UNIVERSITY OF CALIFORNIA SAN DIEGO

Delayed feedback might, but not collaborative exams, improve long-term student retention of course content

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Science

in

Biology

by

Martín Arteaga

Committee in charge:

Professor James Edward Cooke, Chair

Professor Stanley Malvin Ming-him Lo, Co-Chair

Professor David Kacev

2021

The thesis of Martín Arteaga is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

Dedicated to the memory of my best friends

Irvin Rodriguez

and

Mario Bahena

I thank them both for teaching me how to be resilient

and to pursue my passions

Special thanks to Limón, my family, and friends

for the unconditional support and love

Lastly, to James Cooke,

who showed me a different way to look at education

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT OF THE THESIS

Delayed feedback might, but not collaborative exams, improve long-term student retention of course content

by

Martín Arteaga

Master of Science in Biology

University of California San Diego, 2021

Professor James Edward Cooke, Chair
Professor Stanley Malvin Ming-him Lo, Co-Chair

Improvements in student learning and the retention of learned content are possibly the most important outcomes that educators strive to achieve. Exams have been widely utilized in efforts towards gauging the students' overall understanding and performance. However, exams can also serve as a means for learning and solidifying content. While some studies suggest collaborative exams are helpful at improving retention (Gilley and Clarkston, 2014; Cooke et al., 2019) others have found no effect (Leight et al., 2012; Vojdanoska et al., 2009). Multi-stage collaborative exams provide students with interactive, rewarding, and feedback experiences. The current thesis strives to determine whether collaborative exams improve retention of course

content in an upper division physiology course. Utilizing a mixed-crossover design similar to previous studies and increasing student motivation, collaborative exams do not improve retention of course content for upper division physiology students at UCSD.

GENERAL INTRODUCTION

With student learning and retention of content being the core objectives for institutions, educators often find themselves exploring ways to improve student outcomes. How students experience a course - with its given content and the student's interactions with others in the course - could possibly shed more light on increasing content retention. Students that undergo a teamwork-based approach to solve complex problems actively learn the course content and can retain more compared to students that do not work in groups (Pendergrass et al., 2001; Purzer, 2011; Menekse et al., 2013). The use of collaborative learning has gained momentum as of late and the technique has been applied across academic disciplines (Adolphus et al., 2013; Nkechinyere et al., 2018; van Eijl et al., 2005). Although exams are often a tool to evaluate and assign grades, the practice can also serve as an additional step towards solidifying concepts and knowledge. As courses and students vary, the teaching approach often gets analyzed and adjusted for new ways to engage students with a goal of improving learning.

One area gaining prevalence is the effects that exams contribute towards student learning. The testing effect as described by Karpicke and Roediger, 2006, offers an additional teaching strategy when it comes to testing students. The authors' results suggest that when students take an exam, they enhance their learning in the longer term (ie: 7 days) compared to controls who studied the material for the same amount of time. The type of exam that students encounter can also play a role in improving the retention of content. Exams can take on many different shapes and contain varying levels of complex test questions. True or false, multiple choice, fill in the blank, and free response questions allow educators to gauge student learning gains on the course content. Compared to multiple choice questions, short answer questions lead to improved content retention on subsequent open-ended and multiple choice questions (McDaniel et al., 2007).

McDaniel and colleagues (2007) postulate that these results occur because students are having to recall information to answer questions, rather than recognizing correct answers when they are presented.

Collaborative exams have also been gaining recognition within education. Students that work in group exams provide differing perspectives towards the exam questions and can provide or receive insightful information. Group exams offer students a form of social teaching and learning by allowing students to work towards a common goal (Pendergrass et. al., 2001; Vojdanoska et. al., 2010; Donovan et. al., 2018). However, group exams are not often part of conventional teaching methods. Generally, students are taught in groups (classes) and these classes serve as the primary environment for student learning. In traditional courses students are taught for weeks at a time before they are tested individually on the material covered in the class. The exams pose as measurements of content retention and student success heavily depends on how well they do on them. By testing students in groups, educators can gain another form of providing learning gains. That is: the exam is not only utilized as a measurement tool, but an opportunity to solidify content. Testing students in groups allows for continued reinforcement of a similar environment as they learn in class. By applying the testing effect with short answer questions and including a group portion towards these exams students might gain substantial learning and retention.

Do Collaborative Exams Improve Content Retention?

