by the performance of group averages. The researchers randomly divided the class in half at the beginning of the semester

and used an alternating treatment design, making sure that each group was exposed to both conditions equally during the experiment. The experiment demonstrated social significance by showing that there was an increase in homework submission and high quiz grades during the points condition versus the no-points condition. The increased quiz grades boosted the students' overall grade averages by 10% amounting to a full letter grade increase. The authors discovered that there was experimental control over the group average performance data and not so much by individual participants' data (Ryan & Hemmes, 2005). Rehfeldt et al., (2010), then replicated this study with graduate students and produced the same results as Ryan and Hemmes (2005). The participants reported that homework assignments were an excellent studying tool to prepare for their quizzes. Students continued to do the homework during the no-point conditions and reported not submitting the assignments because they did not meet all the requirements. These studies did not use a continuous measurement to compare the amount of time engaged with homework/studying and improved performance with class material which amounted as a major limitation in the experiments' design (Rehfeldt, Walker, Garcia, Lovett & Filipiak, 2010).

Students have also shown to increase academically through atypical grading scales. Research has suggested that curving students' grades towards 100% has rewarding effects. Solley (2011) conducted a study with elementary students to strengthen their reading by using a curved method of grading. The teacher used an accelerated reader program that tested students' comprehension of books at their current reading level. If a student received an 80% or better on their quiz, they were rewarded with 100% as a grade and then moved to the next reading level. If a student did not receive an 80%, they were rewarded with the grade they achieved on their quiz and continued to stay at their current reading level. The curved method of grading was compared with another classroom that used the traditional grading method. The class that used the

accelerated reader program produced statistically significant results of progress in reading compared to the other classroom that did not use the accelerated reader program (Solley, 2011).

A study done by Jakupcak and colleagues (1996) included students with typical academic abilities and students with learning disabilities. The researchers assessed inquiry learning, instructional strategies, student-oriented curriculum, and different learning styles, to increase quality and quantity of student achievement. By having all types of student learners in the same experimental group, the teacher was able to hold all students to the same mastery criterion of 80% or better. This study demonstrated that the grades from the test and homework assignments increased approximately 10% by expecting all students to obtain at least an 80% mastery criterion (Jakupchak, Rushton, Jakupcak,, & Lundt, 1996).

Not only is an 80% mastery criterion used within the educational setting, but it is also used in applied settings for skill acquisition and maintenance. Two studies compared different effects of mastery criteria (Richling, Williams, & Carr, 2019; Fuller & Fienup, 2018). Richling and colleagues (2019) first conducted a survey to determine what the most common mastery criteria was being used among clinicians. The survey yielded that the most commonly used criterion was, in fact, 80%. Therefore, the researchers sought to examine the effects of this mastery criterion compared to other mastery criteria such as 60%, 90%, and 100%. The researchers found that the 80% criterion required about seven sessions and produced more presentation of reinforcement than the 60% criterion. The most effective mastery criterion demonstrated in this study was 100% -- as 100% requires absence of any error in performance and was achieved in an average of 14 sessions. No error in performance can be seen as unrealistic as individuals of all skill levels and ages encounter errors at times. In follow up sessions, between 80% criterion (average of seven sessions) and 90% criterion (average of 13

sessions), the outcome of performance is nearly identical. As 100% criterion is unrealistic and with 90% criterion resulting in identical performance as 80% criterion, the participants' time can be maximized when mastery criterion is set to 80%. Fuller and Fienup (2018) discovered that higher levels of maintenance can be predicted by mastering skills at higher performance levels (Fuller & Fienup, 2018). Their results yielded the same as the previous study in regard to participants' follow up performances of 80% and 90% criteria.

Even with the research on homework growing, Trautwein and Köller (2003) wonder if "homework assignments can foster achievement" and to "what extent homework behavior improves grades" (Trautwein & Köller, 2003, p. 134). The authors conducted their own literature review, discovering that homework is a form of studying and that studying is a key factor in school achievement (Trautwein & Köller, 2003). Trautwein and Koller, (2003) found that the amount of time individuals spent on homework determined their level of achievement. However, self-regulating learning factors are at play due to the students' responsibility for regulating their own behavior in regards doing and/or completing homework. A component of self-regulation is motivation and with high motivation an individual may engage in "good homework behavior". Intrinsic and cost are values associated with motivation. These values contribute to "good homework behavior" in the form of time spent on homework. Intrinsic value assesses the enjoyment an individual receives from engaging in an activity. Some students may already find homework to be intrinsically motivating, whereas, most students do not find homework to hold a strong intrinsic value due to the high cost value. Cost value is when homework and studying get in the way of hanging out with friends and doing other highly intrinsic motivating activities. Most individuals want to encounter preferred items/activities without a large delay in time. Both

values contribute to time spent on homework and should be considered in further experiments on homework and achievement (Trautwein & Köller, 2003).

