An Investigation of Knowledge Retention Using Two-Stage Exams in Undergraduate Biology

by

Abby Judge

A thesis submitted to the Department of Biological Sciences, University of Manitoba,
in partial fulfilment of the requirements for the course
BIOL 4100 (Honours Thesis)
for the degree of
Bachelor of Science (Honours)
April 2024

Abstract

Although active learning is typically thought of as an in-class experience, this concept can be further applied to assessments as well. Two-stage exams allow for the unique experience of students working collaboratively during examinations, following the completion of the individual stage. Although two-stage exams have been shown to improve student learning gains, the effects of two-stage exams on retention of course content remains variable. I examined the effects of two-stage exams on knowledge retention at various Bloom's level (Remember, Understand, and Apply) to determine whether two-stage exams promote knowledge retention at various levels of cognitive thinking. A two-stage in-class test followed by the re-testing of questions answered individually or in groups at subsequent time points (5, 48, and 85 days) allowed testing for knowledge retention. Our results indicate that group questions improve knowledge retention at relatively long time periods across all Bloom's levels while also promoting retention at more complex Bloom's levels at intermediate and relatively long time periods. Additionally, our analysis reveals that an average of 40% of individual only questions are forgotten by the final exam, while only an average of 19.5% of group questions are forgotten by the final exam. These results indicate that two-stage exams promote the ability to retain complex information at relatively long time periods.

Acknowledgements

I would like to thank Cassandra Debets and Kevin Scott for their continued support and encouragement throughout the entirety of my thesis project. I would also like to thank Celine Latulipe and John Markham for their valuable feedback and contributions to the final thesis.

Table of Contents

List of Tables

List of Figures

Introduction

Active learning is described as any interactive and engaging process that may include student collaboration and group work (Lombardi *et al.,* 2021). Active learning has been shown to improve student performance on examinations and promote student retention in Science, Technology, Engineering, and Mathematics students (Freeman *et al.,* 2014). Active learning can be further applied to assessments as well. Two-stage exams allow students to complete an exam individually and then proceed to write the same exam, or a similar version, in small groups. Two-stage exams allows for the unique opportunity where students collaborate during examinations and therefore become active participants in their learning journey during examinations. Two-stage exams have been shown to improve student learning gains measured by exam scores (Cortright *et al.,* 2003, Leight *et al.,* 2012, Gilley and Clarkston, 2014). Improved learning gains may help explain the positive student perceptions of two-stage exams that have been reported (Cortright *et al.,* 2003, Leight *et al.,* 2012) as well as the ability to minimize achievement gaps in historically underrepresented groups (Meaders and Vega, 2022). However, the effects two-stage exams have on knowledge retention at various levels of Bloom's taxonomy, a hierarchical system for classifying learning outcomes, remains poorly understood. Some STEM educators have reported that two-stage exams improve knowledge retention (Cortright *et al.,* 2003, Cooke *et al.,* 2019), while some report no significant improvements in knowledge retention (Leight *et al.,* 2012). Various factors including course level and exam format have been investigated and may influence student's retention and learning gains. Firstly, the level of course, either introductory

(Cooke *et al.,* 2019) or fourth-year students (Cortright *et al.,* 2003) may impact the results. If greatly dissimilar, the style of questions may also contribute to the variable results (multiple choice questions; Leight *et al.,* 2012 or open-ended questions; Cooke *et al.,* 2019). Additionally, retention may be tested at various times such as one week following a pre-assessment (Cooke *et al.,* 2019) or eight months later (Eastwood *et al.,* 2020). As a result, the impacts of two-stage exams on knowledge retention remains unclear.

Due to relatively few studies and the numerous factors that may influence knowledge retention, the impact of collaborative testing on retention in the literature remains inconclusive. In short term time periods, two-stage exams have shown to significantly improve retention measured after 3 days (Gilley and Clarkston, 2014) as well as one-to-two weeks (Ives, 2014). However, a study by Vojdanoska *et al.* (2003) studied retention and found no significant improvement in retention seven days after the assessment. In long term time periods, studies have shown that two-stage exams result in retention benefits when measured at four weeks (Cortright *et al.,* 2003) as well as eight months later (Eastwood *et al*., 2020). Meanwhile, others have found no benefit in retention after the implication of two-stage exams when measured at four weeks (Woody *et al.,* 2008) and six weeks (Ives, 2014). These conflicting results indicate the importance of continued investigation on this topic.

In general, student perceptions on two-stage exams are largely positive. These positive perceptions may be explained by students reporting deeper understanding of the material, improved student relations, and the ability to receive immediate feedback on their individual performance (Cortright *et al.,* 2003, Leight *et al.,* 2017). However, a small

proportion of students feel negatively affected by two-stage exams. These negative perceptions may be explained by feelings of confusion following the group stage and the student's ability to be convinced of the wrong answer (Eastwood *et al.,* 2020). There is also the possibility that improved exam scores in lower-performing students could be the result of higher-performing students leading the group-based discussion and making the final decision for the group (Khong and Tanner, 2020). Despite these infrequent negative perceptions and group discussions, two-stage exams have still been reported to improve student learning, especially in minority groups, which may help address achievement gaps in students historically underrepresented in science based on ethnicity or race (Meaders and Vega, 2022).

Bloom's taxonomy allows instructors to characterize the cognitive level of learning that students achieve on examinations. In introductory biology courses, many assessments are targeted at lower cognitive levels (Momsen *et al.,* 2010), leading to the memorization of facts rather than gaining the desired deeper understand of the material presented. Indeed, studies have found that males and middle/high socioeconomic status students are disproportionately favoured to answer questions correctly as Bloom's level increases (Wright *et al.,* 2017, Stanger-Hall, 2012). While the ability of two-stage exams to impact different levels of cognitive thinking has not been studied specifically, collaborative exams may allow for students of all backgrounds to achieve higher levels of cognitive thinking. Additionally, collaboration through two-stage exams may allow more students to retain content material at higher Bloom's levels regardless of their background.

With 40% of American students entering university with an interest in STEM programs, studies reveal that only 20% of those students graduate with STEM degrees (Freeman *et al.*, 2014). Therefore, it is critical that post-secondary institutions and faculty members promote the retention of students within STEM programs. In this study, I aim to build on previous literature to investigate the use of two-stage exams as it pertains to knowledge retention. I hypothesize that collaborative exam structure influences knowledge retention in first year biology students. I predict that questions answered as a group compared to individually answered questions, will improve student learning gains, improve knowledge retention, and promote higher levels of cognitive thinking.

Methods

Course description and assessment

This study was completed in two introductory biology class sections (BIOL 1020, A05 and A06) at the University of Manitoba in Fall 2023. BIOL 1020 is required by all Biological Sciences Major/Honours students and is prerequisite for all subsequent biology courses in the Department of Biological Sciences. As per the syllabus, students completed a multiple choice in-class test (September 28th), two midterm exams (October 5th and November 20th), and a final exam (December 22nd). The in-class test was completed in the two-stage format where students first answered 14 questions on the topic's osmosis, permeability, and transportation. The students were asked four questions from the Bloom's level 'Remember', six questions were asked from the Bloom's level 'Understand', and four questions were asked from the Bloom's level 'Apply'. Immediately afterwards, the students formed groups of three to four students and answered 7 of the original 14 questions in small groups, creating two types of question treatments (individual only and

individual + group referred to herein as individual and group questions, respectively). Similarly worded questions from the in-class test were then asked on the first stages of three subsequent exams to test for retention and reduce the effects of repeated exposure. On each subsequent exam, three individual questions and three group questions from the in-class test were retested (**Figure 1**). By the completion of the final exam, every question from the in-class test was retested once or twice. Additionally, for both treatment groups on each re-test, one question from each Bloom's level (Remember, Understand, and Apply) was re-examined. During these subsequent examinations, a question on the topic that was never seen on the in-class test was asked as a control question and to measure the effects of repeated exposure. Additionally, the second midterm exam does not include the topics above as traditional testable topics and the students were not informed of the re-test prior to the exam, therefore, each re-tested question on the second midterm acted as a bonus question for grading purposes. Despite the midterms and final exam also taking place in the two-stage format, the re-tested questions were only asked on the individual portion of these exams.

