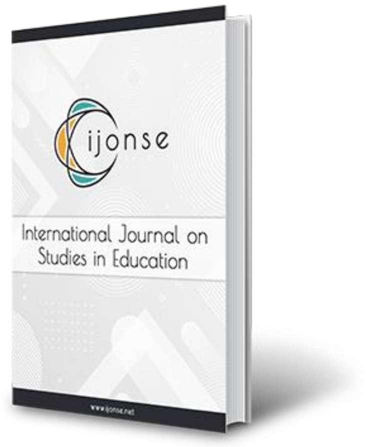

www.ijonse.net

Extra Credit as a Spaced-Study Motivator

Michael Rocha
United States Military Academy, United States of America

Margaret Reynolds
United States Military Academy, United States of America

Si Park
United States Military Academy, United States of America

To cite this article:

Rocha, M., Reynolds, M., & Park, S. (2023). Extra credit as a spaced-study motivator. *International Journal on Studies in Education (IJonSE), 5*(4), 496-514. https://doi.org/10.46328/ijonse.166

International Journal on Studies in Education

2023, Vol. 5, No. 4, 496-514 https://doi.org/10.46328/ijonse.166

Extra Credit as a Spaced-Study Motivator

Michael Rocha, Margaret Reynolds, Si Park

Article Info

Article History

Received:
17 January 2023
Accepted:
13 June 2023

Keywords

Bulk-learning
Spaced-learning
Bonus points
Extra credit
Classroom experiment

Abstract

A big challenge in academia is motivating students and helping them store information long-term. One known method to help students retain information is spaced learning, which allows for certain time intervals in which concepts are re-tested versus waiting until the end of a block to test a learning objective. One method to motivate students is implementing bonus opportunities that keep students engaged in the material. We combine these methods via a spaced daily versus bulk review program incentivized with bonus points. To investigate this method, we designed and implemented a classroom research experiment during Fall of 2022 at the United States Military Academy in three courses. We found that students who were incentivized to do either daily or bulk review bonus opportunities were primarily those who already were doing well in the class. We also found that those who did not pursue bonus opportunities, regardless of section, indicated that time was their limiting factor. We present our findings including quantitative and qualitative results via course grades, mid-semester and end-of-semester surveys, and anecdotal experiences with our students. We then discuss the merit and potential improvement of the experiment to better understand the relationship between student motivation and long-term information retention.

Introduction

As instructors at The United States Military Academy at West Point, we get to teach in a unique learning environment with small student populations per class section, classrooms with chalk boards for encouraged collaboration, and the opportunity for pedagogical classroom experiments and research. Many of the instructors are passionate about improving their own teaching as well as cadet learning. Many instructors utilize bonus points for completing extra work or participating in learning events outside of the classroom. We have tried different avenues for earning those bonus points. For instance, when teaching advanced multivariable calculus to freshmen in Spring of 2022, M. Reynolds (co-author of this work) included challenge problems at the end of each lesson for which cadets could earn bonus points if they completed and turned them in by the next class period. She found she got good participation throughout the semester, and it even pushed students beyond what was tested on exams for the course. This led us to consider the following questions:

- *Research Question 1*: Can we use extra credit to motivate cadets to conduct more ‘spaced’ behaviors?
- *Research Question 2*: Will these be an effective means of practice that will reflect on the final exam?

- *Research Question 3*: How will cadets perceive the different assignments and is it worth the change?

Our initial response and hypothesis to questions 1 and 2 was ``yes." For question 3, we hypothesized that they would show improvement throughout the semester and see the worth in spaced study habits. We wanted to test our theories on daily bonus point incentives and ultimately structured our thoughts into a tangible research question. We then set up a classroom experiment to be conducted throughout the Fall of 2022 in our three different courses (S. Park in MA103: Mathematical Modeling, M. Rocha in MA206: Introduction to Statistics, M. Reynolds in MA371: Linear Algebra). We present our research questions followed by a literature review of bonus point incentive programs as well as spaced versus bulk learning. We then describe our research experiment and methods more clearly, follow-up with an analysis of the quantitative and qualitative results of our experiment, and end with a final discussion and conclusion of the challenges, benefits, relevance, and reproducibility of the experiment.

Literature Review

There have been a wide range of articles comparing study behaviors in students and incentive programs. In their exam study techniques paper, Fini et al. (2010) compared groups of students and varied their study behaviors and recall of course material after being tested. Results show that in the immediately held test, the group with mass studying had better scores, but in delayed tests, the distributed studying technique groups had better scores. The researchers concluded that retention is better with repeated, distributed sessions than a single, massed session immediately before the exam. It also found that mass-studying retention decays at a much faster rate.

In another study, Bagiati et al. (2017) examined what they termed the "forgetting effect". They identified that it declines as information is recalled more often and the information grows stronger the more often it is recalled. For example, the authors estimate that a mechanical engineer freshman will lose about 50% of what they learn by their senior year unless it is reused. If it is reused, the students will recall much more information and understand it better, resulting in better performance. The authors conclude that students who stay engaged with concepts will show improvement, even if it is only similar, not identical, concepts being reviewed.

The idea of 'spaced learning' was investigated further in Carpenter et al. (2012), where the authors reviewed multiple other studies and the effectiveness of different spacing gaps and testing gaps for material, ranging across multiple disciplines from language and mathematics to motor skills in surgical residencies. Studies have consistently shown that students who used spacing techniques perform better on tasks than those who cram, and in general a longer 'spacing gap' results in better retention. They also explore the 'testing gap,' the time between the last study session and a tested event and found that there exists a limit before students begin to forget and performance drops off again. In general, the optimal spacing gap has been found to lie between 10% and 20% of the testing delay. Studies have also shown that an expanding schedule, one where spacing gaps grow longer across the period of study, performs better than static spacing gaps. Finally, they conclude that there are some obstacles to incorporating these findings into a live educational environment as opposed to laboratory findings. One primary finding was that textbooks and other educational materials typically promote segmented topics with little review or overlap. It is then recommended that instructors add to these materials to promote spacing and review, such as cumulative quizzes and exams and offering review in class for previous lessons.

Another study looked at spacing effects on types of learning beyond memory processes, such as the acquisition and generalization of new concepts (Vlach & Sandhofer, 2012). They found that spacing lessons out in time resulted in higher generalization performance for both simple and complex concepts. Specifically, they looked at the concept of food chains because lesson plans generally emphasize generalizations and concept learning across multiple biomes. In the experiment, 36 students were assigned massed, clumped, and spaced techniques. The study found that children in spaced conditions had a greater increase in performance on simple generalization tasks compared to massed and performed better than both massed and clumped concerning complex generalization tasks. The conclusion is that even for complex generalization tasks, spacing promotes improved retention and performance when compared to massed study habits.