Motivation is a leading factor in time spent on studying. A different method examining environmental contingencies needs to be put in place. Enhanced studying and homework is crucial, because the majority of learning in college is done through studying (Michael, 1991). Neef et al., (2011) expanded on their 2007 research pertaining to the implementation of study sessions in the format of a game, such as Jeopardy® or Who Wants to Be a Millionaire® (Neef, Perrin, Haberlin & Rodrigues, 2011; Neef et al., 2007). The expanded research consisted of study questions created by the students in comparison to the original study which consisted of instructor generated questions. Both studies used an alternating treatment design which counterbalanced across two class sections. The participants scored higher on quizzes during the game conditions with average quiz scores at 84.4% (range, 68.3% to 96.1%), with the non-game condition producing average quiz scores of was 72.1% (range, 53.1% to 88.6%). This study also showed social significance with improved quiz scores during the game condition by an entire letter grade equivalent. The researchers noted that the competitive aspect of the game was not the motivating factor during the experiment as participants reported “feeling bad” for other students who didn’t receive the bonus points during the study. Further research should be conducted with different game features to add engagement with course material (Neef et al., 2011).

Based on previous research, it is essential that instructors motivate their classes as traditional teaching methods seem to be ineffective. The use of games built inside the structure of a course can help students acquire new study skills. This may lead students to be more motivated for classwork in comparison to a more traditional style (Silva, Rodrigues & Leal, 2019). Adding games to a classroom is known as gamification, defined as “…an integration of game elements

and game thinking in activities that are not games" (Kiryakova, Angelova & Yordanova, 2014, p. 1). Kiryakova and colleagues (2014) list six fundamental features that gamification may use, users must be participants (such as students), there needs to be some sort of challenges or task, accumulation of points, levels based on those points, rewards, and ranking of users. Games are a simple and familiar mechanism to the everyday person as games are always within reach such as apps on cellphones and laptops. Gamification can affect individual behaviors, which leads to improved commitment, motivation, a gain in new knowledge and skills for success. Gamification in an educational setting is most successful when the following are implemented; feasibility, multiple attempts, levels of difficulty, and various learning paths (Kiryakova, Angelova & Yordanova, 2014).

Yildirim (2017), conducted a study with 97 undergraduate sophomores from the department of mathematics education at a university in south Turkey. The control and experimental groups were taught using a blended procedure of 60% of the class was conducted as a traditional setting and 40% of the class was taught through distance education. Both groups used the same Moodle platform to access the material needed for the distance education portion. The experimental group differed from the control group by the added gamification element. The experimental participants were told that the course, overall, was just a game and by finishing the game they would receive a reward at the end. The experimental group did not receive their online material in bulk but instead had to complete the first section to move on to the second section and so forth. The experimental group had optional activities throughout their coursework compared to the control group. The results of this study showed that the experimental group had a mean test score of 73.44 and the control group had a mean test score of a 68.06. Although theses score are the difference of 5.38 points, these results are socially significant. A class grade

of 68 could be considered as failing and a grade of 73 could be considered as passing per typical school standards. The author considered the large class size and distance learning to be a limitation to this study. These factors made it hard to monitor the use of outside sources that could have affected the experimental condition of the two groups, causing a possible confounding variable in the data. With gamification becoming a new trend within all levels of the educational setting, Silva and colleagues (2019) demanded more empirical research is needed to validate the effectiveness of gamification (Yildirim, 2017).

Bharamgoudar (2018) compiled multiple studies that demonstrated the benefits of the addition of gamification elements to the educational setting. One of the studies mentioned was conducted by Lamb et al., (2017) that incorporated social media and a point reward system in the delivery of homework questions. These gamification elements increased participants examination ranking by about 14%. The students who did not participate in the experiment had their percentile ranking fall approximately 10%. Lamb et al, (2017) demonstrated that not integrating gamification into educational settings produces negative effects on students. (Lamb, DiFiori, Jayaraman, Shames, & Feeney, 2017). Chen et al., (2017) conducted a study that implemented interactivity, instantaneous feedback, and short time limits. With the addition, students were able to analyze radiographs more quickly and accurately compared to students that did not participate (Chen, Roth, Galperin-Aizenberg, Ruutiainen, Gefter, & Cook, 2017). Another one of the studies reviewed by Bharamgoudar (2018), piloted by El Tantawi et al., (2018), implemented gamification elements in an undergraduate classroom and revealed an overall grade increase of 44% (El Tantawi, Sadaf, & AlHumaid, 2018). Gamification elements motivate students to work harder and achieve more in their coursework whether they are competing against themselves or each other (Bharamgoudar, 2018).

Another element in gamification is random point assignment. Random point assignment is also known as probabilistic reinforcement or as variable ratio schedule of reinforcement. Random point assignment has shown to increase appropriate responding behavior in animals, children and adults (Martens, Ardoin, Hilt, Lannie, Panahon, & Wolfe, 2002; Sherman, & Thomas, 1968; Rost, 2018). Martens and colleagues (2002) implemented a lottery day, with probabilistic reinforcement in a fourth-grade classroom to increase the appropriate behavior of completing more math problems. The lottery system implemented consisted of children drawing four colored chips (red or white) for the chance to exchange "easy problem slips" for a prize. The procedures stated that each student had to do a certain number of hard problems before they could complete an "easy problem slip". If a student pulled a white chip out of the bag, then they were not allowed to exchange any slips that day. If a student pulled a red chip, they could exchange a maximum of two slips for a prize. Each lottery day the ratio of red to white chips were lowered without the knowledge of the students. Every lottery day experienced different ratio of colored chips. The study resulted in higher completion of complex math problems on the lottery days compared to their baseline (Martens et al., 2002).

