

**THE USE OF GROUP TESTS TO PROMOTE COLLABORATION
AND LEARNING: DO THEY WORK?**
**L'UTILISATION DE TESTS DE GROUPE POUR PROMOUVOIR LA
COLLABORATION ET L'APPRENTISSAGE : EST-CE QU'ILS
FONCTIONNENT?**

Kimberley Ann Gilbride

Volume 55, Number 1, Winter 2020

Regards croisés de la recherche sur l'accompagnement professionnel
en milieux scolaire et universitaire
Research on Professional Development in School and University
Settings: Intersecting View

URI: <https://id.erudit.org/iderudit/1075728ar>

DOI: <https://doi.org/10.7202/1075728ar>

[See table of contents](#)

Publisher(s)

Faculty of Education, McGill University

ISSN

1916-0666 (digital)

[Explore this journal](#)

Cite this article

Gilbride, K. (2020). THE USE OF GROUP TESTS TO PROMOTE COLLABORATION
AND LEARNING: DO THEY WORK? *McGill Journal of Education / Revue des
sciences de l'éducation de McGill*, 55(1), 237–257.
<https://doi.org/10.7202/1075728ar>

Article abstract

The research was carried out to determine whether the use of group tests for undergraduate science students to augment lecture material in a second-year core course in microbiology would improve the retention of material on a subsequent regular mid-term/final exam. On three separate occasions, the students were asked to complete short multiple-choice tests individually and then were asked to get together in groups of 4 to re-answer the same questions. The discussions they had in the groups improved their individual marks by 10.9% in the first test, 14.5% in the second test and 20.9% in the third test. Overall, the class average was 2.5% better than the previous year. The majority of the students indicated that the group tests improved their understanding and helped them to learn the lecture material.



THE USE OF GROUP TESTS TO PROMOTE COLLABORATION AND LEARNING: DO THEY WORK?

KIMBERLEY ANN GILBRIDE *Ryerson University*

ABSTRACT. The research was carried out to determine whether the use of group tests for undergraduate science students to augment lecture material in a second-year core course in microbiology would improve the retention of material on a subsequent regular mid-term/final exam. On three separate occasions, the students were asked to complete short multiple-choice tests individually and then were asked to get together in groups of 4 to re-answer the same questions. The discussions they had in the groups improved their individual marks by 10.9% in the first test, 14.5% in the second test and 20.9% in the third test. Overall, the class average was 2.5% better than the previous year. The majority of the students indicated that the group tests improved their understanding and helped them to learn the lecture material.

L'UTILISATION DE TESTS DE GROUPE POUR PROMOUVOIR LA COLLABORATION ET L'APPRENTISSAGE : EST-CE QU'ILS FONCTIONNENT?

RÉSUMÉ. Le projet de recherche présenté ici a été conduit afin de déterminer si l'utilisation de tests de groupe pour compléter le matériel présenté dans un cours obligatoire de microbiologie de deuxième année de baccalauréat en science augmenterait la rétention de la matière dans une évaluation de mi-session ou de fin de session. À trois moments différents, les étudiants ont eu à compléter de courts tests à choix multiples et à ensuite se réunir en équipes de 4 afin de répondre aux mêmes questions. Ces discussions en groupes ont mené à une augmentation de 10,9% de leurs résultats individuels au premier test, de 14,5% au deuxième test et de 20,9% au troisième test. Dans l'ensemble, la moyenne du groupe a augmenté de 2,5% par rapport à l'année précédente. La majorité des étudiants ont indiqué que les tests de groupe les ont aidés à améliorer leur compréhension de la matière.

The goal of every educator is to have their students learn and retain information. However, students often find some concepts hard to understand and therefore struggle with the comprehension and reiteration of course content. It has been found that active participation in processing information as opposed to memorizing content without accompanied understanding leads

to better retention of the information (Lujan & DiCarlo, 2006). Engagement through active learning activities has been found to result in higher order thinking on the part of the student and help the student become a more independent learner (Bonwell & Eison, 1991). A variety of activities can promote active learning including making presentations, participating in out-of-classroom projects, think-pair share activities in the classroom, peer instruction, and generating concept maps to name a few.