One challenging task, then, as an instructor is to encourage students to space out their study habits. However, in a 2009 study, Kornell (2009) found that while spacing is more effective than massing for 90% of participants in terms of retention, after seeing the results, 72% of participants still felt that massing was a more effective study habit. It was also found that studying one larger stack of flashcards to space out repetitions was more effective than multiple smaller stacks, where repeating occurred more frequently. Again, however, students believed that they would perform worse after spaced study than those who massed or crammed, despite the opposite being true. It became apparent that once students settled on what they believe to be a superior study model, anecdotal evidence and their own performance was not enough to sway them from their opinions. Students would not change their preferred study method to spacing from massing, even after evidence of their own enhanced performance over the spacing method and would revert to massing despite the negative results to their learning. In terms of incentivizing students with bonus points, Fuad and Jones (2012) found that by relieving students from the mental pressure of taking tests and making those tested events extra credit, students actually performed better in solving harder problems and learning more advanced concepts in a computer science course. The authors used scheduled and unscheduled quizzes to reduce test anxiety and found that with the stress of scheduled quizzes, students performed better on the harder surprise quizzes worth no points than the scheduled quizzes worth points.

Another study found that the best way to successfully get undergraduate students to attend on-campus events outside of regular class time was to offer extra credit (Foltz et al., 2021). The researchers wanted to enhance classroom learning with extracurricular activities and intellectual growth. Researchers found that extra credit outweighed material factors and even personal interest in a student's decision to attend events outside the classroom and recommends that faculty link material gain, interpersonal gain, intellectual gain, and grade performance to influence decision making after class. It was emphasized that extra credit is not the sole motivator, but it was found to be the most influential one.

One specific study in Ireland discusses a bonus point initiative where students are given the opportunity to earn bonus points during lower-level math classes (O'Meara et al., 2020). Those bonus points are then applied to upper-level math class that they can choose to take. This initiative stemmed from an effort to encourage students to pursue higher level mathematics. While the enrollment of students in higher level math classes has increases steadily since the introduction to the program, concerns have been raised that students who are not prepared for those classes still choose to do the assignments, gain the bonus points, and move to the higher level with the extra credit incentive, but then are still unprepared for those classes. Concurrently, grading standards have adjusted to this under-prepared population in the classroom, hence the program's benefit is now being investigated more thoroughly.

An article by Ingalls (2018) specifically investigated a bonus incentive program for online homework systems for pre-calculus and statistics courses at a private university. The results of the program found that students who were incentivized with bonus for achieving mastery on the online homework assignments performed better on final assessments in the course than those who were not incentivized with extra bonus points. These bonus points improved their semester grade and showed strong evidence of improving exam grades.

A study found in by Harrison et al. (2011) investigated whether students who are already doing well in non-introductory psychology college courses are more likely to complete extra credit assignments, and whether gender and class size influence extra credit completion rates. This study had a large sample size of 508 students. The results showed that students with higher grades were more likely to complete extra credit assignments than those with lower grades. Furthermore, female students were more likely than male students to complete extra credit, and students in larger, lecture-style courses were more likely to do so than those in smaller classes. The study discusses the implications of these findings for the usefulness of extra credit and student motivation.

Lastly, Rousu et al. (2015) created a study that investigates whether providing monetary incentives in a prisoner’s dilemma game can improve student learning in a principles of economics college course. The study includes 641 students from four universities who were randomly assigned to three groups: one group played a two-player prisoner’s dilemma game for real money, another group played the same game with no money at stake, and a control group where no game was played. After the game, all students were given a set of common exam questions to measure their learning. The results of the study show that students who played the game for real money earned higher test scores than those who played the hypothetical game or were in the control group where no game was played. This finding challenges the conventional wisdom that monetary incentives are unnecessary in classroom experiments.

Methods

Study Population

The authors had the opportunity to run the experiment simultaneously with multiple sections between three different courses. This experiment was prepared and set up in Spring of 2022 and conducted during the Fall of 2022. The three courses we taught during this time have a wide range in student population and difficulty.

MA103 (Introduction to Mathematical Modeling)

The first course is MA103 (Introduction to Mathematical Modeling) which is a course that provides an introduction to mathematical techniques utilized for modeling natural and scientific phenomena and exploring their behavior. This course is the first in the standard core mathematics sequence, which is required for students who did not qualify for the advanced core mathematics sequence. The majority of students taking this course are freshmen. The main focus is on Discrete Dynamical Systems, which encompass topics such as linear and nonlinear iterations, fixed points, periodic cycles, and stability. Additionally, the course covers basic concepts of vector and matrix arithmetic, complex number arithmetic, determining dynamics in the plane with real and complex eigenvalues, and stochastic modeling using Markov processes. These mathematical concepts and techniques are applied to real-world scenarios in fields like finance, biology, social sciences, and physical sciences. The use of Excel and Mathematica is also emphasized in the course.

MA206 (Probability & Statistics)

The second course in our study was MA206 (Probability & Statistics) which is an introductory course that provides students foundational understanding of statistical concepts and probability. This is the final course in the mathematics core curriculum, required for all students regardless of academic major. The majority of the student population is from the sophomore class, though some engineering majors (mechanical engineering, chemical engineering) take the course during their junior or senior year. Coverage includes data visualization and interpretation, probabilistic models, independence, simulation, random variables and their distributions, hypothesis testing, confidence intervals, and multiple linear regression. The course also introduces engineering applications of probability and statistics techniques. Concepts are explored through real data. The use of the R statistical software enhances understanding, problem solving, and communication.

MA371 (Linear Algebra)

Our third course in the study was MA371 (Linear Algebra) which is a course that covers both computational and theoretical concepts that has applications in a wide variety of subjects. As a result, the students who take this course generally range from majors and minors in mathematics along with majors in physics, a variety of engineering disciplines, economics, and computer science. It is a required course for mathematics majors, and an elective for most other courses. Generally, 35 to 50 students take the course each semester, ranging from sophomores through seniors. The class covers topics starting with the basics of solving a system of equations, converting those systems into matrix-vector systems, and using matrix algebra in order to manipulate the system and find a variety of characteristics defining the given system. It then moves on to discuss vector spaces and subspaces more generally, with a focus on linear independence and dependence, linear transformations, eigenvectors and eigenvalues, as well as applications. We then move into a block of material that takes in-depth investigation into orthogonality, least squares, linear systems, and general inner product spaces. Finally, we round out the course by understanding symmetric systems, quadratic forms, and singular value decomposition, which culminates in principal component analysis (PCA). All of these topics are discussed in detail with both algebraic and geometric views in mind as well as relevant applications. We explore these concepts even further using at least one computer algebra system. In the Fall of 2022, the systems we used were Mathematica by Wolfram and MATLAB by MathWorks.