Some studies have found that collaborative exams improve retention of course content (Cortright et al., 2003; Gilley and Clarkston, 2014; Ives et al., 2010; Cooke et al., 2019). These studies have included the following fields: physiology, earth and ocean science, and physics. While other studies have found that collaborative exams do not improve retention of course

content (Lusk and Conklin, 2003; Vojdanoska et al., 2010; Leight et al., 2012). These studies have included the following fields: nursing, psychology, and biology. Variability in the results may be due to one or more of the following: measuring retention at different times (3 days, Gilley and Clarkston, 2014; 7 days, Vojdanoska et al., 2010; 30 days, Cortright et al., 2003) using multiple choice (Gilley and Clarkston, 2014; Leight et al., 2012; Ives et al., 2010) versus open-ended questions (Cooke et al., 2019; Vojdanoska et al., 2010); not controlling for the test effect (Cortright et al., 2003; Lusk and Conklin, 2003; Leight et al., 2012); using a subsequent exam as a proxy for measuring retention (Cortright et al., 2003; Lusk and Conklin, 2003; Leight et al., 2012) versus using a "pop quiz" as a proxy (Gilley and Clarkston, 2014; Vojdanoska et al., 2010; Cooke et al., 2019).

Student motivation has long been studied in order to better understand what truly drives students to perform at their best. Students may often determine the work they are doing or need to do is not as meaningful and thus lead to work avoidance and an overall decrease in motivation (Seifert, 2004). Exams cause students to feel many different ways, to say the least. However, students must perform well on exams in order to demonstrate their understanding of the material. Patiwael and colleagues (2021) found that collaborative exams allow students to increase motivation by interaction, thinking for themselves, and active participation and that these are required for students to feel autonomous, competent, and relatable. Their study also found that teaching methods play an important role in learning and have a large impact on student motivation, engagement, retention, and achievement. The use of collaborative exams allows educators to introduce more ways for students to gain motivation by presenting students an opportunity to share what they know, test and correct their understanding, and listen to differing perspectives.

This thesis sought to determine whether multi-stage collaborative exams might improve student retention of course content in an upper division physiology course at UCSD. To assess a role for motivation and delaying feedback, we decided to run three experiments: a model similar to those employed previously (Cooke et al., 2019), a model to improve student motivation, and a model that employs delayed feedback. I will address each of these experiments separately before a general discussion of the entire work.

EXPERIMENT #1: Multi-stage collaborative exams using an already existing model.

Brief Intro/overview

To determine the effect of multi-stage collaborative exams on student retention, at the University of California San Diego, we tested undergraduate human physiology students using an adopted mixed-crossover design, similar to the design utilized in Cooke et al., 2019 study (figure 1).

METHODS

Course Context

BIPN 100: Human Physiology I, is an upper division-level course that introduces students to fundamental concepts in membranes and neurophysiology, cell signaling and endocrine system, and renal physiology.

Course Structure

The majority of the course was composed of sophomores and juniors seeking degrees in biology or life sciences-related disciplines. The study took place over the course of two academic quarters and the number of students who participated in all elements of the study are the following: Fall 2015 participated: 240, Spring 2016 participated: 111.

Research Design

A quasi-experimental crossover design was implemented to measure the effect of multi-stage collaborative exams on content retention. This design allows for each student to participate in both the control (individual retest) and treatment (collaborative retest) conditions (Figure 1). Exams were constructed into three phases that occurred in the following order: an individual test, an individual retest, and a group retest. The individual test was administered for 40 minutes and was composed of four short answer questions. The class was randomly split after the students submitted their individual tests; in which half the class received question 1 for individual retest and the other half received question 2 for individual retest. The second stage is controlling for exposure to the test questions a second time. Each student had 10 minutes to complete their individual retest question. Subsequently, the half that received question 1 for individual retest received question 2 as a group retest and the half that received question 2 for individual retest received question 1 as a group retest (treatment condition). Groups consisted of 3-5 students and were self-selected. Groups had 10 minutes to answer their group question.

Retesting the students twice after their individual tests provides a means for controlling the second exposure which resolves the issue of eliminating the influence of a testing effect. To evaluate content retention a retention test (post test) was administered to the students 3 weeks following the exam. The retention test was given as a pop quiz and the students were informed that the quiz would not impact their overall grade but should be viewed to self-evaluate their knowledge and understanding of the material in preparation for the final exam. The post test consisted of two isomorphic short answer questions associated with the topics tested experimentally during the individual test. Students were told that the second stage (individual re-write) would replace their original score on that question if it was higher than the original score.

Students were told that the group exam (stage three) would count for 10% of their exam score if it was higher than their original score on that question. That is: for both stage 2 and 3, the scores could not hurt the students' grades.