One component of gamification is the use of bonus games. Bonus games are a relatively effortless way to incorporate the element of probabilistic/variable reinforcement schedules into the classroom using point delivery. Bonus games are defined as simple games or features that are embedded within a bigger system or game (Harrigan, Collins, Dixon, & Fugelsang, 2010). A bonus game can appear randomly throughout an interaction or sometimes in a more predictable way, such as in between levels or attempts. During a bonus game, an individual is typically asked to pick an object from an array. Each object has a predetermined outcome that is completely random to the individual (Harrigan et al., 2010). Each outcome is commonly in the

form of a positive experience, meaning the individual always "wins" (earns points). When the individual engages in choice making the individual is given the false feeling of having the skill of selecting the "correct" object. Bonus games keep the individual motivated within the task at hand by altering the engagement for a moment of time before returning to the original task (Harrigan et al., 2010).

Saudargas et al., (1977) compared amount of homework assignments completed during a fixed and variable- time schedule in a third-grade classroom. During the fixed-time schedule students were given a packet of assignments on Fridays to be completed at home and turned in by the following Friday. The students were informed that they should do a minimum of four assignments each day. During the variable-time schedule, students were given their typical weeks' worth of assignments to be done at home. However, this time assignments were not due by the following Friday. Instead, seven to nine students were chosen at random throughout the week to turn in specific assignments. This study demonstrated higher percentage of homework assignments completed, both required and voluntary during the variable schedule. The students improved academically with the variable schedule during the experiment. During the follow up the teacher reported, "The variable schedule produced a consistent responding rate even after the assignments for the week were completed" (Saudargas, Madsen, & Scott, 1977, p. 678).

Another study examined student on-task behavior with a fixed and variable schedule of reinforcement. Houten and Nau (1980) assessed five deaf students' attentiveness and decreases in disruptive behavior while in class. During the fixed ratio schedule, students would earn checkmarks for the appropriate targeted behavior. Once the students earned twelve checkmarks, they could exchange their earnings to draw a prize. During the variable schedule the students did not "collect" or visualize their earned checkmark, however, the teacher kept track of their

behaviors. For about every twelve targeted behaviors the students engaged in, they were given the opportunity to select a prize. The study demonstrated nearly 100% rates of attentiveness during the variable ratio schedule and demonstrated increased percentage of completed class work (Houten, & Nau, 1980).

McDougall and Granby's (1996) experiment explored students' preparation, participation, and overall recall of class material in the classroom using a variable schedule of reinforcement. The experiment split an undergraduate statistic class into two groups. The experimental group was told that they were going to be participating in random oral questioning, which meant they would not know when the teacher would call on them to answer a question about class material. Whereas, the control group could answer questions voluntarily. After this experiment ended, the results determined that the students in the variable schedule group demonstrated better recall of class material, completed more class work, and reported more time spent preparing for the class overall (McDougall, & Granby, 1996).

Purpose and Rationale

It is important to provide college students with an enriched learning environment to strengthen their performance in their classes. According to Yildirim (2017) increasing a student's motivation and engagement with class materials, such as quizzes and assignments, are correlated with higher achievement in their classes. Based on previous research, integrating gamification into higher education is favorable. However, more research on the effectiveness of gamification in enhancing student's motivation and engagement is needed (Silva, Rodrigues, & Leal, 2019).

A common mechanism in game design is the application of probabilistic reward or outcomes for engagement. In behavior analytic research, probabilistic reward is an element

sometimes used in classroom settings. Previous studies have demonstrated higher amounts of completed class assignments during a probabilistic reward condition compared to baseline or control conditions in students (Houten and Nau, 1980; Martens et al., 2002; Saudergas et al., 1977). Research suggests there is higher levels of achievement in a classroom setting through the implementation of assignments and probabilistic reward or outcome compared to a fixed ratio schedule of reinforcement. With no known research evaluating the implementation of gamification on quizzes, probabilistic reinforcement schedules should be examined on college-aged individuals through the uses bonus games applied during assigned quizzes. The purpose of the current study is to evaluate the effectiveness of a probabilistic reinforcement schedule, as a gamification element, in increasing interaction with course material and subsequently increasing exam scores in a college classroom setting.

Methods

Participants, Setting and Materials

This study included 40 students enrolled in an introductory undergraduate course on behavior analytic principles and concepts at an upper Midwest university in the United States of America. The participants (f = 38) were split into one of two groups, green or gold. All students were enrolled in the same class section that met twice a week for one hour and forty minutes. From January 13th through March 16th the class was conducted in person, except on dates the university was closed, due to scheduled breaks or snow days. On March 16th, the class structure was switched to all online delivery due to COVID-19.

All participants were assigned a working university laptop with internet connection. Students had access to the university's open source learning management system (LMS), where

they accessed their quizzes, exams, and class material. Each week one quiz was assigned as homework and had to be taken outside of the classroom.

Experimental Design and Dependent Variables

A counter control A-B-A design was used to evaluate the effectiveness of a probabilistic reinforcement schedule, as a gamification element on the average number of quiz attempts and average exam scores. Average attempts on quizzes and average exam scores were collected for each group. The average number of quiz attempts was defined as the summed number of quiz attempts that each participant submitted for the same quiz and then divided by the total number of participants within the group. The average number of quiz attempts was measured to compare the amount of interaction with course material between the two groups during each phase. Average exams scores were defined as the sum of participants' exam scores within one group, on one exam, and divided by the number of scores for that exam. Average exam scores were measured to evaluate if there was an increased in achievement under the two conditions (control or gamification).