Experiment Methodology

In this study, instructors taught three courses comprising three to four sections each. Each instructor randomly divided their sections into two categories: “bulk” and “spaced.” Students in the spaced sections had the opportunity to earn bonus points by completing daily board sheets for extra credit, while students in the bulk sections earned bonus points through review-based activities, such as correcting submitting correction for exam problems or completing a bulk-review sheet prior to the exams. For future reference, the Academy calls mid-semester exams “Written Partial Reviews” (WPR) and final exams “Term End Exams” (TEE). The total number of bonus points available was equal across all sections and the primary course instruction remained unchanged. We also conducted two surveys throughout the semester to gather additional quantitative and qualitative feedback from students. At the end of the semester, we analyzed both the test grades and survey results.

Each course split their sections into two categories. Two sections per class were assigned to be the spaced-learning sections, and the remaining sections (one for MA371 and MA103 and two for MA206) were assigned to be the bulk-learning sections. For the bulk-learning sections, each course provided opportunities to do a bulk review opportunity for bonus points. For MA371, this consisted of review sheets to be completed the week prior to each exam throughout the semester (four total). In MA103, students were required to take three exams during the course. To support their bulk-learning, students were provided with two opportunities to revisit and redo the problems they missed for bonus points. Since MA206 is structured with only a single midterm and final, the bulk review consisted of test corrections after the midterm and a cumulative review assignment leading up to the final exam. Groups were assigned a topic block and given the option to either develop a comprehensive board sheet of problems covering the important concepts or creating a video review session covering the same topics.

The spaced-learning sections for each course consisted of completing daily board sheets. Across all courses, we taught or reviewed the main concepts of the lesson for that day, then provided these board sheets with questions and problems related to the lesson. Each course would assign bonus points for completing the lesson notes either by the next class period or within a week of that lesson. This ensured close to daily review of concepts for the students in these sections. The total amount of points available to earn, regardless of the section being bulk-learning or spaced-learning was equal to 1% of the course points, so there were opportunities to earn the same amount of bonus points regardless of the type of section a student was in.

To gauge student perceptions of the bonus opportunities and assess why students chose not to participate, a survey was given to all sections at two different points during the semester. The survey was administered halfway through the fall semester and again at the end of semester, just before final exams. In the survey, we asked questions involving their study routine, why they chose to take advantage of the bonus opportunities or not, and ranking the balance between time spent and bonus earned on these bonus opportunities. Furthermore, we explored questions involving whether or not students felt that doing the bonus increased their understanding of the course material and if the bonus opportunities influenced their study habits. In particular, we asked the following open response questions:

1. What is your typical study routine for an exam? (e.g., repeated practice over time or cram right before

it?)

2. If you have not taken advantage of the bonus opportunities/opportunity, why not?
3. How helpful are the bonus opportunities/opportunity when preparing for class?
4. How did the bonus opportunities influence your study habits and how you approached the class?

We also included the following quantitative Likert scale questions which students could answer on a scale of 1 to 5, with clarification provided for the scale for each question:

5. Rank the balance between time spent and bonus earned on these bonus opportunities/opportunity (where 1 is too much time compared to the points earned, 5 is too generous compared to the points earned, and 3 is a good balance of time spent and bonus earned.)
6. On a scale of 1 to 5, how much did doing the bonus further your understanding of the material covered? (where 1 did not impact your understanding and 5 greatly enhanced your understanding.)

We asked question 1 to set a baseline for students' understanding of their own study habits. Question 2 gave us an understanding of why students chose to partake in bonus opportunities or not, which could range anywhere from time management to difficulty level. We asked questions 3 and 4 to give the students an opportunity to reflect on how useful their time spent on completing the bonus was toward their learning and if it influenced them to change anything about their study habits from the beginning of the semester, to middle of semester, and finally at the end of the semester. For the Likert scale questions, we wanted to give students the opportunity to numerically evaluate their time spent on bonus opportunities and whether it was useful. Specially, we wanted to give a limited number of values to choose from (1 to 5) in order to get a broad overview of their opinion which they could expand upon in the earlier free response questions. We also chose to include an odd number of options to choose from (1, 2, 3, 4, or 5) so that students could choose "3" as a perfect balance between time spent and bonus earned in question 5 or as measure of a moderate impact on their understanding in question 6.

To assess academic performance, deidentified grades were obtained through the Academy's registrar office after the semester ended. Administrative information such as the course number, section number, and instructor were included. Sections were grouped such that they could be identified as "bulk" or "spaced," but multiple sections falling in the same group were not separated. We pulled multiple grades to construct our data set: placement exam scores, semester exam scores, final exam scores, and bonus points.

The first grade we pulled was the mathematics placement exam results which every cadet takes upon entering the Academy to determine which mathematical track they will take (remedial, regular, or advanced). This exam, known as the fundamental placement exam or FCE for short, is scored on a range from 0 to 50, with a perfect score placing into the advanced math courses and scores of 0 tending towards placement in the remedial math course for their first semester. The data set also included exam scores for each class throughout the semester as well as the final exam scores. MA103 and MA371 both conducted 3 exams throughout the semester while MA206 had only one midterm. All three courses also included a final exam conducted in December of the Fall 2022. To account for the difference in number of exams for each course, we combined exams to generate an average exam grade for MA103 and MA371 in order to compare to the one midterm exam in MA206. Lastly, we collected the number of bonus points each student earned throughout the semester.

Results

Preliminary Survey Results by Course

Before comparing the overall group performance, we analyzed the overall conduct of the classes to ensure it is reasonable to collectively compare bulk and spaced study groups. First, based on survey results, we assessed the difference in perceived balance between extra credit effort and point payoff between courses. The question asked to all three classes was "Rank the balance between time spent and bonus earned on these bonus opportunities/opportunity (where 1 is too much time compared to the points earned, 5 is too generous compared to the points earned)." The initial results are shown in Table 1. We conducted a Welch ANOVA test on means of the responses, resulting in $F_{2,80.651} = 1.7219$ and p-value equal to 0.1852. This confirmed no significant difference between courses. With these results, we can confidently assess perception of balance between groups.