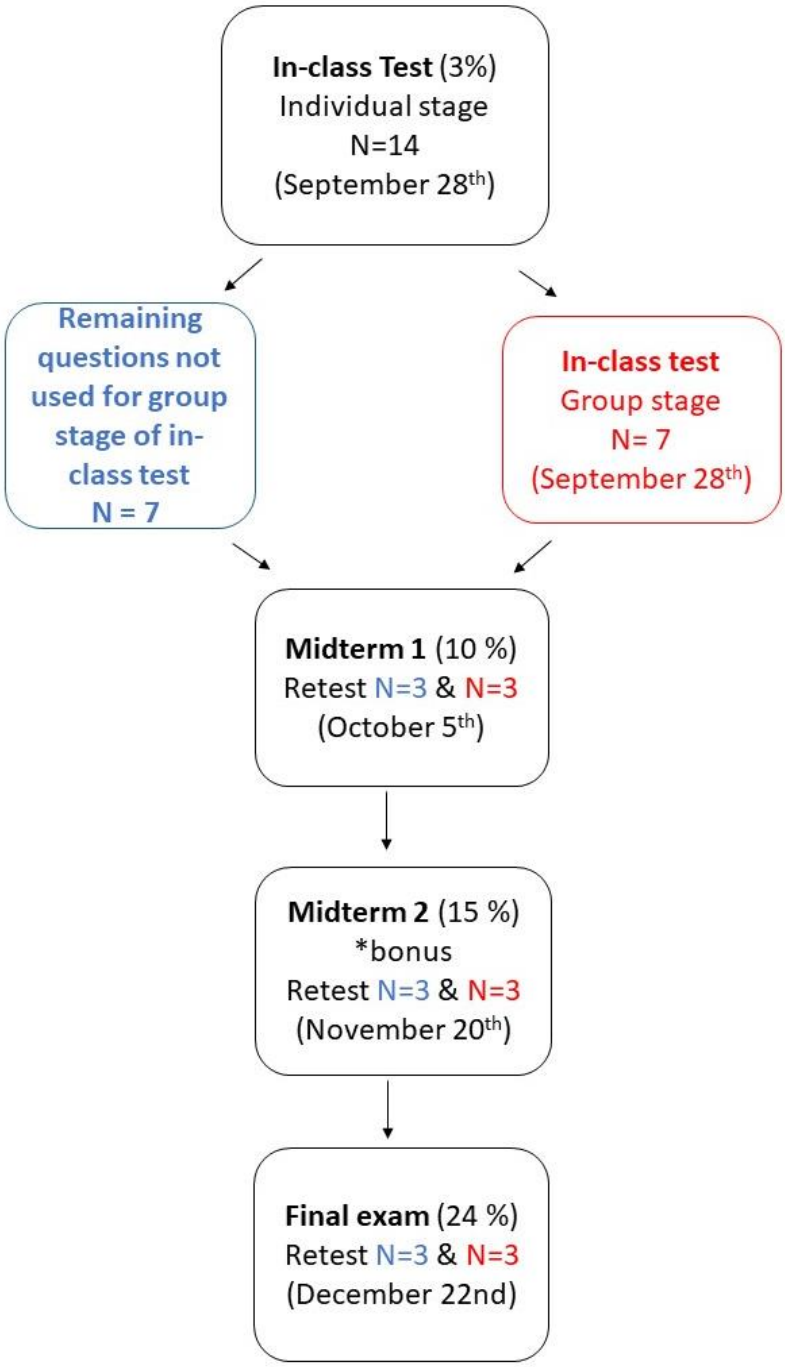


Figure 1. Experimental design of the study. Includes all tests and exams with percentages of course weight taken by students throughout the semester. The individual stage of the in-class test consisted of 14 questions, followed by a retest of 7 of the 14 questions in small groups. Subsequent exams retested three questions answered from the individual stage and three questions from the group stage of the in-class test.

Data collection

The experimental protocol is approved by the Manitoba Research Ethics Board (HE2023-1905) for collection of student data. Students were recruited through a recruitment letter and completed a Likert style in-class survey to collect student perceptions on two-stage exams and student demographics (see Appendices) on November 28th. The recruitment letter assured the students that their student information and test scores would be kept anonymized for research purposes. Additionally, the students had the option to provide consent to take part in this study at this time. Only those students who consented and completed every assessment had their test scores, perceptions, and demographics analyzed.

After completion of the in-class test, the individual question treatment allowed us to compare student scores to the group question treatment. Learning gains were assessed by comparing the proportion of students who answered the questions correctly in the individual and the group stage questions. The performance on questions from either the individual or group treatment was then assessed again on subsequent examinations to investigate for knowledge retention throughout the semester. Student scores from the re-tested questions on the midterms and final exam were compared to the scores on the in-class test. All questions were asked in a multiple-choice format, resulting in correct or incorrect responses with no partial mark opportunities. Additionally, student demographics allowed us to determine if specific groups of students were more highly impacted by the two-stage exams compared to other groups of students.

Statistical analysis

The student scores were obtained and coded as correct or incorrect for each question. The in-class test results were analyzed using a T-test statistical analysis to determine the differences between the individual and group treatments. Additionally, an ANOVA statistical analysis was used to determine the difference in score at each examined Bloom's level. To measure the effects of retention in each treatment, the average score of the individual and group questions on the midterms and final exam were calculated. An ANOVA statistical analysis was used to measure the difference of scores on each assessment in each treatment. Additionally, to measure student learning gains more accurately, the average for each question on each subsequent re-test was normalized back to the original average score of the question on the individual stage of the in-class test. This analysis was completed using the equation = ((X/QTI)*100)-100 where X is the average score of the question on the retested assessment while QTI is the average score of the question on the individual portion of the in-class test.

Retention analysis

By analyzing the correctness of a specific question throughout the semester, a student may fall into one of five categories per question. The first category is defined by the student retaining the information by answering the question correctly throughout the semester. The student may answer the question correctly on every assessment through the semester (In-class test, individual (TI)-In-class test, group (TG)-Midterm (M)-Final (F); 1-1-1-1) or the student may answer the question incorrectly on the individual stage of the in-class test and then proceed to answer the question correctly on the following

assessments (TI-TG-M-F; 0-1-1-1). Additionally, retention can be defined as student answering the question correctly on the individual stage of the in-class test, incorrectly on the group stage of the in-class test, and then correctly again on the midterm and final exam retest (TI-TG-M-F; 1-0-1-1) as the group may have decided on the incorrect answer, however, the individual still retained the correct information.

The second category is defined by the student studying for the question. In this category, the student answered the question incorrectly on the in-class test and the midterm then answered the question correctly on the final exam (TI-TG-M-F; 0-0-0-1). Secondly, the student answered the question incorrectly on the in-class test, then proceed to answer the question correctly on the midterm and final exam retest (TI-TG-M-F; 0-0-1-1). Finally, the student may answer the question correctly on the in-class test, forgot the information on the midterm and answered the question incorrectly, and then proceeded to answer the question correctly on the final exam by studying (TI-TG-M-F; 1-1-0-1).

In the third category, the student forgot the information. The student may have answered the question correctly on the in-class test then proceeded to answer the question incorrectly on the midterm and final exam retest (TI-TG-M-F; 1-1-0-0). Additionally, the student may have answered the question correctly on the in-class test and the midterm but answered the question incorrectly on the final exam (TI-TG-M-F; 1-1-1-0). Finally, the student may have answered the question incorrectly on the in-class test, proceed to answer the question correctly on the midterm retest and then proceed to answer the question incorrectly by the final exam (TI-TG-M-F; 0-0-1-0).

The fourth category was defined as the group carrying the individual to answer the question correctly. The student only answered the question correctly during the group stage of the in-class test (TI-TG-M-F; 0-1-0-0). Lastly, if a student did not answer the question correctly at least once throughout the semester, they do not know the correct information (TI-TG-M-F; 0-0-0-0). The student's response histories may comprise of multiple pathways when analyzing all the questions, therefore an average proportion of students representing each pathway for every question was determined. The average proportion of each pathway that defines one of the five categories was summed and then averaged to determine the average proportion of questions that was retained or forgotten. A Chi-squared statistical analysis test was used to determine the difference in proportion on each reassessment.

Table 1. Summary of retention pathway definitions. TI = In-class test, individual stage. TG = in-class test, group stage. M = midterm. F = final exam.

Category	Group question path (TI-TG-M-F)	Individual question path (TI-M-F)
Retained	1-1-1-1	1-1-1
	0-1-1-1	
	1-0-1-1	
Studied	0-0-0-1	0-0-1
	0-0-1-1	0-1-1
	1-1-0-1	1-0-1
	1-0-0-1	
Forgot	1-0-0-0	1-0-0
	0-1-0-0	0-1-0
	1-1-0-0	1-1-0
	1-1-1-0	
	0-0-1-0	
	0-1-1-0	
Group carried	0-1-0-0	
Does not know	0-0-0-0	0-0-0

Statistical Analysis

R Studio (2023) was utilized for various statistical tests, including a t-test, ANOVA, chi-squared, and logistic regression functions. I used a logistic regression model to statically analyze the student scores in R Studio. The logistic regression determined the probability of a discrete outcome when given an input variable. With various factors that could influence the correctness of student scores, including the individual or group treatment, the date of the assessment (in-class test, midterms, or final exam), the various Bloom's levels (Remember, Understand, and Apply), the number the student was exposed to a question, and whether the student studied for the assessment, a logistic regression best captures the probability of a student answering a question correctly or incorrectly while accounting for these factors. In this model the student ID was a random effect while the correctness of answers, treatment, date, Bloom's level, exposure, and studying were all fixed effects. By using student ID as a random effect, it removed any influence of differences among students from the fixed effects that were not directly associated with the tests. Initially, the effects of all variables were analyzed to examine the overall effects of the group stage on individual performance. Subsequent models only measured the effects of specific variables and how they impacted the proportion of questions answered correctly.