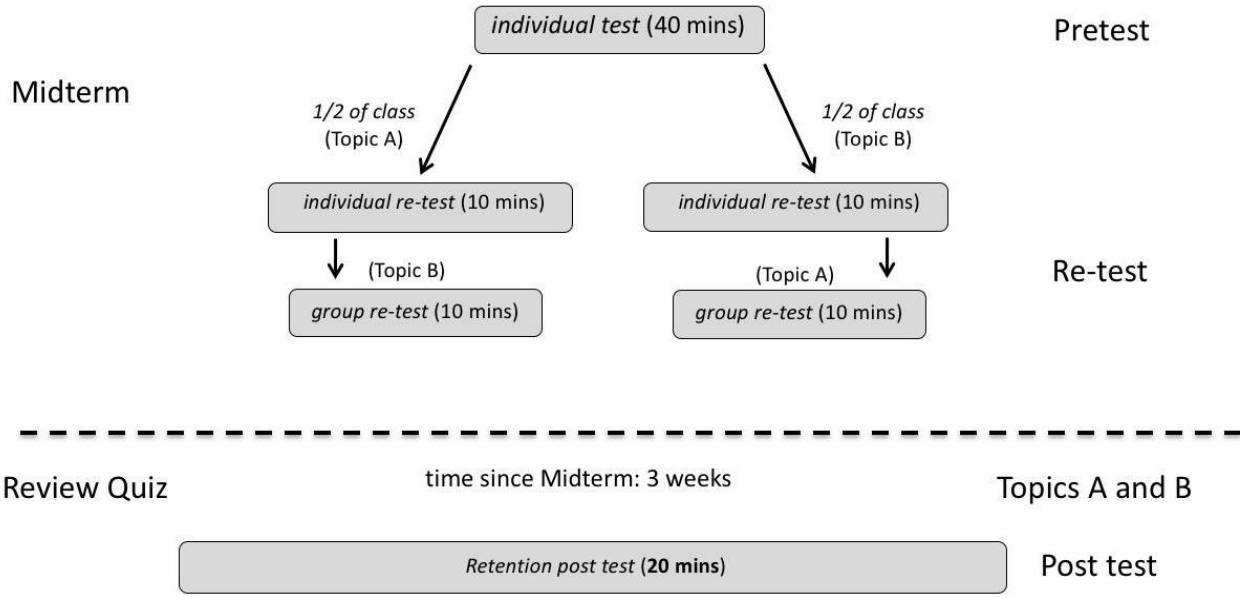

FIGURE 1. Adopted experimental design of study (Cooke et. al,. 2019).

RESULTS

Our study analyzed the difference in score between the retention test (Post-test) and midterm (Pre-test). The mean difference scores were lower on the post-test than on the pre-test in both treatment and control conditions (Figure 2). We evaluated whether group exams had higher retention than the control group utilizing a mixed effects model and found no significant difference between treatment and control conditions. With the possibility that there might be benefits to a particular group of students, we separated the students into low-, mid-, and high-performing students based on their individual test scores for the group exam questions. Again, we found no significant effect of group exams for low-, mid-, and high-performing students.