The participants were randomly assigned to one of two groups, green group or gold group. All participants took Exam 1 before the study began as a baseline for comprehension of the class material. After Exam 1, The green group started the semester with the gamified condition quizzes, the gold group started the semester with the control condition quizzes (Phase A). The condition stayed the same for both groups until Exam 2. After Exam 2, the groups switched conditions. The green group was now in the control condition and the gold group was now in the gamified condition (Phase B). Phase B lasted through Exam 3. After Exam 3, the two groups switched back into Phase A, until the end of Exam 4, the end of the experiment.

Procedure

Before the study began the participants were informed that they would be assigned homework in the form of weekly quizzes related to that week's material. The students would access the quizzes through the LMS. The participants were informed that they may attempt the quizzes as many times as they would like while the quiz was open. The instructor also informed the participants' that there were two different quiz formats and each quiz had the same bank of questions and all participants were required to answer the same number of questions for both types of the quizzes. The quiz questions were in the possible format of multiple choice, true/false or matching. All quizzes had predetermined correct answers to each question and were graded electronically immediately after the participant submitted the quiz. Quiz grades were awarded immediately upon completion of the quiz.

Control Condition

During the control conditions, the participants took a 5-10 question(s) quiz each week. Each question answered correctly was worth 1-2 point(s) and any question answered incorrectly was worth 0 points, with the opportunity to earn up to 10 points. The quizzes were assigned to the participants to access them on the LMS and had five days to interact with the quiz as many times as they would like. The attempt with the highest number of points was submitted for their class grade. Every quiz contained a different arrangement of quiz questions pulled from a question bank to decrease the likelihood that students would encounter the same questions, arrangement of questions, or arrangement of question responses on each attempt.

Gamification Condition

During the gamification conditions, the quiz was broken up into two parts. The first part of the quiz consisted of the participants answering the quiz questions. The questions were pulled from the same question bank as the control condition, along with the quizzes containing the same number of questions, and worth the same amount of points. The participants must have correctly answered 80% of the questions, or more, to move on to part two. If the students did not obtain at least an 80%, they may reattempt the questions, until they did, if the quiz was still open to be taken. Once the participant reached the 80% or higher, part two was unlocked and visible to the participant. Part two was an embedded bonus game that asked the participants to choose one of four options from an array. Each option had a random percentage attached to them, either 80%, 90%, or 100%, that the participant was unaware of. The option the participant chose determined the amount of points awarded for their overall quiz score. Participants' could interact with the entire quiz (successfully completing part one and two) up to five times, if the quiz was still open. The highest points awarded during one quiz attempt, was submitted for their class grade.

Results

Figure 1 shows baseline exam scores, the average number of quiz attempts, and the average exam scores for the green group and gold group during the control and gamification conditions. Exam 1 depicted the two groups exam averages before the conditioned were introduced to the participants. The green group had a mean exam score of 64.77 (SD = 13.26, range 42.04- 85.71). The gold group had a mean exam score of 74.15 (SD = 13.84, range 35.92-93.88). Exam 1 was used as a covariate when analyzing Exams 2, 3, and 4 mean scores and the number of attempts for both groups, $F(1,37) = 11.51$, $p = .002$, ($\alpha < 0.05$).

During phase A, the gold group was under the control conditions and the green group is under the gamification conditions. During phase B, the green group is under the control condition and the gold group is under the gamification conditions. The green group attempted the quizzes during the gamification condition for a total mean of 5 times (SD = 4.97, range 0-23) and attempted the quizzes during the control condition for a total mean of 3.63 times (SD = 3.39, range 0-18). The gold group attempted the quizzes during the gamification condition for a total mean of 6 times (SD = 5.30, range 0-20) and attempted the quizzes during the control condition for a total mean of 3.57 times (SD = 3.59, range 0-25). An ANCOVA was ran to determine the effect of control and probabilistic reinforcement gamified quizzes on the number of quiz attempts to compute the overall interaction with course material. After accounting for Exam 1 scores, there was not a statistically significant difference in the overall interaction with course material between the conditions, $F(1,37) = .280$, $p = .600$ ($\alpha < 0.05$).

The green groups total mean score on the exams that followed the gamification condition was a 59.47 (SD = 27.88, range 0-99.17), and the gold group total mean scores on the exam that followed the gamification condition was a 76.08 (SD = 25.94, range 0-95.38). The green groups total mean on the control condition exams, was a score of 68.32 (SD = 27.07, range 0-93.85) and the gold group total mean on the control condition exams, was a score of 78.76 (SD= 16.47, range 50-100). An ANCOVA was ran to determine the effect of control and probabilistic reinforcement gamified quizzes on exam scores to compute the overall in class achievement. After accounting for Exam 1 scores, there was not a statistically significant difference in the overall in class achievement between the conditions, $F(1,37)=2.23$, $p=.144$ ($\alpha < 0.05$).