Results

For the group questions on the in-class test, students scored 22.94% higher on the group stage than individual stage ($T_{160} = 6.7$, $p = < 0.001$) (**Figure 2**). Additionally, during the individual stage, the individual questions scored significantly higher than the individual

questions that would then be answered again during the group stage (T_{160} = -13.07, p = <0.001). When separating the in-class test responses into the Bloom's levels that were analyzed, similar results were observed. At every Bloom's level tested, average group scores were significantly higher than the average individual scores (**Figure 3**). The Bloom's level Understand saw the largest change in score between the two stages at 26.18%.

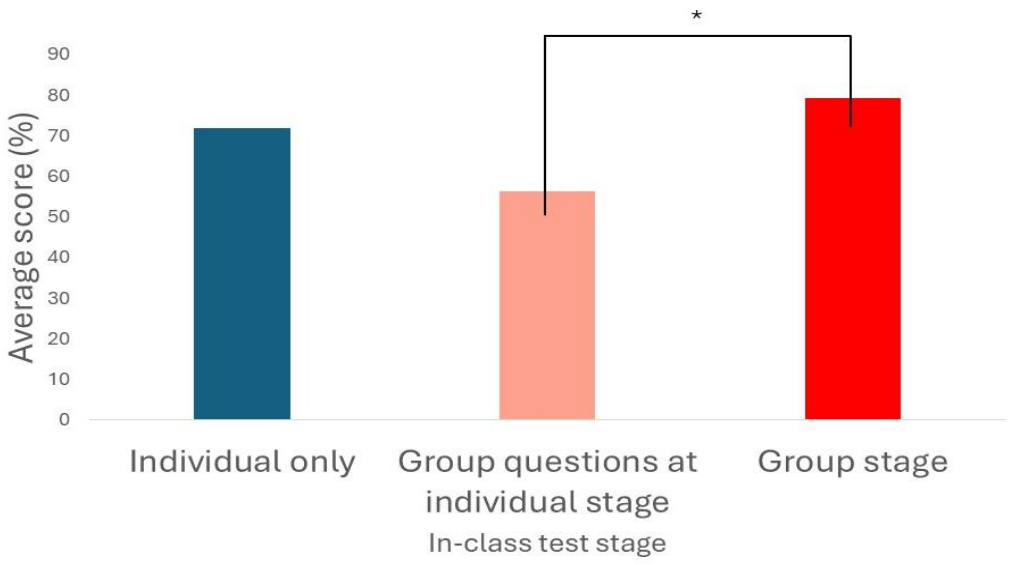


Figure 2. The average test score (%) for individual questions, individual stage questions that would be asked on the group stage, and the collaborative group stage score of the in-class test. Asterix indicates a significant difference with 95% confidence. ($p < 0.001$, $n = 160$)

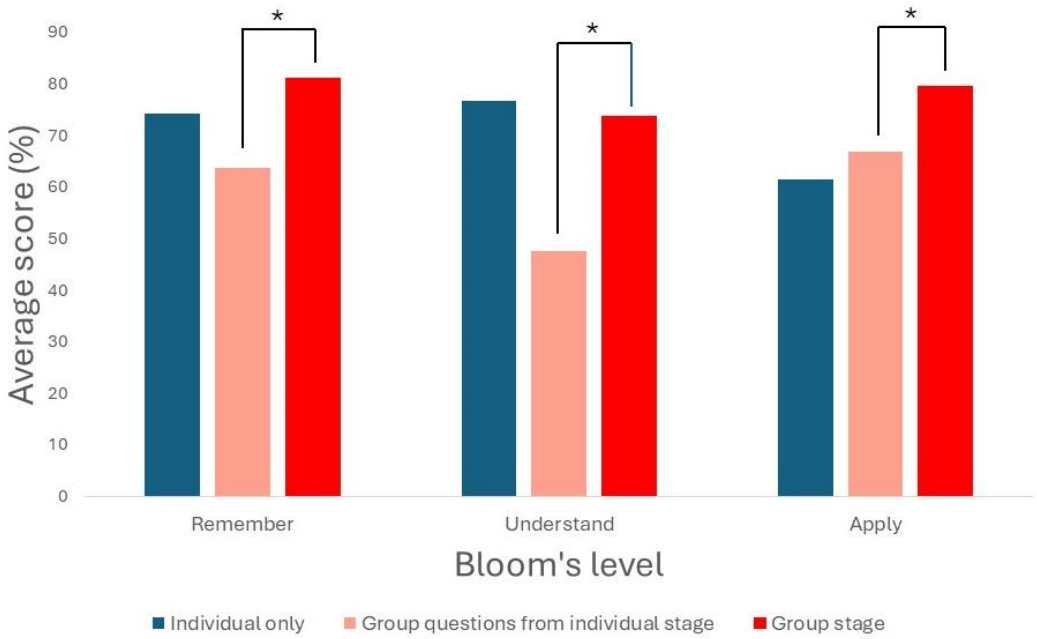


Figure 3. The average test scores (%) for individual only questions, individual only questions that would be asked on the group stage, and the collaborative group stage score of the in-class test. Questions are separated based on Bloom's level hierarchy Remember ($p < 0.001$), Understand ($p < 0.001$), and Apply ($p < 0.001$). Asterix indicates a significant difference with 95% confidence. (n=160)

Individual and group stage scores were also analyzed throughout the semester to test for retention. On the reassessment of questions on the final exam (day 85), the group questions performed significantly better than the individual questions at the Bloom's level Remember ($p < 0.001$) and Apply ($p < 0.001$) (**Figure 4**). In addition to the final exam within the Apply Bloom's level, the group stage questions also resulted in a higher average score on midterm two (day 53) at 64% compared to the individual questions with a score of 48% ($p < 0.001$) (**Figure 4B**).

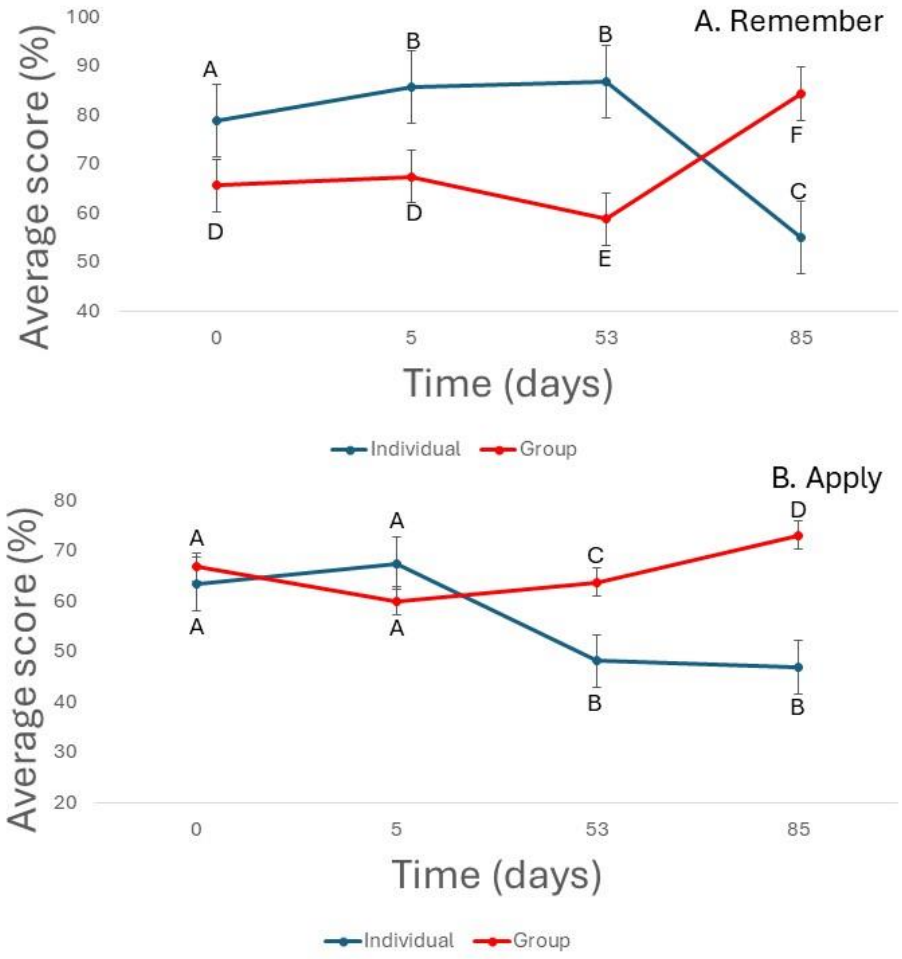


Figure 4. The average test scores (%) of individual only (blue) and group stage (red) questions at time points throughout the semester. Questions are separated into Bloom's hierarchy levels Remember (A) and Apply (B). Error bars are SE. (N=160)

By normalizing the average score change of the re-tested scores to the original in-class test scores from the individual stage, it is revealed that students performed better on group questions compared to individual questions at all three Bloom's levels on the final exam. The normalized re-tested group scores at the Bloom's level Remember also improved from the in-class test scores on midterm one (**Figure 5A**). On the final exam, the average score change differential between the group and individual questions was 74.5%.