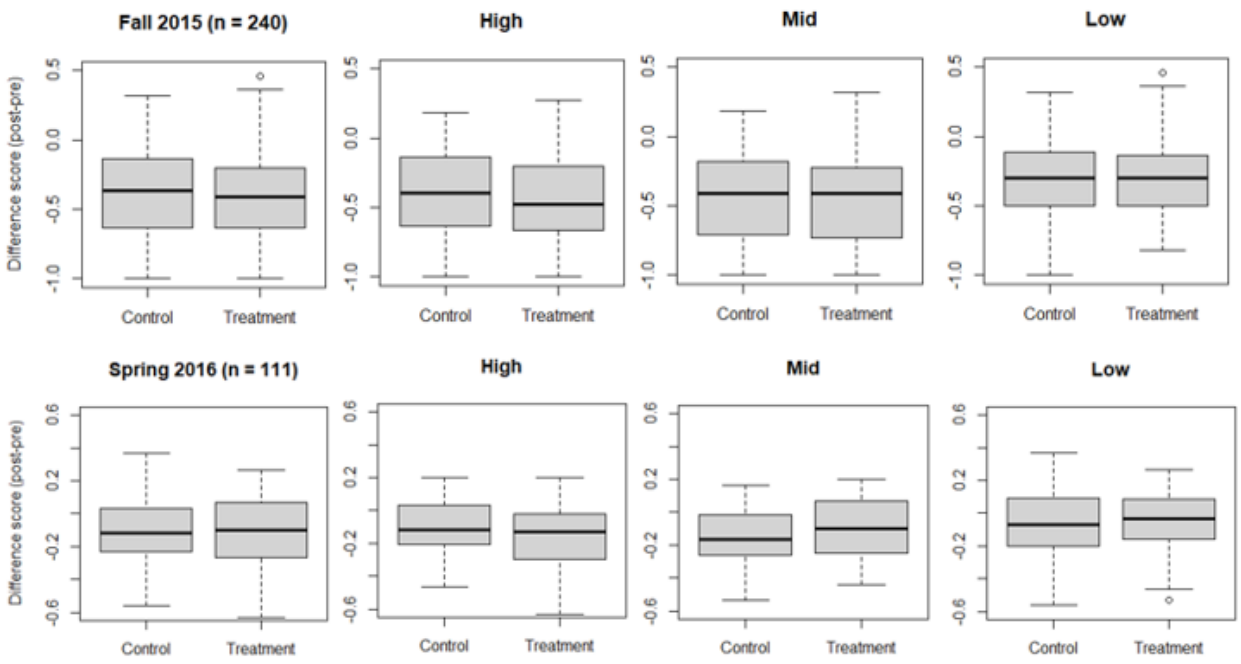

FIGURE 2. Average difference in scores (post – pretest) for midterm 1 of Fall 2015 (n = 240) and Spring 2016 (n = 111) quarters, using the mixed-crossover design. Left most graph: the class as a whole. Students' abilities ranked on their pre-test scores were classified into three competency groupings, High-, Mid-, Low-performing students.

DISCUSSION

The study for the Fall 2015 and Spring 2016 quarters suggests that collaborative exams have no effect on improving student retention of course content; the classes as a whole or low-, mid-, or high-performing students had no statistically significant differences between the group and individual scores. At a student-centered university, UBC, (Cooke and colleagues, 2019) measured retention of course content at 21 days and for the class as a whole and recorded a significant effect between group and individual scores for low-performing students. Our study aimed to replicate the study from UBC to determine the effects of group exams at a differing institution. Although the study conducted at UBC utilized the same experimental methods and design, our results were different. It is worth noting that the student population at UBC were

introductory biology students whereas we examined upper division physiology students. The differing student populations could be the reason our results were different. Based on previous literature studies with differing student populations found varied effects when it comes to multi-stage collaborative exams. Looking at similar studies conducted with undergraduate lower division students, (Leight et al., 2012 and Vojdanoska et al., 2010) both found that groups scored similarly to individuals deeming group exams not an effective tool for improving student retention. In contrast, (Cortright et al., 2003; Gilley and Clarkston, 2014) performed similar studies on undergraduate upper division students looking at group exam effectiveness and determined that group scores were significantly higher than the individual scores. While our study fits within the undergraduate upper division student population criteria, our findings diverge from these similar studies.

Another element in consideration is the campus culture in terms of education. At UBC the students were often exposed to group learning and group exams. Whereas UCSD students do not often experience group elements within their classes. The lack of exposure to group assessments or tasks could have made the students uncomfortable with the group portion of the exam. In addition, if UCSD students grew accustomed to individualistic classroom environments and roles then this could be an explanation for their effectiveness to work in groups for the exam.

The University of California, San Diego tends to house competitive STEM students looking towards professional programs and careers. Previous research on group dynamics and competitive students within these groups suggests that more competitive students have less motivation towards helping their peers. Hodges (2005) found that some students when placed into group exams may fear that their peers might not contribute as much or have the desire to

share what they know. Competitive students tend to recognize which students are freeloading off others and their motivation to contribute decreases.

In addition to student motivation, students often evaluate course assignments and tasks to determine the level of worth it has on their overall grade. Chevalier and colleagues (2018) determined that student participation substantially increases when assessments count towards the final grade. Our study does not place any value in terms of grade changes for the students' performance on the post-test evaluation, the pop quiz. Although the students were informed to do their best because similar questions would appear on the final exam and that the pop quiz could help them determine gaps in their understanding, student motivation for an assessment that was not effecting their overall grade could have led to the fact that we did not see improvement in retention. To increase incentives towards more fruitful group interactions, our study aims at increasing fruitful interactions by incentivizing students to contribute more with their groups.

EXPERIMENT #2: motivational tools to improve student effort during the post-test.

Brief Intro/overview

A common technique many students often implement is gauging how meaningful or impactful given tasks or assignments are towards their overall grade (Chevalier and colleagues, 2018). They found that student motivation towards their performance increases or decreases based on how much they believe these are worth. Experiment 1 results could have been due to the lack of motivation the students had on either the group portion or the post-test stage of the exam that was not going to increase their chance for a higher grade on the exam.

METHODS

Generally, the methods remained the same as experiment 1 but with a few key differences:

Course Structure

The study took place over the course of two academic quarters, Fall 2016 (n = 66) and Spring 2017 (n = 124).