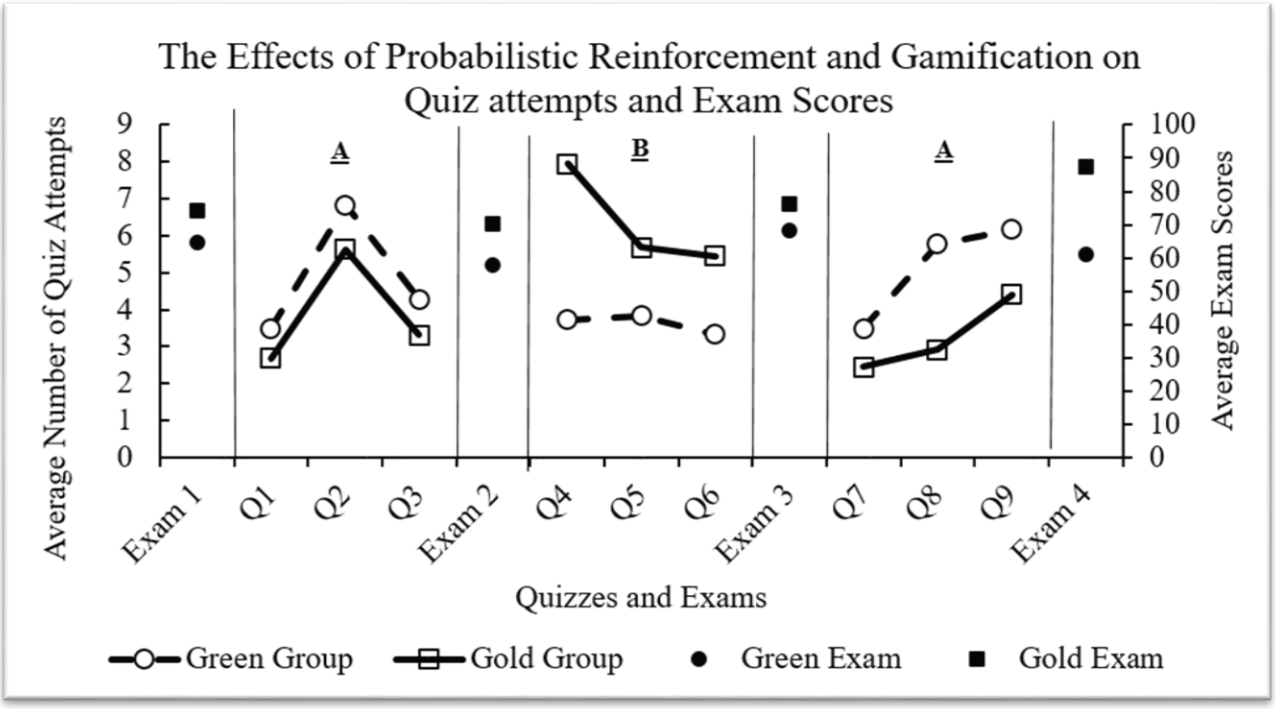

Figure 1. Shows the average number of quiz attempts and average exam score exhibited by both the green and gold group. Exam 1 was used as baseline. During phase A, the green group was under the gamification condition and the gold group was under the control condition. During phase B, the gold group was under the gamification condition and the green group was under the control condition.

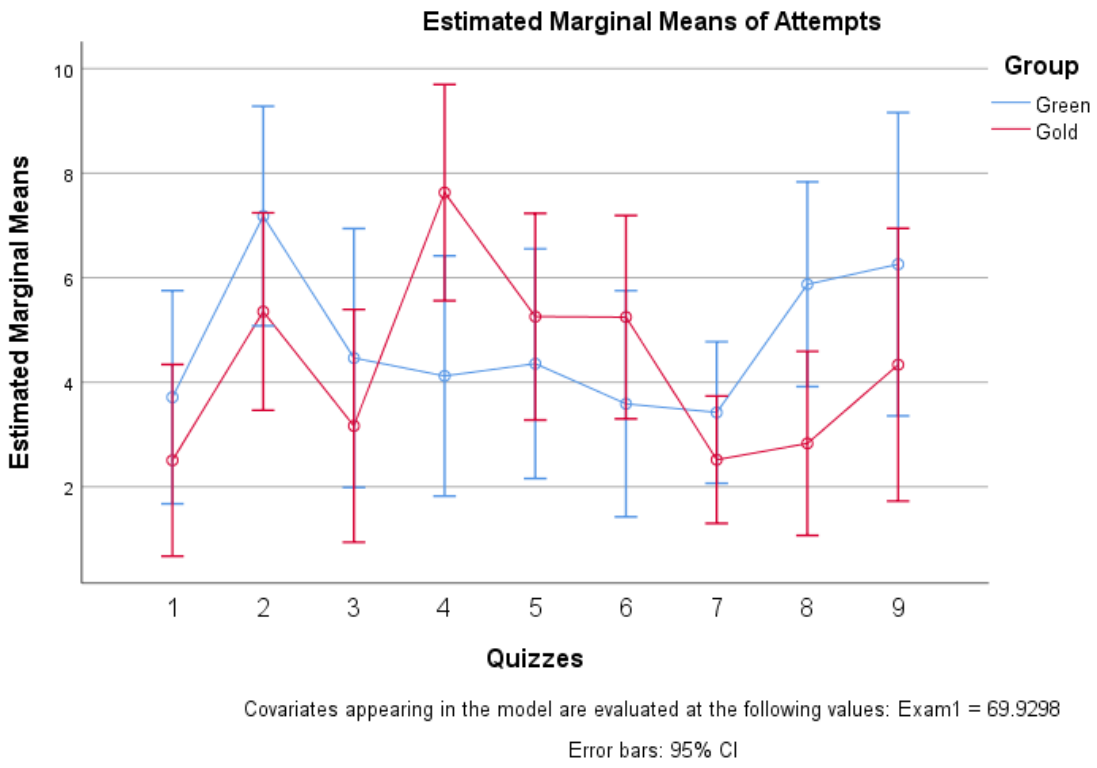

Figure 2. Shows the estimated marginal means as a function of the assigned quizzes and groups in the participants' mean number of quiz attempts.