Table 1. Summary Statistics of Balance Responses by Course

Class	Max	Med	Mean	Min	n
MA103	5	3.0	2.62	1	42
MA206	5	2.5	2.32	1	50
MA371	4	2.0	2.29	1	35

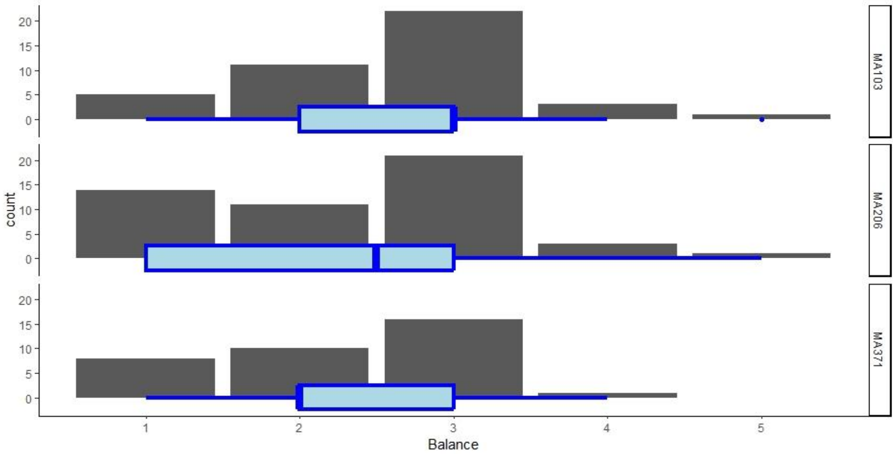

Figure 1. Histogram of Balance Responses by Course

Next, we conducted the same analysis for the perception between the extra credit assignment and gained level of understanding between all three courses. The question asked in the survey was "On a scale of 1 to 5, how much did doing the bonus further your understanding of the material covered? (where 1 did not impact your understanding and 5 greatly enhanced your understanding.)". The initial results are shown in Table 2 and Figure 2. Similarly, we conducted a Welch ANOVA test, resulting in $F_{2,71.143} = 1.0073$ and p-value of 0.3704. We concluded there was no significant difference between courses in perception of gained understanding through bonus assignments. Given that neither understanding nor perception of balance is significantly different between the three courses, we felt justified in assessing the impacts of different bonus methodologies on final exam performance.

Table 2. Summary Statistics of Understanding Responses by Course

Class	Max	Med	Mean	Min	n
MA103	5	3	2.93	1	41
MA206	5	3	2.86	1	49
MA371	4	3	2.53	1	32

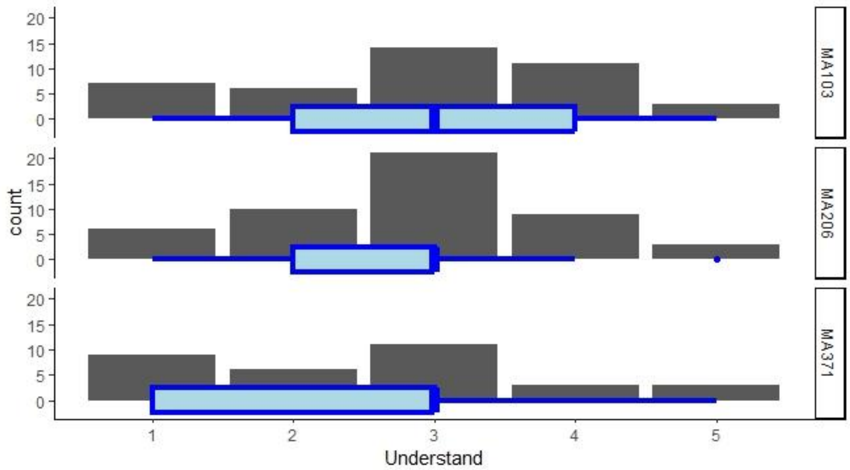

Figure 2. Histogram of Understanding Responses by Course

Survey Results by Bonus Method

After concluding no difference between the three classes, we then analyzed the same survey results broken up by bonus methodology. “Bulk” study involved larger assignments such as test corrections or a cumulative review session. “Spaced” students had smaller but more numerous board sheet assignments every night. A total of 59 responses were collected from students in the bulk group and 83 from the spaced groups. The Figures for both Balance and Understanding are shown below in Figure 3 and Figure 4.

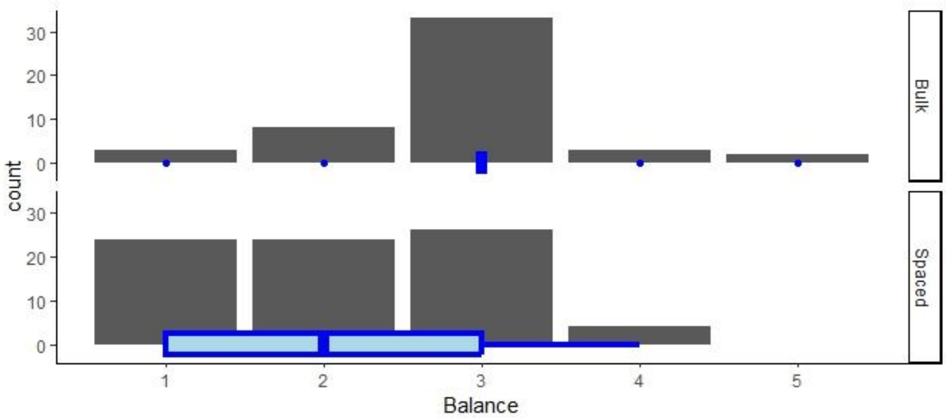

Figure 3. Histogram of Balance Responses by Method

We conducted a two-sample t-test comparing the means for the Balance responses (Figure 3). The resulting t-statistic is 4.75 with a p-value of 5.43×10^{-6}, indicating a significant difference in the mean response. We further considered the potential nonlinear nature of Likert scale responses and conducted an exact Wilcoxon-Mann-Whitney test for the median, resulting in z equal to 4.29 and a p-value of 1.17×10^{-5}. This is congruent with our t-test, indicating a significant difference between students' perception of bulk and spaced bonus opportunities. Specifically, there is evidence that students felt spaced bonus assignments were less balanced between the time spent on the assignment and the bonus points earned.