Research Design

The research design was identical to that of experiment 1, only we provided two incentives to increase motivation: an altered grading scheme; and the post-test could replace the score of the original test question if it was higher. To impact group performance students were informed that the group portion of the exam could increase their original score for that question on the exam if they achieved a higher score as a group. The students' corrected score was derived from taking the difference between the group's score and the average score of all students' individual responses within that group. The difference was then added to the students' individual score for that particular question. The altered grading scheme allows for all students to benefit if they actively participate in increasing their performance in their groups.

RESULTS

Again, the mean difference scores between post-test and pre-test were lower on the post-test than the pre-test (Figure 3). There was no significant difference between treatment and control conditions in this high motivation model. We separated the students into low-, mid-, and high-performing students based on their individual test scores for the group exam questions, and again, we found no significant effect of group exams for low-, mid-, and high-performing students.

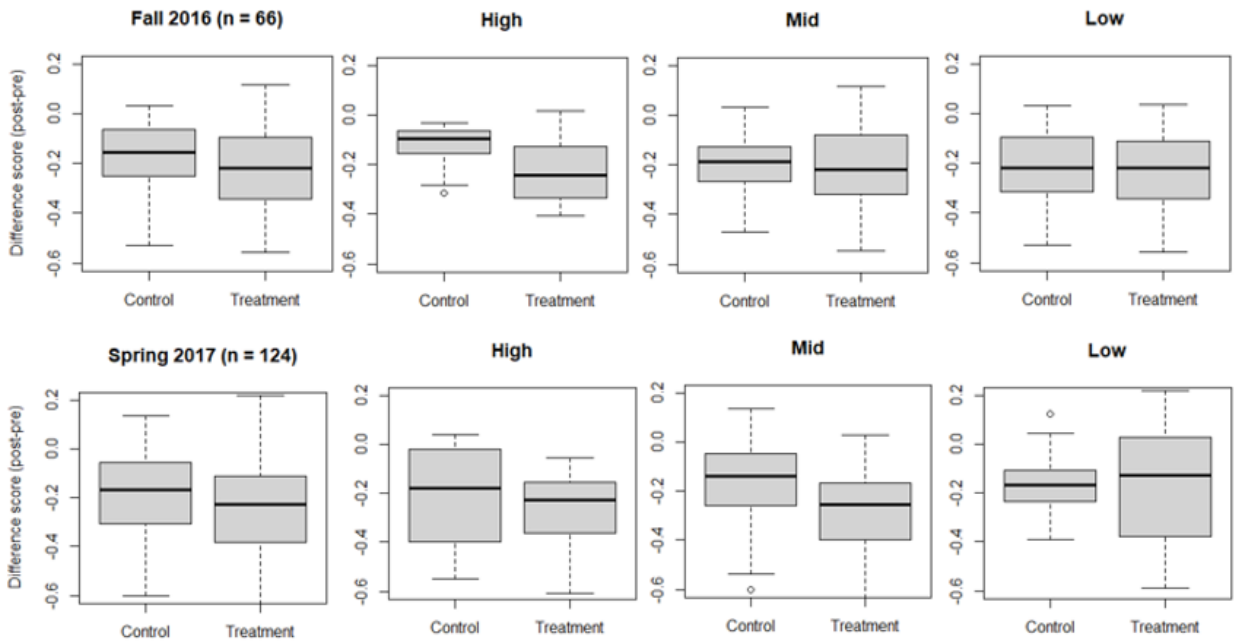

FIGURE 3. Average difference in scores (post – pretest) for Fall 2016 (n = 66) and Spring 2017 (n = 124) quarters. Left most graph: the class as a whole. Students' abilities ranked on overall performance in the course were classified into three competency groupings, High-, Mid-, Low-performing students. Outliers were plotted as individual points. The class as a whole and all three competence ranking students had no significant difference between control and treatment for both quarters.

DISCUSSION

With changes to some of the experimental conditions, student motivation was driven by incentives for students to increase their participation in their groups (a survey was conducted after the completion of the course and students reported these incentives increased motivation). Although the results clearly indicate that collaborative exams demonstrate no significant difference towards retention; based on observational evaluations and student testimonies the students appeared and reported higher levels of involvement. The literature points to student participation being an important factor for academic success (Douglas and Alemanne, 2007; Poole, 2000; Voelkl, 1995). Student participation has been widely researched and some key factors have been the role it plays in causing students to actively learn. As students participate

more their understanding of content can lead to retention. However, something that has yet to be researched as extensively is how much a student needs to participate in order to gain the academic benefits. Perhaps the level of participation had not yet met the amount necessary in order for longer lasting effects. In addition, student affect towards their contribution being important also plays a role in group participation (Meyers, 1997). Students' belief in understanding and worth often leads to whether a student will speak up and become an active member in a group or not. This phenomenon often goes undetected, and students base their group involvement by gauging their self-confidence and the understanding that others contribute to their group.