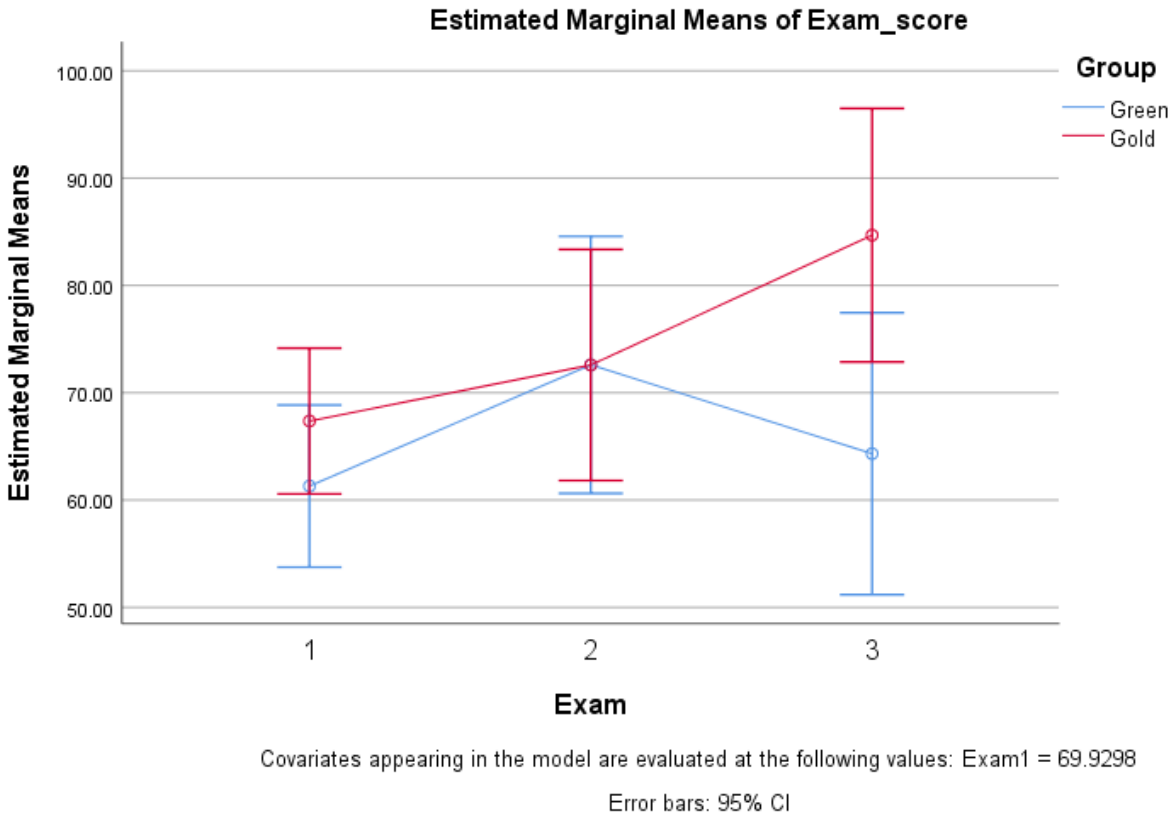

Figure 3. Shows the estimated marginal means as a function of the exams and groups in the participants' mean exam scores. It is to be noted that "Exam 1" listed on the graph represents the scores from Exam 2, as Exam one was baseline before conditioned were introduced. Therefore, "Exam 2" is representing Exam 3 scores and "Exam 3" is representing exam 4 scores.

Discussion

The purpose of the current study was to demonstrate that by incorporating gamification elements, such as probabilistic reinforcement, into a college class setting, would establish an improvement in class participation. If students' participation with class assignments increased, it is more likely that the students will demonstrate a better understanding of the material and perform better on exams. Adding gamification into a typical undergraduate class increases students' motivation, which commonly correlates with higher achievement in the class (Yildirim, 2017). Pervious literature suggests an increase in appropriate behavior in adults and children with the use probabilistic reinforcement (Martens et al., 2002). This is due to the variable ratio schedule involved with probabilistic reinforcement. McDougall and Granby (1996), found that with the use of variable schedules, students were more prepared for class overall i.e. completed more class work and had better recall of class material. Previous literature on gamification has been discovered to increase undergraduate grades by as little as one letter grade up to 44% (Bharamgoudar, 2018; Yildirim, 2017).

The results of the current study suggest that implementing gamification in the form of probabilistic reinforcement in undergraduate students' quizzes through the delivery of bonus games, had no statistically significant impact on students' interaction and achievement in class. These results are not consistent with the finding of Martens et al., (2002), their study resulted in more work done by the students during the probabilistic reinforcement condition when compared to the control condition or baseline. The current experiment's results were also not consistent with the findings of Jakupcak et al., (1996) and Yildirim (2017) regarding exam scores being positively affected. Exam scores in the present study stayed consistent to the baseline exam. The gold group exams scores constantly stayed higher than the green group's exams regardless of the

conditions, and achievement was best predicted by baseline exam scores. It should be noted that the current studies used participants that were enrolled in a freshman level, introductory course. As mention in the literature review, the participants may not have developed study skills. While the current study may or may not have prepared them in shaping new study skills for future classes. This study could be reexamined with participants who already demonstrate strong study skills.