When normalizing the re-tested scores at the Bloom's level Apply, the group stage average score change on the final exam was higher than the original in-class test average score, however, the individual average score change had a larger improvement (**Figure 5B**). Additionally, the group stage questions maintained a nearly negligible change in score on midterm two while the individual only questions resulted in a decrease in average score by 44.8%.

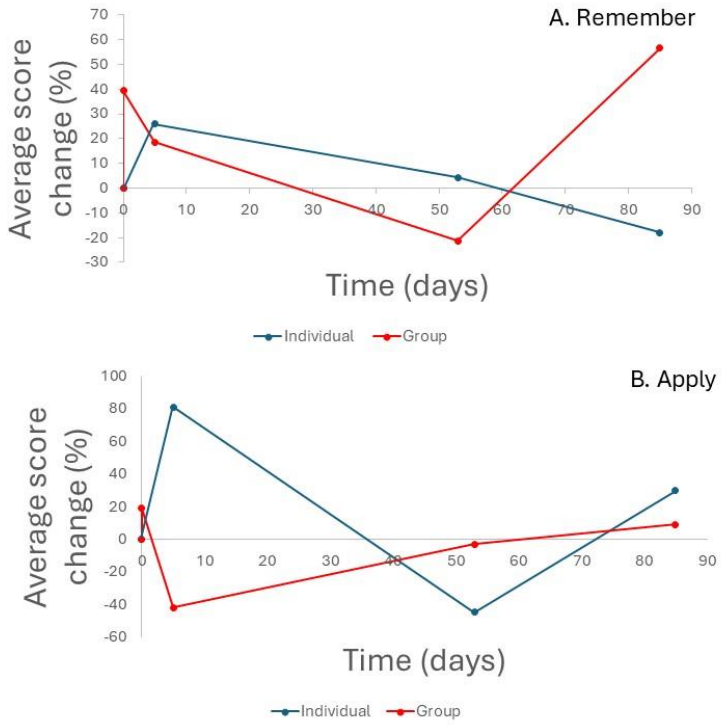


Figure 5. The average score change (%) of individual (blue) and group (red) questions of assessments throughout the semester. Questions only include Bloom's level Remember (A) and Apply (B). Positive values represent an increase in average score while a negative value represents a decrease in average score. (n=160)

When analyzing the possible retention pathways throughout the semester, group questions had a higher proportion of students retain the concepts on midterm one ($X^2_{(1,\ 160)} = 66.8$, $p < 0.005$) and the final exam ($X^2_{(1,\ 160)} = 97.8$ $p < 0.005$) compared to the

individual questions (**Figure 6A**). By the final exam, 28.1% of students answering individual questions retained the information, while 45.5% of students answering group questions retained the information. The proportion of forgotten questions throughout the semester was also analyzed by our retention pathways, revealing that students were less likely to forget group questions on the final exam compared to individual questions ($p < 0.005$) (**Figure 6B**). On the final exam, students forgot individual questions 9.2% more than the group questions.

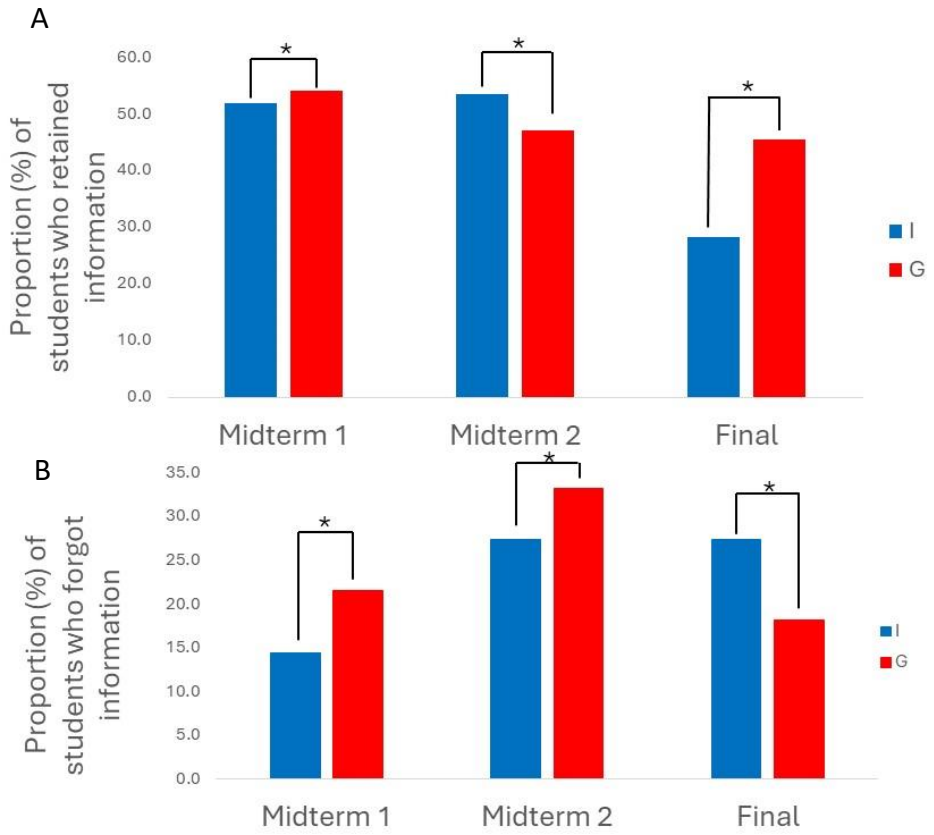


Figure 6. The average proportion of students who retained (A) and forgot (B) individual (blue) and group (red) questions according to our definition of retention in the retention pathway analysis. Includes assessments on midterm one ($p < 0.005$), midterm two ($p < 0.005$), and the final exam ($p < 0.005$). Asterix indicates a significant difference with 95% confidence. (n=160)

By analyzing self-reported student demographics from the in-class survey, it was revealed that all students benefitted from the group stage of the in-class test. While significant differences between individual and group questions in student cohorts were negligible (gender, indigenous status, familial academic history, nationality), all cohorts still experienced higher average scores in re-tested group stage questions compared to re-tested individual only questions. (**Figure 7**). Additionally, the students who reported Indigenous status saw an improvement in average score by 7.2% on the group stage questions compared to the individual only questions (T_{342} = -1.404, p = 0.1612) (**Figure 7B**).

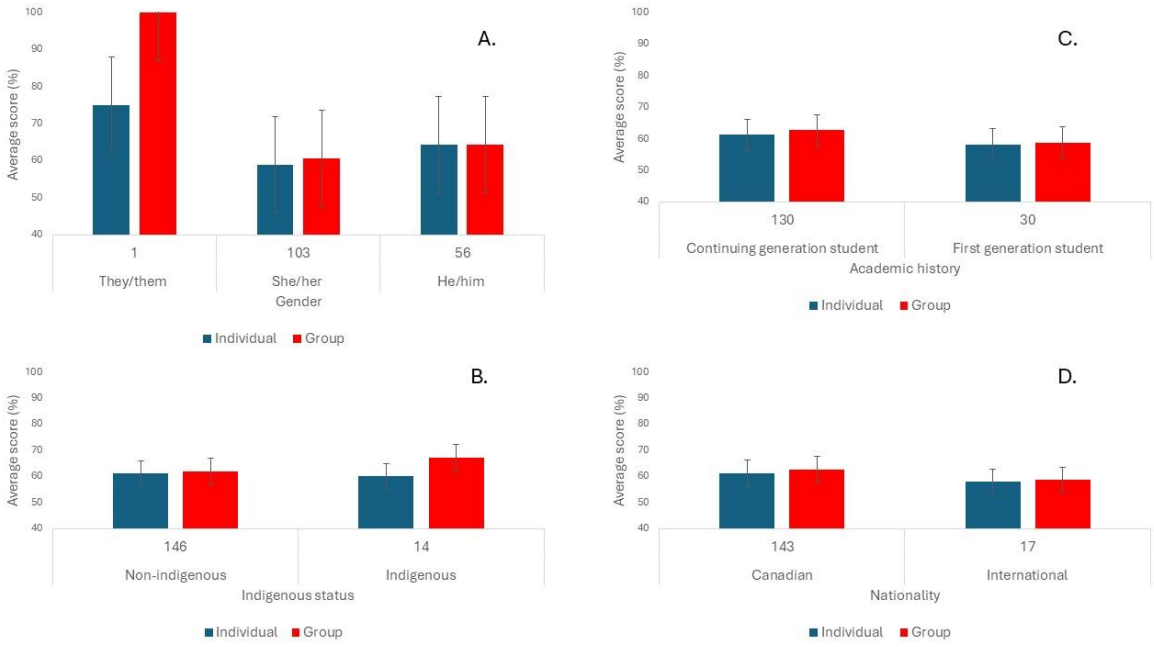


Figure 7. The average score of re-tested individual questions and re-tested group questions throughout the semester separated by self-reported demographics. Cohorts include A. Gender, B. Indigenous status, C. Academic history, D. Nationality. Value in each category represents the number of individuals.