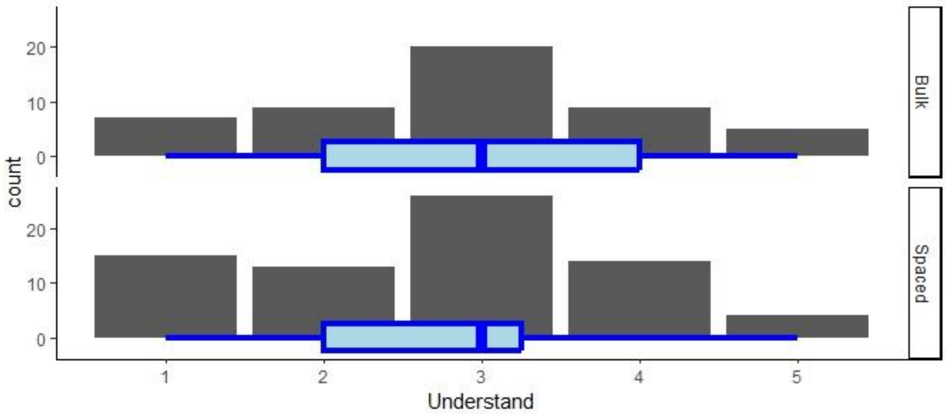

Figure 4. Histogram of Understanding Responses by Method

Similarly, we conducted a two-sample t-test comparing the means for the Understanding response and a Wilcoxon-Mann-Whitney test for the median (Figure 4). These resulted in t equal to 0.99 (p-value equal to 0.324) and z equal to 0.89 (p-value equal to 0.892), respectively, indicating no statistically significant difference between students assigned bulk bonus opportunities and students assigned spaced opportunities with regards to their perception on the assignment and how it helped further their understanding of course concepts.

World Clouds and Text Analytics

In addition to the students' perception of the bonus activities, we wanted to investigate if the bonus point opportunities had the intended effect of influencing student behavior. One question asked on the survey was "What is your typical study routine for an exam? (e.g., repeated practice over time or cram right before it?)." The results, shown in Figure 5 and Figure 6, indicate similarities in free- response answers and no significant difference between methodologies. Of note, repeated practice and building a note sheet or study sheet were high in both groups, followed by a general time range of 2 to 5 days when students generally begin studying. Interestingly, board sheets specifically show up in the top 10 responses for the spaced groups, but not in the bulk groups, potentially indicating an increased awareness of the problems as a study tool even if they are not taking advantage of the bonus opportunities or feel that they are balanced.

Another question asked on the survey was "If you have not taken advantage of the bonus opportunities/opportunity, why not?" Overall, the most consistent answer across all groups was "time," as seen in Figure 7. It was overwhelmingly the number one response across both groups and may be indicative that, regardless of the medium presented, time is a major factor in a student's willingness to do additional work outside of the classroom for bonus points.

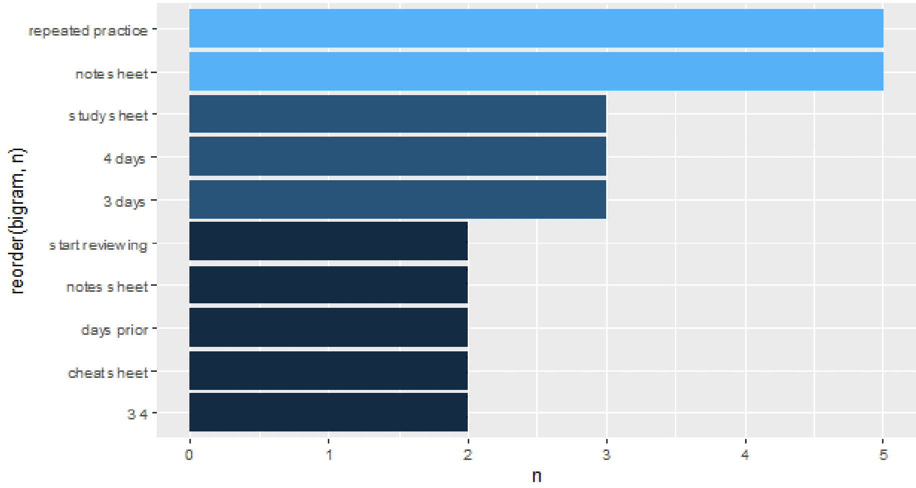

Figure 5. Response frequencies of Bulk Study Routine

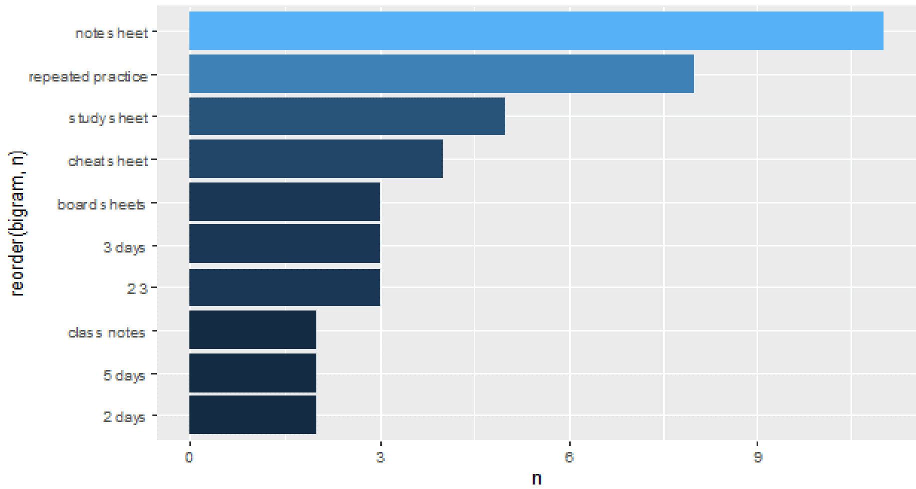

Figure 6. Response frequencies of Spaced Study Routine

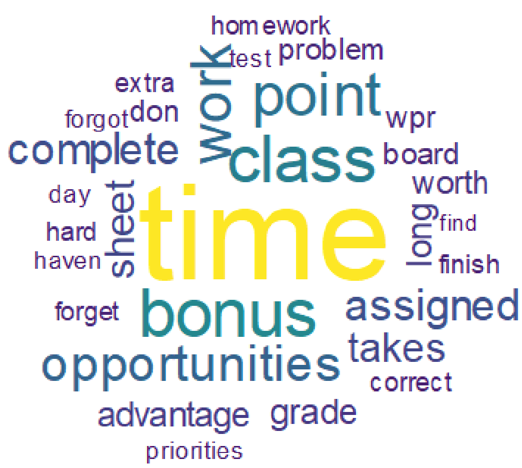

Figure 7. Word Cloud of Why Not (Consolidated)

Finally, we compared the free text responses separated between the two groups, as seen in Figure 8. A pattern emerged where, in general, responses from the bulk group focused on the number of opportunities and major graded events. Conversely, those from the spaced group focused more on the length of the assignments, perceived difficulty, and how many points they were worth. This is consistent with the disparity in views on balance seen in our earlier section on survey results by bonus method.