This study resulted in four notable limitations that should be addressed. The first limitation in the study was the lack of counterbalance between the groups. At the beginning of the study, before the baseline exam was delivered, the class was randomly dived into two groups by the LMS system. It is revealed that the gold group had higher exam scores than the green group from the baseline through each condition, ranging an average of 8-26%. The study could have benefited from a pre-test covering the class material to determine if students had any prior knowledge of the subject. After the pre-test was graded, students could have been equally split into the two groups based on their pre-test performance. The delivery of a pre-test would ensure that one group would not be populated with higher achieving students that may already be familiar with the material being taught throughout the semester. By not implementing a pre-test and counterbalancing the two groups could be the reason as to why there was no variation in the average exam scores throughout the study. While this limitation was present, the use of Exam 1 as a covariate within the data analysis served to reduce the impact of the unbalanced but random assignment.

The second limitation would be the counter control A-B-A design used. There is a possibility that sequence effects could have occurred due to the lack of a second Phase B. In the present study the green group was able to experience the gamification condition followed by the control

condition and vice versa, whereas, the gold group only experienced the control condition followed by the gamification condition. Without the second Phase B, the gold group never experienced the gamification condition before the control condition. Expansion on this experiment should be conducted to determine whether equal exposure across all phases to both groups would yield statistically significant results.

The third limitation in the study was the university's LMS used to deliver the gamified quizzes. The programing behind the LMS did not allow for the students to interact with both phases in the gamification condition on the same page. Meaning, the students would open the internet page that contained phase one, then take their quiz. Once they have completed phase one, the LMS would bring them back the beginning of the phase one quiz. If the student scored an 80% or better, a new link for the part two bonus game would be in the bottom right corner, or the students could access the part two bonus game by returning to the home page. From this set up there is a possibility that students may not have noticed the new internet link in the corner or may not have gone back to the home page after attempting the quiz each time. A better setup to reduce this problem would be for the LMS to immediately open part two of the quiz once the student scored 80% or better. With multiple internet pages needing to be accessed to complete both parts of the gamification condition, we were only able to set up five attempts to complete the entry of the quiz. This may have caused a ceiling effect on the participant when interacting with the gamification condition. This may have affected the results from demonstrating higher number of average quiz attempts. It also can be noted that the response effort needed to interact with the gamified quizzes was much higher than the efforts needed to interact with the control condition. The control condition only had one part of quiz that could be continuously accessed immediately after finishing an attempt, with no limit of quizzes or multiple internet pages.

The fourth and last limitation that should be mentioned was the impact of the global pandemic of COVID-19. At the start of the pandemic school was canceled on March 11th- 13th. During this time the university announced that classes would resume on March 16th in an online only format. As the weeks to come the university kept pushing back plans for when classes would be resumed in person. Eventually it was determined that the rest of the semester would finish out in the online format. As soon as the class switch to the online delivery some participants stopped interacting with the class material. As the weeks went on so do the lack of the participants' interaction, even leading to a handful of students withdrawing the class. This may be another factor why the average exam scores did not differentiate among groups. The increasing lack of participation and withdrawal of participants was not evenly distributed across the two groups. There were less participants and more zeros recorded in the green group by the end of the experiment compared to the gold group. It should be noted that any students that withdrew from the class at any point throughout of the semester was not include in the study. Additionally, all students were given the opportunity to select a Pass/Fail grade instead of the normal percentage-based letter grade in the course, i.e. many students may have recognized that a lower exam score would result in the same recorded passing grade for the course, thus resulting in a lack of engagement with the quizzes and exam material.

Future studies should include a more diverse sample size and improved gamification elements. Including a more diverse sample size could be done in several ways such as gender, race, education level, subject matter, and geographical location. The current study only had two male participants and could have been influenced differently if the two groups were equally made up of both male and female participants. With the study taking place in a rural area that is mainly populated with Caucasian individuals, the results could have also been influenced by the

culture. In the future the study could compare these findings with results collected from a more Urban area, along with using a more ethnically diverse sample. Lastly, future studies could also compare these findings with studies conducted at various educational levels and subject matters such as graduate students, grade school students or even undergraduate students in different fields of study.

The current study lost the "fun" or motivating factor in the gamification condition, due to the high response effort needed to access the bonus game. Future studies could also expand this current experiment delivering probabilistic reinforcement schedules in a different way. Another method of probabilistic reinforcement could be implemented with quizzes such as the incorporation of a lottery system. Martens et al., (2002) demonstrated higher levels of student engagement with class material when a lottery system was added into a grade school math class. A future study could also include a drawing for a chance to win preferred items such as extra credit points towards the class. Students would earn entries into the drawing by achieving a particular score on a quiz. The suggestion of a drawing is a simpler way to integrate gamification rather than implementing the use of a learning management system. A lottery system possibly motivate the participants to study their class material so that their chances of being entered into the drawing are higher. The addition of these subtle variables could influence the outcome and produce significant results.

References

Bharamgoudar, R. (2018). Gamification. *The Clinical Teacher*, *15*(3), 268–269.

Fleming, V. M. (2002). Improving students' exam performance by introducing study strategies and goal setting. *Teaching of Psychology*, *29*(2), 115-119.