Overall, student perceptions towards two-stage exams were positive **(Table 2)**. The in-class survey revealed that 70% of students agree that two-stage exams help them retain

the information **(Figure 8A)**. Additionally, many students reported that students contributed equally to the group stage portion of the exam while many students disagreed that some group members may unfairly benefit from the exam **(Figure 8 B & C)**.

Table 2. Summary of Likert-style survey responses regarding two-stage exams (n=250).

Statement	% Strongly Agree	% Agree	% Neutral	% Disagree	% Strongly Disagree
Working with other students helps me learn.	31	47	18	5	1
I enjoy working with the group during the group stages of assessments.	37	35	18	8	2
The group stage of the exam helped me understand the concepts more clearly than if we did not have the group part of the exam.	34	39	15	9	3
Group exams should be used in other courses.	49	25	16	6	4

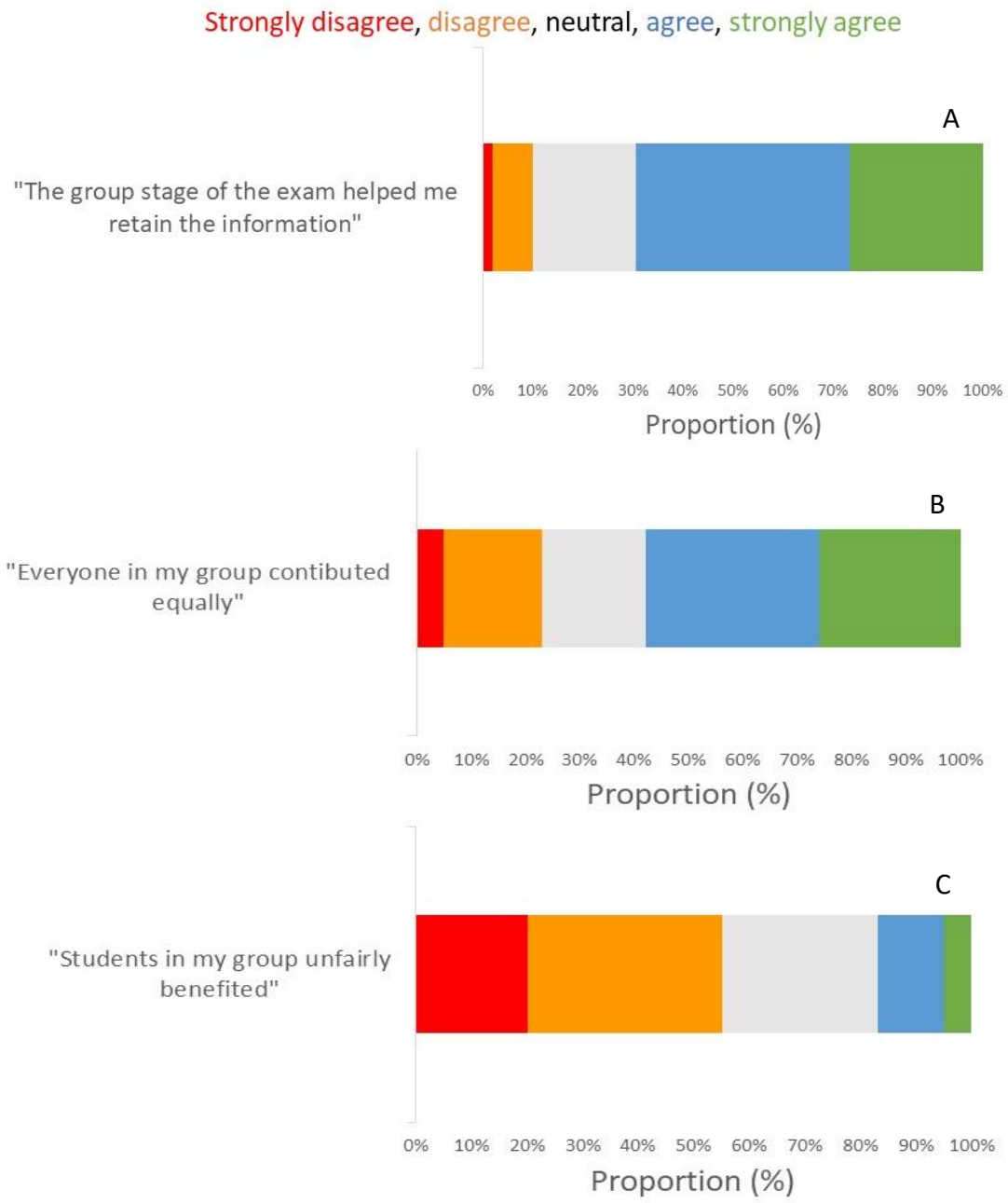


Figure 8. Proportion of student responses of "strongly agree, agree, neutral, disagree, and strongly disagree" to the prompted statements "Everyone in my group contributed equally to the group part of the exam" (A), "Students in my group unfairly benefited from the group part of the exam" (B), and "The group stage of the exam helped me retain the information." (C) (n=250)

Questions from the individual stage of the in-class test and re-tested questions from midterm one, midterm two, and the final exam were included in a logistic regression model in R Studio. The results revealed that models with the fixed variables "Treatment, Bloom's, Exposure, Date, Study, and Student ID" (**Table 3.**) and fixed variables "Gender, Nationality, Indigenous status, and Academic history" (**Table 4.**) were the most appropriate models.

Table 3. Various R Studio models, log ik, and AIC values.

Model	Log ik	AIC
Treatment + Bloom's + Exposure + Date + Studied + |Student ID	-3853.2	7728.4
Treatment + Bloom's + Exposure + | Student ID	-3895.7	7807.4
Bloom's + Exposure + | Student ID	-3897.7	7809.5
Treatment + Exposure + | Student ID	-3952.1	7916.2
Treatment + Bloom's + | Student ID	-3972.4	7954.9
Date + Studied + | Student ID	-3983.5	7978.9

Table 4. Various R Studio student demographic models, log ik and AIC values.

Model	Log ik	AIC
Gender + International + Indigenous + Firstgen + | Student ID	-3870.9	7755.7
Gender + Indigenous + | Student ID	-3889.4	7788.7
Gender + Firstgen	-3896.4	7802.8
Gender + International + | Student ID	-3932.7	7875.5
International + Indigenous + | Student ID	-3986.1	7980.3
International + Firstgen + | Student ID	-3993.5	7995.0

Discussion

Our results indicate that two-stage exams improve knowledge retention at relatively long time periods, especially at higher levels of Bloom's taxonomy. When investigating knowledge retention at the Remember Bloom's level, group questions had a higher average score than individual questions only on the final exam re-test at 12 weeks. Additionally, group questions at the Apply Bloom's level had a higher average score than

individual questions when measured 6.5 and 12 weeks later. While our results from the Remember Bloom's level align with Ives (2014) who also measured retention at six weeks and saw no improvement in retention, our results from the Apply Bloom's level do not follow the same pattern. This difference from the Apply level may be explained by the types of questions used in the Ives (2014) study which did not elaborate on the types of questions used despite specifying the use of multiple-choice questions as well as ensuring the clarity of questions through graduate student revisions. Although not a direct comparison, observed retention benefits of two-stage exams at four weeks more closely follows our results obtained at the Apply Bloom's level which resulted in improved group scores on both midterm two (7.5 weeks) and the final exam (12 weeks) (Cortright *et al.,* 2003). Unfortunately, the types of questions used in the Cortright *et al.* (2003) study was not specified, making a direct comparison difficult. Meaders and Vega (2022) were some of the first to identify Bloom's hierarchy in the questions tested on. However, Meaders and Vega (2022) asked higher order Bloom's taxonomy questions only on the group stage of the exam that may have benefitted from collaborative effort, and as a result did not find significant improvement from the application from two-stage exams in two out of three exams. Additionally, our study's unique measurement of retention at 12 weeks makes it difficult to produce direct comparisons. However, the results of two-stage exams improving retention at relatively long time periods is supported by Eastwood *et al.* (2020) who observed that retention was higher in group stage questions when measured eight months later. These results indicate that two-stage exams improve knowledge retention at relatively long time points across all Bloom's levels, and especially higher levels of

cognitive thinking. These results reveal the importance of collaborative examinations in facilitating higher order thinking skills. However, I suggest that these results are interpreted with caution as the final exam was weighted more heavily than other assessments and may have promoted increased levels of studying resulting in the increased scores compared to the original in-class test.