Meyers (1997), also pointed out that students must have incentives for contributing to the group. The incentives we added to the students' exam grade allowed for all performing level students to gain from increased participation in their groups. Even with the increased incentives in place, student retention did not improve. This could be due to the incentives not being high enough for the students to want to participate more. More so, the students' understanding of how the modified grading for the group portion could have led them to believe that the impact or weight on their overall grade was not worth their additional effort. Again, students often determine their effort based on the value they place on course assessments and tasks (Chevalier et al., 2018).

EXPERIMENT #3: Delaying the timing of the group portion of the exam.

Brief Intro/overview

Up to this point the students experienced feedback once they got into their groups. The literature on feedback has shown that feedback can greatly increase retention (Menekse et al., 2013); more

so, studies have found that the type of feedback students receive is of great importance. Corbett and (Anderson,2001) found that immediate feedback was effective for developing procedural skills, but that delayed feedback enhanced the transfer of learning into understanding the content. Both (Soderstrom and Bjork, 2015) and (Butler, 2007) indicated that delaying feedback was more beneficial for long-term course content retention relative to immediate feedback. Experiment 3 tests whether delaying the feedback would improve retention.

METHODS

Again, much of the experiment remained the same as the first two experiments but with a few changes to account for delayed feedback.

Course Structure

The study took place over the course of one academic quarter, Winter 2020 (n = 240).

Research Design

Conditions were identical to those in Experiment #2 except that there was a delay of ~48 hours between the midterm (pre-test) and the re-tests (both group and individual). The post-test was administered ~ 3 weeks after the re-tests.

RESULTS

Student retention was evaluated based on whether the treatment group had higher retention than the control group utilizing a mixed effects model and found no significant difference. With the possibility that there might be benefits to a particular group of students, we separated the students into low-, mid-, and high-performing students based on their individual test scores for the group exam questions. Again, we found no significant effect of group exams for low-, mid-, and high-performing students. The median difference scores in this experiment are much higher than those of Experiment #1 and Experiment #2. All previous experiments resulted in a

difference in scores median values below zero, whereas experiment 3 resulted in difference in scores median values above zero.

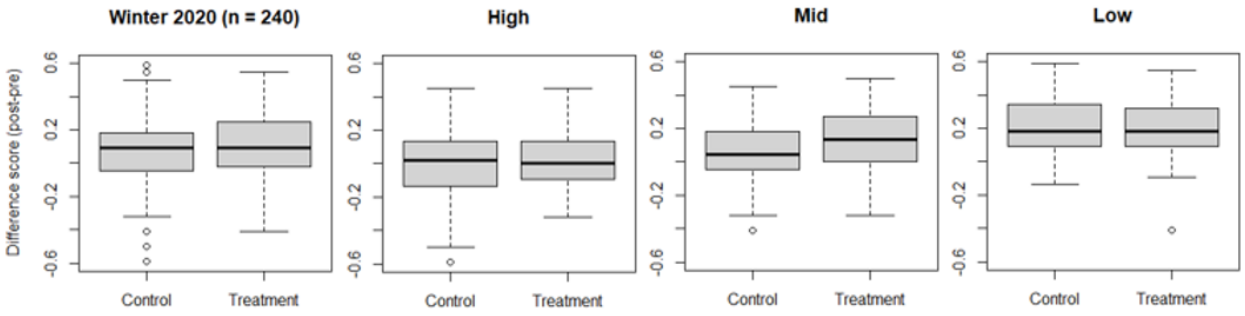

FIGURE 4. Average difference in scores (post – pretest) for Winter 2020 (n = 240) quarter. Left most graph: the class as a whole. Students' abilities ranked on overall performance in the course were classified into three competency groupings, High-, Mid-, Low-performing students. The students' median (difference in Post-test – Pre-test) scores depicted on average how much content was lost between the midterm (Pre-test) and the pop quiz (Post-test). Outliers were plotted as individual points. The class as a whole and all three competence ranking students had no significant difference between control and treatment for both quarters.