Fuller, J. L., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion level: Effects on response maintenance. *Behavior Analysis in Practice*, *11*(1), 1-8.

Harrigan, K. A., Collins, K., Dixon, M. J., & Fugelsang, J. (2010). Addictive gameplay: What casual game designers can learn from slot machine research. *Proceedings of the international academic conference on the future of game design and technology* 127-133.

Houten, R. V., & Nau, P. A. (1980). A comparison of the effects of fixed and variable ratio schedules of reinforcement on the behavior of deaf children. *Journal of Applied Behavior Analysis*, *13*(1), 13-21.

Kiryakova, G., Angelova, N., & Yordanova, L. (2014). Gamification in education. Proceedings of 9th International Balkan Education and Science Conference.

Jakupcak, J., Rushton, R., Jakupcak, M., & Lundt, J. (1996). Inclusive education. *The Science Teacher*, *63*(5), 40.

Latham, G. P., & Dossett, D. L. (1978). Designing incentive plans for unionized employees: A comparison of continuous and variable ratio reinforcement schedules 1. *Personnel Psychology*, *31*(1), 47-61.

Martens, B. K., Ardoin, S. P., Hilt, A. M., Lannie, A. L., Panahon, C. J., & Wolfe, L. A. (2002). sensitivity of children's behavior to probabilistic reward: Effects of a decreasing-ratio lottery system on math performance. *Journal of Applied Behavior Analysis, 35*(4), 403-406.

Mazur, J. E. (1989). Theories of probabilistic reinforcement. *Journal of the Experimental Analysis of Behavior, 51*(1), 87-90

McDougall, D., & Granby, C. (1996). How expectation of questioning method affects undergraduates' preparation for class. *The Journal of Experimental Education, 65*(1), 43-54.

Michael, J. (1991). A behavioral perspective on college teaching. *The Behavior Analyst, 14*(2), 229-239.

Neef, N. A., Cihon, T., Kettering, T., Guld, A., Axe, J. B., Itoi, M., & DeBar, R. (2007). A comparison of study session formats on attendance and quiz performance in a college course. *Journal of Behavioral Education, 16*(3), 235-249. https://doi.org/10.1007/s10864-006-9037-3

Neef, N. A., Perrin, C. J., Haberlin, A. T., & Rodrigues, L. C. (2011). Studying as fun and games: Effects on college students' quiz performance. *Journal of Applied Behavior Analysis, 44*(4), 897-901. https://doi.org/10.1901/jaba.2011.44-897

Rehfeldt, R. A., Walker, B., Garcia, Y., Lovett, S., & Filipiak, S. (2010). a point contingency for homework submission in the graduate school classroom. *Journal of Applied Behavior Analysis, 43*(3), 499-502. https://psycnet.apa.org/doi/10.1901/jaba.2010.43-499

Richling, S. M., Williams, W. L., & Carr, J. E. (2019). The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. *Journal of Applied Behavior Analysis*, *52*(3), 701-717.

Rost, K. A. (2018). Reinforcement uncertainty enhances preference for choice in humans. *Journal of the Experimental Analysis of Behavior*, *110*(2), 201-212.

Ryan, C. S., & Hemmes, N. S. (2005). effects of the contingency for homework submission on homework submission and quiz performance in a college course. *Journal of Applied Behavior Analysis*, *38*(1), 79-88. https://doi.org/10.1901/jaba.2005.123-03

Saudargas, R., Madsen, C., & Scott, J. (1977). Differential effects of fixed- and variable-time feedback on production rates of elementary school children. *Journal of Applied Behavior Analysis*, *10*(4), 673-678. https://doi.org/10.1901/jaba.1977.10-673

Sherman, J., & Thomas, J. (1968). Some factors controlling preference between fixed-ratio and variable-ratio schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *11*(6), 689-702. https://doi.org/10.1901/jeab.1968.11-689

Silva, R. J. R. D., Rodrigues, R. G., & Leal, C. T. P. (2019). Gamification in Management Education: A Systematic Literature Review. *BAR-Brazilian Administration Review*, *16*(2).

Solley, K. (2011). Accelerated reader can be an effective tool to encourage and bolster student reading. *Knowledge Quest*, *39*(4), 46-49.

Trautwein, U., & Köller, O. (2003). The relationship between homework and achievement—still much of a mystery. *Educational Psychology Review*, *15*(2), 115-145.

Yildirim, I. (2017). The effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. *The Internet and Higher Education, 33,* 86-90.

IRB APPROVAL FORM

Memorandum

TO: Tiffanie Weeden
Psychological Science Department

CC: Jacob Daar
Psychological Science Department

FROM: Lisa Schade Eckert
Dean of Graduate Education and Research

DATE: January 23, 2020

SUBJECT: IRB Proposal HS20-1089
"Evaluation of quiz format on course performance."
IRB Approval Date: 1/23/2020
Proposed Project Dates: 1/23/2020 – 5/1/2020

RE: Modification to HS20-1089
Original IRB Approval Date: 1/23/2020
Modification Approval Date: 2/7/2020
"Evaluation of quiz format on course performance."

VITA

Graduate School

Northern Michigan University

Tiffanie Weeden

Tweeden@nmu.edu

Northern Michigan University

Bachelors of Science, Psychology, May 2018

Thesis Title:

Evaluating the Effectiveness of Probabilistic Reinforcement, As a Gamification Element, In a College Class Setting

Major Professor: Jacob Daar