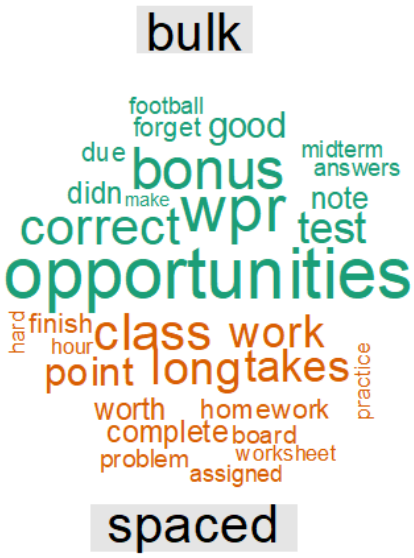

Figure 8. Word Cloud of Why Not (Comparison)

Analysis on Final Exam Performance and Improvement between Groups

After analyzing the students' qualitative perspectives of perception on balance, understanding, and reasons for not partaking in bonus opportunities, we analyzed potential quantitative impacts on the students' performance. For each student, we had their placement exam results upon admission to the academy, their exam performance history for the course, and their final exam scores.

Final Exam Grades by Method

We first analyzed the final exam scores from the different courses to confirm we can properly group them together by bonus methodology. We conducted a Welch ANOVA analysis across final exams for all three groups, resulting in $F_{2,81.656} = 0.98984$ and a p-value of 0.376. With no significant difference between the three courses, we justified the decision to group course data and compare by bonus methodology group. The confidence intervals of the differences between classes can be seen in Figure 9.

The final exam averages for the bonus method groups are shown in Table 3. We conducted a two-sample t-test for a difference in means, resulting in t equal to 2.35 and a p-value of 0.020. This indicates that the average final exam score for those in the spaced group was higher than the average final exam score for the bulk bonus group. More specifically, the 95% confidence interval yields a true difference between 0.007 and 0.075 on the final exam.

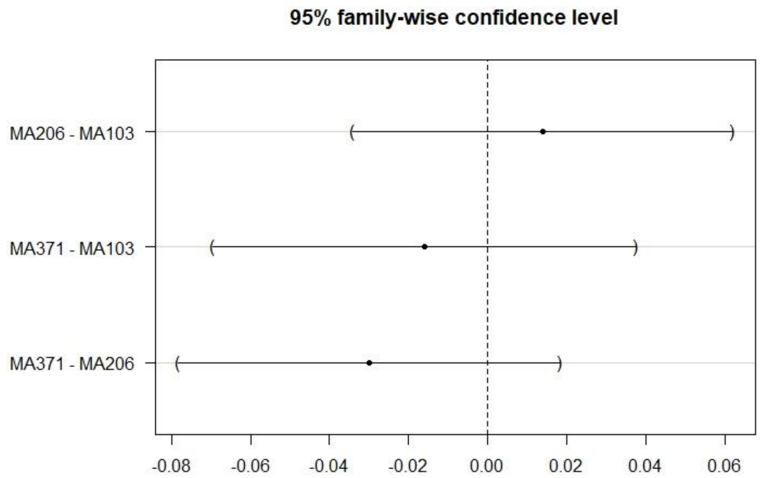

Figure 9. 95% Confidence Interval of Final Exam Differences

Table 3. Summary Statistics of Final Exam Scores by Method

Method	**Mean**	**sd**	**Count**
Bulk	0.819	0.1092	62
Spaced	0.860	0.0979	91

Improving on these results, we conducted a linear regression model to control for variation within the different courses and placement exam results so that we could isolate the effects of the bonus method from innate math ability. We conducted stepwise selection by AIC to reduce our factors. An ANOVA test resulted in a p-value of 0.604, indicating no significant difference between the full model and our reduced one, so the reduced model was used (see Table 4).

Table 4. Linear Regression Results of Final Exam Performance

		Final Exam	
Predictors	*Estimates*	*CI*	*P*
(Intercept)	0.68	0.37 — 0.99	<**0.001**
Placement	-0.50	-0.95 — -0.06	**0.025**
WPR Avg	0.10	-0.28 — 0.48	0.596
Method [Spaced]	0.06	0.01 — 0.11	**0.028**
Course [MA206]	0.08	-0.11 — 0.26	0.397
Course [MA371]	-0.46	-0.71 — -0.21	**<0.001**
Placement x WPR Avg	0.67	0.16 — 1.17	**0.010**
WPR Avg x Course [MA206]	-0.05	-0.29 — 0.19	0.687
WPR Avg x Course [MA371]	0.52	0.22 — 0.82	**0.001**
Method [Spaced] x Course [MA206]	-0.04	-0.10 — 0.02	0.188
Method [Spaced] x Course [MA371]	-0.08	-0.15 — -0.02	**0.013**
Observations	153		
R^2/ R^2 adjusted	0.651 / 0.627		

The linear regression model indicates that after controlling for WPR performance and placement test performance, there is a difference in final exam score by bonus group assigned. Interestingly, the effect is different depending on the course assigned to. The largest positive effect is found in MA103, with a reduced effect in MA206 and a potentially negative effect in MA371. This may be indicative of course difficulty or concepts impacting the effectiveness of the bonus opportunities or a diminishing effectiveness as students develop further into their academic career.

Individual Growth by Method

After analyzing average final exam grades, we also wanted to assess the bonus opportunity impacts on individual growth and knowledge retention of the students. To assess this, we calculated the difference between average WPR performance throughout the semester and the performance on the final exam. The purpose was to assess the impacts on individual growth and reduce variability between students in the performance data. The summary statistics for the differences are shown in Table 5 and Figure 10. We conducted a two-sample t-test, resulting in a t-statistic of 0.887 and p-value of 0.376, concluding that there is not a significant difference between the bulk and the spaced group with regards to a difference in average WPR performance and final exam performance.