Following two-stage exams, very few studies have measured knowledge retention at multiple time points, making it difficult to observe retention over time. Interestingly, two study designs have investigated retention at multiple time points found opposing results. Cooke *et al.* (2019) found that two-stage exams improve retention at relatively long time periods (three weeks) with no benefits at short time periods (one week). Conversely, Ives (2014) found that two-stage exams benefits retention at short time periods (one week) but not at long time periods (seven weeks). Throughout our study, I expected a general decrease in retention at each time point measured following the in-class test based on the decay of memory performance over time (Sayre and Heckler, 2009). During the instruction of a topic, student scores rapidly increase, however, once the instruction period has ended and the semester continues, students experience an exponential decay in exam scores (Sayre and Heckler, 2009). The exponential decay in memory indicates how important yet difficult it is to retain information for long-time periods. Indeed, the results revealed a large reduction in the proportion of students who retained information in the final exam, especially in individual only questions, while the proportion of students retaining information was relatively constant on both midterms. Following a similar pattern, I observed a larger proportion of students forget information

on midterm two and the final exam. The larger proportion of students forgetting information on midterm two compared to the final exam may be explained by the midterm's surprise factor, where students did not know they were going to be tested on the topic in investigation for this study. Nonetheless, group questions were forgotten at a lower proportion than individual questions on the final exam, supporting our hypothesis that two-stage exams improve knowledge retention. Additionally, collaborative assessment also promotes retention at relatively long time periods, potentially defying the exponential decay in memory.

The results revealed that all cohorts of students had higher average scores on re-tested group stage questions compared to re-tested individual stage questions, indicating that group questions benefits student learning gains in all students. Meaders and Vega (2022) investigated how students historically underrepresented based on ethnicity or race were impacted by two-stage exams. There were no significant differences on the group exam scores for these groups of students (Meader and Vega, 2022). Additionally, students who did not identify as part of a historically underrepresented group received higher average final grades (Meaders and Vega, 2022). The present study investigated the effects of two-stage exams on four specific cohorts of students and found similar results, therefore, two-stage exams may not proportionally benefit any specific group of students. However, some have also reported that lower-performing students achieve larger learning gains due to collaborative exam implementation compared to higher-performing students (Khong and Tanner, 2020, Cooke et al., 2019). Regardless of personal background or academic standing, all students still benefit from two-stage exams by experiencing larger

learning gains. Like Meaders and Vega (2022), I recognize that equity gaps may still exist despite the implementation of two-stage exams, however, I can report that two-stage exams benefit all groups of students.

Our in-class Likert-style survey allowed us to gather student perceptions on two-stage exams which revealed positive results. The survey results revealed that 73% of students agree that two-stage exams improve concept clarity, while 74% of students agree that two-stage exams should be used in other courses. These positive reviews align with many other sources of literature (Cooke *et al.,* 2019; Himbeault and Latulipe, 2023; Rempel *et al.,* 2023). The variety of questions that may be questioned on a survey further validifies that students positively perceive two-stage exams for multiple reasons. Some report that two-stage exams are positively perceived due to a deeper understanding of the material (Himbeault and Latulipe, 2023) while others report positive perceptions based on strengthened relationships with peers (Rempel *et al.,* 2023). The present study found that students agree that two-stage exams help them retain information as well as improve clarity on the topic. These positive perceptions indicate that students do not only enjoy two-stage exams based on the notion that two-stage exams improve performance, but rather due to the collaborative style of learning while also making new relationships. Taken together, the data suggest that two-stage exams offer benefits beyond grade performance and the retention of material throughout the semester.

Conclusion

With an increased requirement for individuals in STEM fields, it is critical that educators utilize strategies, such as two-stage exams, to meet this demand. Two-stage exams improve knowledge retention at relatively long time periods, especially at higher levels of Bloom's taxonomy. Additionally, two-stage exams allow for the combined effect to improve learning gains, knowledge retention, and the ability to make impactful relationships with peers. However, the wide variety of methodologies within two-stage exam testing makes it difficult to directly compare the present study to others in a similar field, especially within retention. I encourage future studies to identify the level of Bloom's hierarchy that they are examining to provide further support that two-stage exams and collaborative efforts improve knowledge retention at higher levels of cognitive thinking. Additionally, I encourage other researchers to investigate how small groups are managed and how order and control are determined within these groups.

Literature cited

Cooke, J. E., Weir, L. K., & Clarkston, B. (2019). Retention following Two-Stage Collaborative Exams Depends on Timing and Student Performance. CBE- Life Sciences Education. 18(2).

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

Cortright, R. N., Collins, H. L., & DiCarlo, S. E. (2005). Peer instruction enhanced meaningful learning: ability to solve novel problems. Advances in Physiology Education. 29(2): 107– 111.

Eastwood, J., Kleinberg, K., & Rodenbaugh, D. W. (2020). Collaborative Testing in Medical Education: Student Perceptions and Long-Term Knowledge Retention. Medical Science Educator. 30(2): 737–747.

Freeman, S., Eddy, S. L., McDonough, M. J., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences of the United States of America. 111(23): 8410–8415.

Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching*, *43*(3), 83-91.

Ives, J. (2014). Measuring the learning from two-stage collaborative group exams. *arXiv preprint arXiv:1407.6442*.

Himbeault, L., & Latulipe, C. (2023). Using a Two-Stage Final Exam in an Intro CS Course: Student Perceptions and Grade Impacts. In *Proceedings of the 25th Western Canadian Conference on Computing Education* (pp. 1-2).

Khong, M. L., & Tanner, J. A. (2020). A collaborative two-stage examination in biomedical sciences: Positive impact on feedback and peer collaboration. Biochemistry and Molecular Biology Education, 49(1), 69–79.

Leight, H., Saunders, C., Calkins, R. A., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a Large-Enrollment Introductory Biology class. CBE- Life Sciences Education. 11(4): 392–401.

Lombardi, D., & Shipley, T. F. (2021). The curious construct of active learning. Psychological Science in the Public Interest. 22(1): 8–43.

Meaders, C. L., & Vega, Y. (2023). Collaborative Two-Stage Exams Benefit Students in a Biology Laboratory Course. *Journal of Microbiology & Biology Education*, *24*(1), e00138-22.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory Undergraduate Biology courses focus on Low-Level Cognitive Skills. CBE- Life Sciences Education, 9(4), 435–440.

Rempel, B., McGinitie, E., & Dirks, M. (2023). The Influence of Two Stage Collaborative Testing on Peer Relationships: A Study of First Year University Student Perceptions. *Canadian Journal for the Scholarship of Teaching and Learning*, *14*(2), 10.

Sayre, E. C., & Heckler, A. F. (2009). Peaks and decays of student knowledge in an introductory E&M course. *Physical Review Special Topics-Physics Education Research*, *5*(1), 013101.

Stanger-Hall, K. F. (2012). Multiple-Choice exams: an obstacle for Higher-Level thinking in introductory Science classes. CBE- Life Sciences Education, 11(3), 294–306.

Vojdanoska, M., Cranney, J., & Newell, B. R. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. Applied Psychology, 24, 1183-1195.

Woody, W. D., Woody, L. K., & Bromley, S. (2008). Anticipated group versus individual examinations: A classroom comparison. *Teaching of Psychology*, *35*(1), 13-17.

Wright, C. D., Eddy, S. L., Wenderoth, M. P., Abshire, E. T., Blankenbiller, M., & Brownell, S. E. (2017). Cognitive difficulty and format of exams predicts gender and socioeconomic gaps in exam performance of students in introductory biology courses. CBE- Life Sciences Education, 15(2), ar23.

Appendices

Appendix 1: Question Version History

Question 1 ICT_i ICT_g **MT2**

Which of the following is *true* of osmosis? #remember

A. Osmosis only occurs in red blood cells.
B. Osmosis is an energy-demanding or "active" process.
C. In osmosis, water moves across a membrane from areas of lower solute concentration to areas of higher solute concentration.
D. In osmosis, solutes move across a membrane from areas of lower water concentration to areas of higher water concentration.