DISCUSSION

Student feedback up to this point in our experiments has been immediate, following the individual questions and by the means of group discussions. Testing benefits such as retention can result when students receive correct answer feedback (Kulik and Kulik, 1988; Bangart-Downs et al., 1991). Students were able to discuss their understanding of the content and engage with others to challenge their knowledge. However, with this feedback in place in the previous experiments, all students at different performing levels did not have improved retention (compared to controls) regardless of correct answer feedback from their groups. In contrast to the immediate feedback the third experiment engaged the students in a delay between individual responses and group responses. Delayed feedback can greatly improve student outcomes when

compared to immediate feedback (Butler et al., 2007). The delay does many things for students. It allows for students to solidify content, re-study material, and most importantly students can come to a comfortable understanding of the topics at hand before they speak with others. The delay could have led to students likely utilizing the time to re-study so that they could engage better with content during the re-test stage.

Data analysis needs to be further explored to determine if introducing a delay between the individual exams and the re-test questions retention has improved. In addition, this study should be cross analyzed with the previous experimental quarters to better understand why the medians for the class and the different performing students' levels were all above the previous quarters.

GENERAL DISCUSSION

Initially, we began our research to determine if collaborative exams, using similar conditions from another study, would yield improvements in retention for BIPN 100 UCSD students. Our findings determined that collaborative exams have no effect on improving student retention of course content even when we looked closer to see if they could be benefiting different performing level students. Several things could be the reason behind the results. The institutions house differing student populations and cultures. Some student contrasts include academic status and culture of collaboration. Even though the students only varied in a few years apart in terms of academic status this could have played a role in how the students would adapt to a collaborative exam. The culture of education differs at each institution. At UBC, students experienced collaborative classrooms and were used to working with others in assessments. Whereas, at UCSD students tend to be competitive, navigate classes independently, and rarely collaborate on exams. These contrasting academic cultures could be a reason why some students

adopt an exam with group elements versus students that lack prior exposure and tend to compete with their peers. By looking closer at the student culture, we questioned if they were motivated enough to actively participate in their groups.

Experiment two added to the first experiment by implementing incentives to increase student motivation in their groups. We turned attention towards increasing student performance by the means of increasing student participation. Many studies have found that academic success heavily depends on student participation (Douglas and Alemanne, 2007; Poole, 2000; Voelkl, 1995). Two factors went into adding incentives to the experiment; we knew the students were competitive and would closely monitor how their grades would be impacted. However, with the incentives in place, collaborative exams did not demonstrate significant improvements in retention. Possible reasons for these results are due to students still not participating to the level that they need to be, or they evaluated these incentives and deemed the task, exam question, not as worthy of their efforts.

In the final experiment we looked at student feedback and its timing. Up to this point we knew that student feedback was playing an immediate role. That is the students received immediate feedback once they got into their groups right after turning in their individual tests. While some literature finds that feedback leads to testing benefits (Kulik and Kulik, 1988; Bangart-Downs et al., 1991). Our previous experiments demonstrated that immediate feedback was not making the difference towards improvements in retention. However, feedback can be presented at different times. Butler et al., (2007) found that delayed feedback can improve student success when compared to immediate feedback. The delay could have caused the students to re-study the material, elicit peer to peer questions, and formulate structured responses to bring to their group. Although the results demonstrate no significant difference towards

retention for collaborative exams with delayed feedback they also demonstrate having effect on the students. Further data analysis needs to be conducted to determine these effects.

A few things could be the reason why group exams do not work at UCSD for BIPN100 students. The students tend to be competitive and above freshmen status. If the campus environment leans towards these competitive paradigms then their ability to perform in non-competitive assessments can be hard to change in a student that has experienced the culture for a few years. The students could determine these incentives to low for them to engage more in their groups. Lastly, the delayed feedback might hold the key to demonstrating improvements in retention. Delaying could create enough time for students to discuss the material and prepare for their group question. In future experiments it would be worth performing similar experiments at community colleges or primarily undergraduate institutions to investigate if students at institutions that promote collaborative learning environments demonstrate better outcomes.

REFERENCES

Adolphus, Telima & Alamina, Jane & Aderonmu, Temitope. (2021). The Effects of Collaborative Learning on Problem Solving Abilities among Senior Secondary School Physics Students in Simple Harmonic Motion. *Journal of Education and Practice* ISSN 2222-1735 (Paper) ISSN 2222-288X (Online) Vol.4

Albert T. Corbett and John R. Anderson (2001). Locus of feedback control in computer-based tutoring: impact on learning rate, achievement, and attitudes. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '01). Association for Computing Machinery, New York, NY, USA, 245–252. DOI:https://doi.org/10.1145/365024.365111

Chevalier, A., Dolton, P. and Lührmann, M. (2018), 'Making it count': incentives, student effort and performance. *J. R. Stat. Soc. A*, 181: 323-349. https://doi.org/10.1111/rssa.12278

Cooke, J. E., Weir, L., & Clarkston, B. (2019). Retention following Two-Stage Collaborative Exams Depends on Timing and Student Performance. *CBE life sciences education*, 18(2), ar12. https://doi.org/10.1187/cbe.17-07-0137

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advanced Physiological Education, 27*(3), 102–108.