Table 5. Summary Statistics for Difference between Final Exam and WPR Average

Method	Mean	sd	Count
Bulk	0.01421	0.0881	62
Spaced	0.00131	0.0887	91

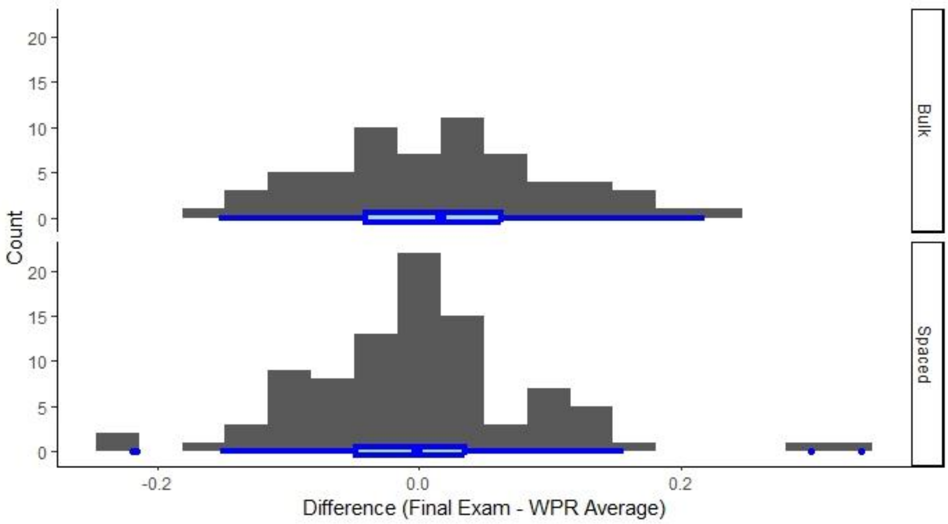

Figure 10. Distribution of Personal Growth by Bonus Method

We then wanted to investigate an association after controlling for placement results and course. We ran a linear regression model on the difference each individual student observed within the class based on method and placement exam. We also ran an interaction model between bonus method group, class, and placement results and used stepwise selection while keeping method. The comparative results are shown below in Table 6. We found that there is no significant association between methods and the difference between the WPR average and final exam grades.

Table 6. Linear Regression Results for Individual Growth

	Difference			Difference		
Predictors	*Estimates*	*CI*	*P*	*Estimates*	*CI*	*P*
(Intercept)	-0.01	-0.11 — 0.08	0.763	0.04	-0.01 — 0.10	0.132
Method [Spaced]	-0.01	-0.03 — 0.02	0.629	-0.01	-0.04 — 0.02	0.460
Course [MA206]	0.15	0.04 — 0.27	**0.011**			
Course [MA371]	-0.14	-0.28 — -0.01	**0.039**			
Placement	0.06	-0.08 — 0.21	0.406	-0.04	-0.11 — 0.03	0.276
Course [MA206]x Placement	-0.20	-0.37 — -0.03	**0.022**			
Course [MA371]x Placement	0.08	-0.11— 0.27	0.411			
Observations	153			153		
R^2/ R^2 adjusted	0.269 / 0.239			0.013 / -0.000		

Paired Comparison of Final Exam Scores by Method

Recall from Table 4 that placement exam results and WPR results both had significant effects when modeling final exam grade. We took that knowledge and paired similar students based on the Euclidean distance between these results to conduct a paired t-test between similar students. Students from the bulk groups were paired with a student from the spaced group in the same course.

As a result of uneven groups, not every student was grouped and thus not every student's score was used in the analysis. This resulted in 50 matched pairs. The average WPR grade for the bulk members of the paired test was 82.06, while the average WPR grade for spaced members was 82.86%. A two-sample t-test resulted in a p-value of 0.7375 indicated the two group's averages were not significantly different. Concerning placement exams, the average for the bulk group was 73.56% and spaced was 75.28%. A two-sample t-test resulted in a p-value of 0.6571, indicating the two groups' averages were not significantly different. We therefore feel confident we can move forward with our paired analysis for our groups as the pairing process resulted in similar behavior.

The distribution of the paired differences is shown in Figure 11 in the order of spaced participant to bulk participant. The resulting average difference between the two groups was 0.0286, with those in the spaced groups performing slightly better than their matched counterparts in the bulk sections. A paired t-test resulted in t equal to 2.276 and a p-value of 0.0272, indicating a significant difference between the matched pairs of groups. A 95% confidence interval is (0.0033, 0.0538), indicating a small but statistically significant improvement from the spaced group compared to the bulk group on final exam performance.

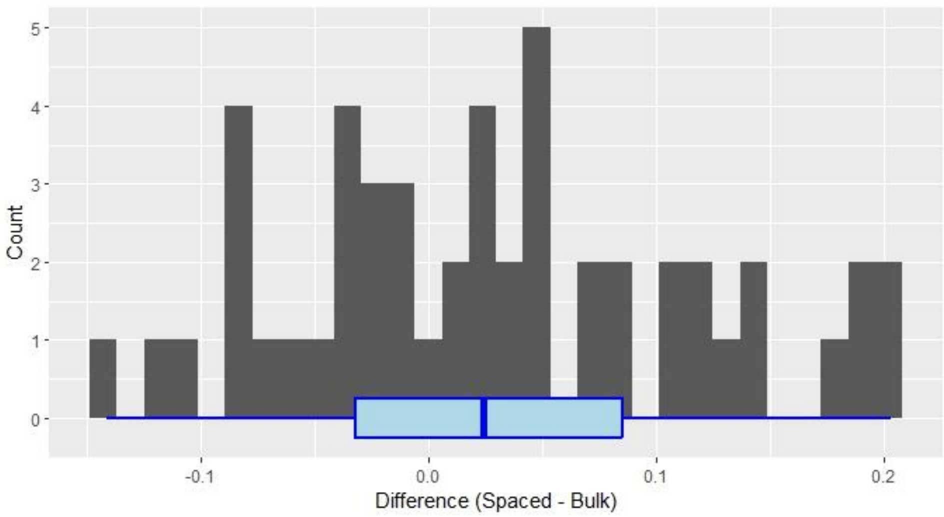

Figure 11. Distribution of Final Exam Differences between Matched Pairs

Discussion and Conclusion

We sought to analyze if extra credit could be used to motivate spaced study behavior in our students across three different math courses which included freshmen through seniors in undergraduate programs of various disciplines. We used both self-reported survey results as well as quantitative test performance to answer three questions:

- *Research Question 1*: Can we use extra credit to motivate cadets to conduct more ‘spaced’ behaviors?
- *Research Question 2*: Will these be an effective means of practice that will reflect on the final exam?
- *Research Question 3*: How will cadets perceive the different assignments and is it worth the change?

Anecdotally, bonus opportunity participation was higher among bulk classes than among spaced. Additionally, consistent with previous research (Harrison, 2011), it was generally the students who understood the material already and did not need bonus points that completed the assignments, especially in spaced classes. Those who participated in the spaced assignments generally displayed better time management practices. Points alone did not produce a perceptible increase in bonus submission participation. However, survey results did indicate that spaced groups were more likely to utilize board sheets as a study reference when preparing for exams than bulk groups. The added emphasis on these assignments might not have influence nightly study, but they did appear to influence the manner in which exams were prepared for, promoting more practice problems for mathematical tests and less reliance on skimming the textbook. While we did not capture if behavior changed in the manner we intended, this does reflect that the emphasis on extra credit did influence behavior outside of class consistent with previous research (Foltz et al., 2021).