Which of the following is true of osmosis? #remember

A. Osmosis requires ATP
B. In osmosis, water move across a membrane from areas of higher solute concentration to areas of lower solute concentration
C. In osmosis, solutes move across a membrane from areas of higher solute concentration to areas of lower solute concentration
D. In osmosis, water moves across a membrane from areas of lower solute concentration to areas of higher solute concentration

Question 2 ICT_i **MT2**

What kinds of molecular pass through a cell membrane most easily? #remember

A. small and uncharged
B. small and charged
C. large polar
D. monosaccharides such as glucose

What properties allow molecules to pass through the membrane easily? #remember

I. Being small
II. Being polar
III. Being charged
IV. Being uncharged
V. Being large

A. I and III
B. I and IV

C. II and V
D. II, III, and V

Question 3 **ICT_i** **ICT_g** **MT1** **F_i** **F_g**

A non-penetrating solute moves across a membrane down it's concentration gradient by which of the following mechanisms? #remember

A. simple diffusion
B. facilitated diffusion
C. active transport
D. antiport

A non-penetrating solute moves across a membrane up its concentration gradient by which of the following mechanisms? #remember

A. Facilitated diffusion
B. simple diffusion
C. symport
D. active transport

Which of the following statements is correct about facilitated diffusion? #understand

A. It is very rapid over long distances.
B. It requires an expenditure of energy by the cell.
C. It is a passive process in which molecules move from a region of higher concentration to a region of lower concentration through a transmembrane protein.
D. It is a passive process in which molecules move from a region of higher concentration to a region of lower concentration on its own.

Question 4 **ICT_i** **MT1** **F_i** **F_g**

Which of the following best describes a hypertonic solution? #remember

A. A solution that causes water to move into the cell.
B. A solution that causes water to move out of a cell.
C. A solution with a higher concentration of all solutes than another.
D. A solution with a higher concentration of penetrating solutes than another.

Which of the following best describes a hypotonic solution? #remember

A. A solution that causes water to move into the cell.
B. A solution that causes water to move out of a cell.
C. A solution with a higher concentration of all solutes than another.
D. A solution with a higher concentration of penetrating solutes than another.

Which of the following best describes an isotonic solution? #remember

A. A solution that causes water to move into the cell.
B. A solution that causes water to move out of a cell.
C. A solution with the same number of penetrating solutes than a cell.
D. A solution with the same number of nonpenetrating solutes than a cell.

Question 5 **ICT_i** **ICT_g** **MT2** **F_i** **F_g**

Which of the following statements *correctly* describes the normal tonicity conditions for typical happy plant and animal cells? #understand

A. The animal cell is in an isotonic solution, and the plant cell is in a hypertonic solution.
B. The animal cell is in a hypertonic solution, and the plant cell is in an isotonic solution.
C. The animal cell is in an isotonic solution, and the plant cell is in a hypotonic solution.
D. The animal cell is in a hypertonic solution, and the plant cell is in a hypotonic solution.

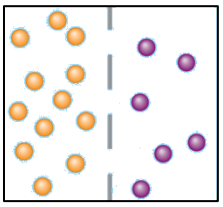
Which of the following solutions cause an animal cell to burst, but makes a plant cell happy? #understand

A. Isotonic solution
B. Hypertonic solution
C. Hypotonic solution
D. A solution that would cause an animal cell to burst and make a plant cell happy does not exist

Which of the following statements *correctly* describes the two tonicity conditions that would make an animal cell crenated and a plant cell turgid? #understand

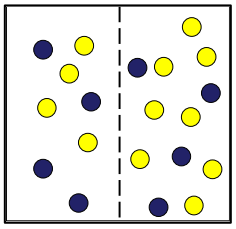
A. The animal cell is in an isotonic solution, and the plant cell is in a hypertonic solution.
B. The animal cell is in a hypertonic solution, and the plant cell is in an isotonic solution.
C. The animal cell is in an isotonic solution, and the plant cell is in a hypotonic solution.
D. The animal cell is in a hypertonic solution, and the plant cell is in a hypotonic solution.

Question 6 ICT_i MT2 F_i F_g



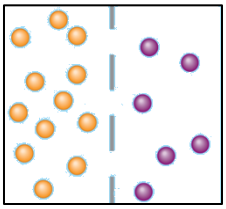
The dashed line is a semipermeable membrane, and the dots are all non-penetrating solutes. As it is drawn, water will move _____. #understand

A. to the right
B. to the left
C. there will be no net movement



Consider the diagram to the right, where the dashed line is a semipermeable membrane and the dots are all penetrating solutes. As it is drawn, water will move _____. #understand

A. To the right
B. To the left
C. Equally in both directions, there will be no net water movement



The dashed line in the diagram to the right is a semipermeable membrane, and both the light and dark dots are penetrating solutes. As it is drawn, water will move _____. #understand

A. to the right
B. to the left
C. there will be no net movement

Question 7 ICT_i ICT_g MT2

Which of the following most accurately describes selective permeability? #understand

A. An input of energy is required for transport across the membrane.
B. There must be a concentration gradient for molecules to pass through a membrane.
C. Only certain molecules can cross a cell membrane.
D. Amphipathic molecules pass through a membrane.

Which of the following most accurately describes selective permeability #understand

A. An input of energy is required for a molecule to cross the membrane.
B. Amphipathic molecules pass through a membrane.
C. There must be a concentration gradient for molecules to pass through a membrane.
D. Only certain molecules can cross a cell membrane.

Question 8 ICT_i ICT_g MT2

The phosphate transport system in bacteria imports phosphate into the cell even when phosphate concentration outside the cell is much lower than the cytoplasmic phosphate concentration. Phosphate import requires proton movement due to a pH gradient across the membrane—more acidic outside than inside the cell. Phosphate transport is an example of _____. #understand

A. facilitated diffusion.
B. primary active transport.
C. antiport.
D. symport.

The H^+/Sucrose co-transporter in plant cells allows the uptake of sucrose even when its concentration outside the cell is much lower than the cytoplasmic sucrose concentration. Sucrose uptake occurs because the co-transporter allows H^+ to diffuse down its electrochemical gradient into the cell. Sucrose transport is an example of _____. #understand

A. Facilitated diffusion
B. Primary active transport
C. Antiport
D. Symport

Question 9 ICT_i MT1 F_i F_g

Which of the following statements is correct about diffusion? #understand

A. It is very rapid over long distances.
B. It requires an expenditure of energy by the cell
C. It is a passive process in which molecules move from a region of higher concentration to a region of lower concentration
D. It is an active process in which molecules move from a region of lower concentration to one of higher concentration.

Which of the following statements is correct about diffusion? #understand

A. It is very rapid over long distances.
B. It requires an expenditure of energy by the cell.
C. It is a passive process in which molecules move from a region of higher concentration to a region of lower concentration.
D. It is an active process in which molecules move from a region of lower concentration to one of higher concentration.

A non-penetrating solute moves across a membrane alone and against its concentration gradient by which of the following mechanisms? #remember

A. simple diffusion
B. facilitated diffusion
C. active transport
D. antiport

Question 10 **ICT_i** **ICT_g** **MT1** **F_i** **F_g**

Which of the following would likely move through the lipid bilayer of a plasma membrane most rapidly? #understand

A. CO_2
B. an amino acid
C. glucose
D. K^+

Which of the following would likely move through the lipid bilayer of a plasma membrane most rapidly? #understand

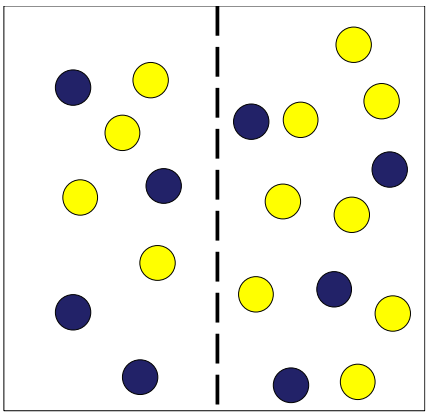
A. O_2
B. H_2O
C. Sucrose
D. Na^+

Which of the following would likely move through a plasma membrane most rapidly? #understand

A. CO_2
B. an amino acid
C. glucose
D. K^+

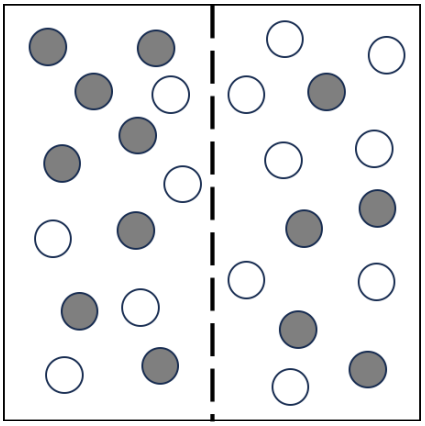
Question 11 ICT_i MT1 F_i F_g

The dashed line is a semipermeable membrane, the dark dots are penetrating solutes, and the light dots are non-penetrating solutes. Which of the following predictions is correct. #apply



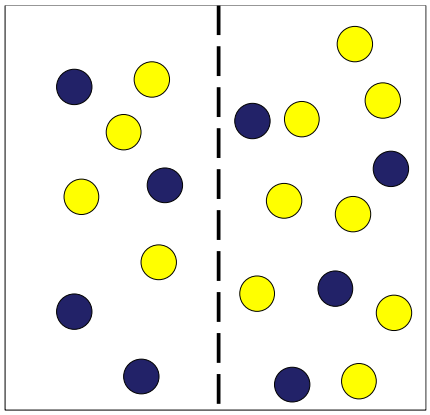
A. The left is solution is hypertonic to the right and water will move to the right.
B. The left is solution is hypotonic to the right and water will move to the right.
C. The left is solution is hypertonic to the right and water will move to the left.
D. The left is solution is hypotonic to the right and water will move to the left.