Douglas, Ian & Alemanne, Nicole. (2007). Measuring student participation and effort. *IADIS International Conference on Cognition and Exploratory Learning in Digital Age*, CELDA 2007.

Felder, Richard & Brent, Rebecca. (1994). Cooperative learning in technical courses: Procedures, pitfalls, and payoffs. *ERIC Document Reproduction Service.*

Gilley, B., & Clarkston, B. (2014). Research and Teaching: Collaborative Testing: Evidence of Learning in a Controlled In-Class Study of Undergraduate Students. *Journal of College Science Teaching*, *043*(03). doi: 10.2505/4/jcst14_043_03_83

Hodges, L.C. (2004), Group exams in science courses. *New Directions for Teaching and Learning*, 2004: 89-93. https://doi.org/10.1002/tl.175

Johnson, D.W., R.T. Johnson and K.A. Smith. (1991). Cooperative Learning: Increasing College Faculty Instructional Productivity, *ASHE-ERIC Higher Education Report No. 4*, George Washington University.

Ives, J. (2015). Measuring the Learning from Two-Stage Collaborative Group Exams. *2014 Physics Education Research Conference Proceedings*. doi: 10.1119/perc.2014.pr.027

Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE—Life Science Education, 11,* 392–401.

Lusk, M. and Conklin, L. (2003) Collaborative Design to Promote Learning. *Journal of Nursing Education*, 42, 121-124.

Mcdaniel, Mark & Anderson, Janis & Derbish, Mary & Morrisette, Nova. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL. 19*. 494-513. 10.1080/09541440701326154.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education, 102*(3), 346–374.

Meyers, Steven A. "Increasing Student Participation and Productivity in Small-Group Activities for Psychology Classes." *Teaching of Psychology*, vol. 24, no. 2, Apr. 1997, pp. 105–115, doi:10.1207/s15328023top2402_5.

Nkechinyere, Okeke Modesta and Okey Kelechi Ordu. (2018). Impact of Collaborative Learning Strategy on the Academic Achievement of Senior Secondary School Chemistry Students in Obio-Akpor Local Government Area. *International Journal of Education and Evaluation* ISSN 2489-0073 Vol. 4 No. 2

Patiwael, J.A., Douma, A.H., Bezakova, N. et al. Collaborative testing in physical examination skills training and the autonomous motivation of students: a qualitative study. *BMC Med Educ* 21, 224 (2021). https://doi.org/10.1186/s12909-021-02618-7

Pendergrass, N., Kowalczyk, R., Dowd, J., Laoulache, R., Nelles, W., Golen, J., & Fowler, E. (2001). Improving first-year engineering education. *Journal of Engineering Education, 90*(1), 33–41.

Dawn M. Poole (2000) Student Participation in a Discussion-Oriented Online Course, *Journal of Research on Computing in Education*, 33:2, 162-177, DOI: 10.1080/08886504.2000.10782307

Purzer,S. (2011).The relationship between team discourse, self-efficacy, and individual achievement: A sequential mixed-methods study. *Journal of Engineering Education,100(*4), 655–679.

Randy Y. Hirokawa (1980). A comparative analysis of communication patterns within effective and ineffective decision-making groups, *Communication Monographs*, 47:4, 312-321, DOI: 10.1080/03637758009376040

Rao SP, Collins HL, DiCarlo SE. Collaborative testing enhances student learning. *Adv Physiol Educ*. 2002 Dec;26(1-4):37-41. doi: 10.1152/advan.00032.2001. PMID: 11850326.

Seifert, T. (2004) Understanding student motivation, *Educational Research*, 46:2, 137-149, DOI: 10.1080/0013188042000222421

Soderstrom NC, Bjork RA (2015). Learning Versus Performance: An Integrative Review. *Perspectives on Psychological Science*.10(2):176-199. doi:10.1177/1745691615569000

Van Eijl, P.J., Pilot, A. & Voogd, P.d. (2005). Effects of Collaborative and Individual Learning in a Blended Learning Environment. *Educ Inf Technol* 10, 51–65. https://doi.org/10.1007/s10639-005-6747-4

Kristin E. Voelkl (1995) School Warmth, Student Participation, and Achievement, *The Journal of Experimental Education*, 63:2, 127-138, DOI: 10.1080/00220973.1995.9943817

Vojdanoska, M., Cranney, J., & Newell, B. R. (2009). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology*, *24*(8), 1183–1195. doi: 10.1002/acp.1630