Quantitatively, the different sections did have statistically significant differences in final exam performance. This supports previous findings (Ingalls, 2018) that students incentivized with bonus for mastery on homework assignments perform better on final assessments. Overall, spaced bonus courses outperformed bulk bonus courses by roughly 3% on the final exam, on average. We found that this difference was consistent even when pairing students across sections based on similar performance. Interestingly, while the overall final exam shows a significant difference, students did not show significant differences in growth between midterm exams and final exams. Therefore, it seems the benefits of the spaced behavior manifested early in the semester and plateaued. This growth manifested even with lower participation rates, which might indicate that an increased emphasis on the practice problems had a positive effect on understanding the material, even if the student did not complete the board sheet for submission.

Qualitatively, students held disapproving views of the spaced bonus opportunities, especially when compared against the bulk opportunities. This is unsurprising given most students' preference to massed methods, even when shown that spaced methods of study produce better results and understanding (Kornell, 2009; see also Fini et al.,2010; Bagiati et al., 2017; and Carpenter et al., 2012). They did not perceive any difference in understanding gained from the different assignments, despite grades being generally higher among spaced groups. Additionally, it was repeatedly expressed that spaced assignments were disproportionately long relative to the points they were worth.

Overwhelmingly, the number one reason students cited for not completing bonus opportunities across both groups was time. This is consistent with anecdotal observations and is a reasonable contributor to the difference in participation rates for spaced bonus submissions. When differentiated between spaced and bulk groups, spaced groups were more likely to emphasize that the points were not worth the effort or that they did not know how to do all the problems. Bulk groups were more likely to specify specific out of class requirements that prohibited them from completing the assignments and pointing out that there were few opportunities.

In general, we found that while the use of spaced opportunities might not have impacted an increase in nightly practice, it did appear to have a positive impact on influencing students to practice more for math exams instead of passive reading. The continued emphasis on practice problems may have also contributed to an overall awareness to understanding the problems in class, though future research would need to confirm this specific feeling. Regardless, there was a positive impact on final exam performance between 0.7% and 7.5% when compared to students from the bulk group of similar mathematical aptitude and performance. This is promising, as the known benefits of spaced versus bulk learning demonstrably influence higher level mathematical learning as well.

Some key areas of future research include location and class. Our study was limited to the United States Military Academy. While our study spanned all four classes across three different courses, the requirements and time commitments of the students are not universal across all schools. Further study in other universities might better investigate the influence of spaced bonus opportunities where there are different requirements competing for time.

More interestingly, the differences between spaced and bulk study methods appeared to decrease as the course levels went up. While Introduction to Mathematical Modeling (MA103), a course primarily composed of freshmen, saw the largest benefit, it is also a core course that every student must complete, regardless of discipline. Conversely, Linear Algebra (MA371) saw the smallest, though still significant, difference. This course is an upper-level course comprised mostly of STEM (Science Technology Engineering and Mathematics) majors in junior and senior years. It is unclear through our study if this downward trend is a result of reduced variance in the population studying the material. One possible explanation is the forgetting effect (Bagiati et al., 2017), where students who stay engaged with concepts will show improvement, even if those concepts are similar but not identical. It might be that by junior and senior year, the students are more experienced in academia and better understand how to study and practice on their own without the need of bonus opportunities to motivate them. It might also be that STEM majors already have a better baseline of the material analyzed here, and thus have less benefit to gain from increased study. Further research should be conducted to isolate which factors most impact the benefits students see.

Finally, future research can be done to see if the increased emphasis on the board sheets alone is sufficient to see these benefits. As many students expressed the points were not worth it, and indeed many did not complete board sheets for points, it might be good practice for instructors of mathematical courses to simply increase emphasis on nightly practice both before and after the lessons, calling back to previous lessons and course concepts. It is also possible that the points alone, even if not seen as sufficient to actually do the work, may have inherently placed more importance than words alone would in the classroom. An understanding of the best way to motivate students to practice will undoubtedly benefit instructors in the classrooms, and their students, to truly master and retain the material.

Acknowledgements

The authors wish to acknowledge the United States Military Academy and their Center for Faculty Development which prompted this research through their Master Teacher Program.

Notes

The authors claim no conflict of interest in the experiment's setup, data collection, analysis, and presentation of results. M. Rocha performed the data analysis. M. Rocha and M. Reynolds primarily edited the final paper. All authors contributed to data collection and initial draft of paper.

References

Bagiati, A., Subirana, B., & Sarma, S. E. (2017). Examining stem learning through memory retention: A research agenda. *European Society for Engineering Education, 45*.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24,* 369–378.

Fini, A. A. S., Zainalipour, H., & Mozafari, S. H. (2010). Comparing the effect of mass and distributed study techniques on remembering level and its relation with the exam date. *Procedia Social and Behavioral Sciences, 5,* 1751–1759.

Foltz, K. A., Clements, M., Fallon, A., & Stinson, A. (2021). Extra credit and decision-making: Understanding college students' motivation to attend on-campus events. *The Journal of Campus Activities Practice and Scholarship, 3*(2), 5–15.

Fuad, M. M., & Jones, E. J. (2012). Using extra credit to facilitate extra learning in students. *I.J. Modern Education and Computer Science, 6*, 35–42.

Harrison, M. A., Meister, D. G., & LeFevre, A. J. (2011). Which students complete extra-credit work? *College Student Journal, 45*(3), 550–555.

Ingalls, V. (2018). Incentivizing with bonus in a college statistics course. *REDIMAT-Journal of Research in Mathematics Education, 7* (1), 93–103.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*(9), 1297–1317.

O'Meara, N., Prendergast, M., & Treacy, P. (2020). What's the point? Impact of Ireland's bonus points initiative on student profile in mathematics classrooms. *Issues in Educational Research, 30*(4), 1418–1441.

Rousu, M. C., Corrigan, J. R., Harris, D., Hayter, J. K., Houser, S., Lafrancois, B. A., Onafowora, O., Colson, G., & Hoffer, A. (2015). Do monetary incentives matter in classroom experiments? Effects on course performance. *The Journal of Economic Education, 46*(4), 341–349.

Vlach, H. A., & Sandhofer, C. M. (2012). Distributed learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Dev, 83*(4), 1137–1144.

Author Information

Michael Rocha

https://orcid.org/0009-0005-0030-2732

United States Military Academy

606 Thayer Road, West Point, NY 10996

United States of America

Margaret Reynolds

https://orcid.org/0000-0002-9762-8851

United States Military Academy

606 Thayer Road, West Point, NY 10996

United States of America

Si Park

https://orcid.org/0009-0004-1301-3019

United States Military Academy

606 Thayer Road, West Point, NY 10996

United States of America

Contact e-mail: *si.park@westpoint.edu*