The dashed line is a semipermeable membrane, the dark dots are non-penetrating solutes, and the light dots are penetrating solutes. Which of the following predictions is correct. #apply



A. The left is solution is hypertonic to the right and water will move to the right.
B. The left is solution is hypotonic to the right and water will move to the right.
C. The left is solution is hypertonic to the right and water will move to the left.
D. The left is solution is hypotonic to the right and water will move to the left.

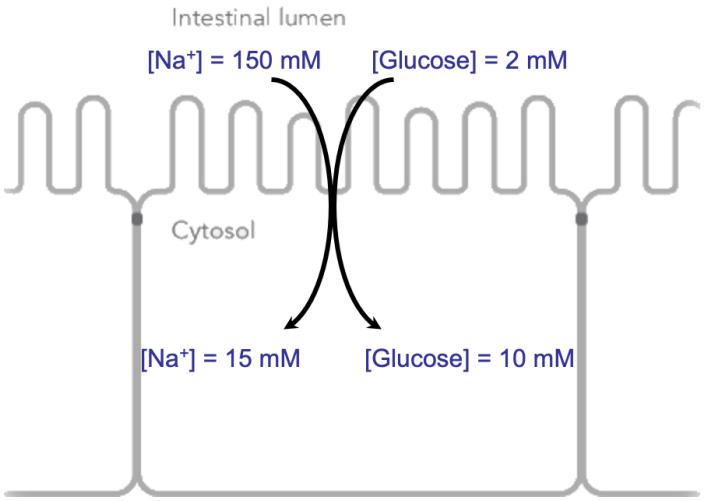
The dashed line is a semipermeable membrane, the dark dots are penetrating solutes, and the light dots are non-penetrating solutes. Which of the following predictions is correct. #apply



A. The left is solution is hypertonic to the right and water will move to the right.
B. The left is solution is hypotonic to the right and water will move to the right.
C. The left is solution is hypertonic to the right and water will move to the left.
D. The left is solution is hypotonic to the right and water will move to the left.

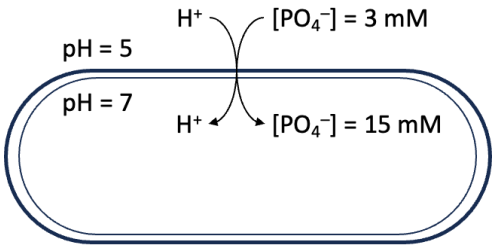
Question 12 ICT_i ICT_g MT2 F_i F_g

Given the diagram below, which of the following is most likely *true* of a protein that cotransports glucose and sodium ions into the intestinal cells of an animal between meals? #apply



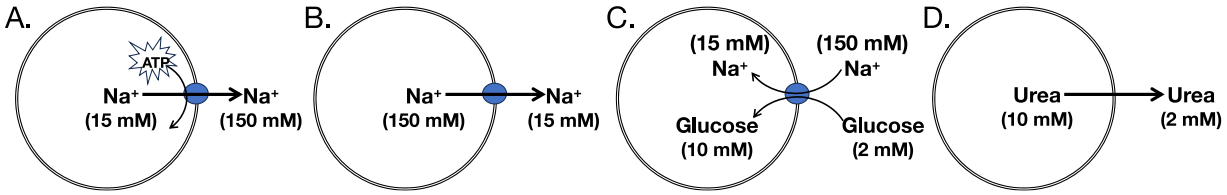
A. The sodium ions are moving up their electrochemical gradient while glucose is moving down.
B. Sodium ions entering the cell down the concentration gradient provides energy for uptake of glucose against its gradient.
C. Sodium ions can move down their electrochemical gradient through the cotransporter whether glucose is present outside the cell or not.
D. The cotransporter can also transport potassium ions.

Given the diagram to the right, which of the following is most likely true during the cotransport of phosphate (PO_4^-) and protons (H^+) into a bacterium? #apply



A. The protons are moving up their electrochemical gradient while phosphate is moving down.
B. Protons can move down their electrochemical gradient through the cotransporter whether phosphate is present outside the cell or not.
C. Protons entering the cell, down their concentration gradient, provides energy for the uptake of phosphate against its gradient.
D. Phosphate entering the cell, down its concentration gradient, provides energy for proton uptake against its gradient.

Which of the following diagrams shows secondary active transport? #apply



Question 13 **ICT_i** **ICT_g** **MT1**

Which of the following statements describes the situation if a freshwater algal cell with a cytoplasmic osmolarity of 400 mM was placed in a solution containing 150 mM NaCl and 150 mM of urea. #apply

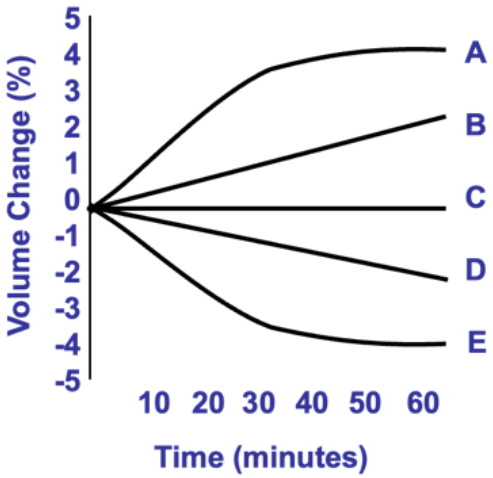
A. The solution is hypoosmotic and hypotonic to the cell, causing water to enter the cell.
B. The solution is hyperosmotic and hypertonic to the cell, causing water to leave the cell.
C. The solution is hyperosmotic but hypotonic to the cell, causing water to leave the cell.
D. The solution is isosmotic but hypotonic to the cell, causing water to enter the cell.

If a saltwater organism with an internal osmolarity of ~1000 mM, was moved into fresh water, what would be the results? #apply

A. Water would rush into the organism because it is hypotonic to the new environment.
B. Water would leave the organism because it is hypertonic to the new environment.
C. Since the organism is isotonic to the new environment, there will be no net water movement.
D. Water would rush into the organism because it is hypertonic to the new environment.

Question 14 **ICT$_i$** **MT2**

Five artificial cells that are impermeable to sucrose, with various internal concentrations of sucrose are placed in separate beakers containing an initial concentration of 0.6 *M* sucrose solution. At 10-minute intervals, the cells' volumes were measured and the percent change of each was graphed to the right. Which line in the graph represents the cell that contained a solution isotonic to the 0.6 *M* solution at the beginning of the experiment? #apply



Artificial cells with membranes impermeable to sucrose are placed in five sucrose solutions (A – E) of different concentrations. The cells' volumes were measured at 10-minute intervals and the percent change of each was graphed to the right. Which line in the graph represents the solution that is hypotonic to and reached equilibrium with the artificial cell? #apply

Appendix 2: Survey and demographic questions

Strongly disagree= 1, disagree=2, neutral=3, agree=4, strongly agree=5

1. Working with other students helps me learn.

2. I enjoyed working with the group during the group stages of assessments.

3. The group stage of the exam helped me understand the concepts more clearly than if we did not have the group part of the exam.

4. Group exams should be used in other courses.

5. Students in my group unfairly benefited from the group part of the exam.

6. Everyone in my group contributed equally to the group part of the exam.

7. The group stage of the exam helped me retain the information.

8. Level of interest Not at all=1, not much=2, somewhat=3, very=4,

9. Preferred pronouns
a. She/her
b. He/him
c. They/them
d. Prefer not to say

10. Are you an international student?
a. Yes
b. No

11. Your home country is or is located in
a. Canada
b. USA
c. Mexico
d. South America
e. Europe
f. Eurasia
g. Africa
h. Indian Subcontinent
i. Asia

12. Do you identify as Indigenous?
a. Yes
b. No
c. Prefer not to say

13. Are you a first-generation university student? (The first person in your family to attend post-secondary education.)
a. Yes
b. No
c. Prefer not to say