Collaborative group testing is an activity that combines think-pair share and peer instruction and has been used as an active learning tool for a number of years. One definition of collaborative learning is that it is an “educational approach to teaching and learning that involves groups of learners working together to solve a problem, complete a task, or create a product” (Laal & Ghodsi, 2012). The use of collaborative group testing has been demonstrated as a way to enhance student learning and increase student retention of the course material (Cortright et al., 2003; Vázquez-García, 2018). Laal and Ghodsi categorize the benefits of collaborative learning into 4 major themes, social, psychological, academic and assessment, to explain why it is a successful learning tool (Laal & Ghodsi, 2012). In science education, Lord explained how cooperative learning can improve skills in at least 11 categories including those related to social skills, science thinking skills and reading and writing proficiency (Lord, 2001). Breedlove and colleagues found that there was a significant association between collaborative testing and concept questions even if they are not combined with collaborative learning (Breedlove, Burkett, & Winfield, 2004). It is thought that this association is due to reduced levels of anxiety and stress on the part of the student by being able to work together to build on each other’s knowledge with positive performance outcomes (Karatzoglou & Weimer, 2011). Furthermore, although it has been shown that collaborative testing is beneficial to all students, it has been found to be significantly more beneficial for low-performing students (Giuliodori, Lujan, & DiCarlo, 2008). However, Breedlove et al. (2004) cautioned that collaborative testing without prior collaborative learning did not improve student’s ability to perform better on theory questions that require higher level cognitive processing. Furthermore, some studies conducted with science (introductory biology class, (Leight, et al., 2012)) and nursing (advanced nursing II course, (Sandahl, 2009) students have noted that collaborative testing did not increase students’ final scores although the students responded very positively to the active learning task. Nonetheless, team learning is becoming an accepted way to enhance learning and may also improve students’ communication and critical thinking skills that can be used in careers after graduation (Loes & Pascarella, 2017). Comparing students that were allowed to re-do a test with the aid of their notes or a textbook versus those that were allowed to collaborate with classmates revealed that collaboration significantly increased the tests scores over individual attempts

with open books (Bloom, 2009). An added benefit was the student's engagement with the material when working collaboratively with their peers (Bloom, 2009). Furthermore, discussion with peers allows students the opportunity to either explain or question the concepts with both activities benefitting the students' comprehension of complex concepts (Giuliodori et al., 2008; Vázquez-García, 2018). Overall, this approach provides students with immediate feedback that is an important component for understanding (Cortright et al., 2003).

In science, many concepts need to be applied in order to understand how they work. Dialogues between teacher and student in a small group setting can be a very effective way to engage students and increase their understanding of course material. However, first and second year science courses in larger universities often contain courses with large class sizes where individual interactions with the instructor may be limited. Allowing students the opportunity to work together to explain and apply their knowledge through application-based questions can increase their understanding of the material or alert them to possible misinterpretations of the facts. Collaborative learning in small groups offers the opportunity for students to make individual contributions and leads to discussions not unlike the principals of scientific inquiry. For these reasons, collaborative group testing has been used with science students to improve their achievement levels and increase their understanding of complex systems (Brame & Biel, 2015; Freeman et al., 2014; Smith et al., 2005). However, while many studies cite a positive response by students to the active learning activities, very few have measured student performance on traditional tests with and without the previous active learning activities. Furthermore, many studies have conducted their collaborative testing on first year science/biology students or nursing students and not many studies have concentrated on more advanced university students (Srougi et al., 2013).

Therefore, the question was, "does the implementation of group tests in a more advanced science such as microbiology benefit the students by improving their retention of the knowledge?" The study was designed to provide students with the opportunity to participate in group testing prior to traditional testing to determine if this added learning tool would result in an increase in performance on the traditional tests. It was thought that the use of group tests as an active learning tool for a large second year microbiology class would augment lecture material and give the students the chance to discuss the material that commonly is misunderstood before being tested with a regular mid-term/final exam. The goals of this study were to use group tests: (1) to provide an alternative testing method that might be less stressful and boost the students confidence in their knowledge through collaboration as measured by the group marks, (2) to increase the students' retention of the course material

as measured by their overall grades on traditional tests, and (3) to encourage networking amongst the students as a career building skill. While this study did not use any definitive metrics to measure this latter skill, information gathered from the post-test questionnaire suggested that the group atmosphere positively affected students and that this interaction was extended beyond the course. The use of the group tests was not a substitute for the regular traditional examination evaluations but, instead, was used to supplement the traditional testing style. An added benefit of collaborative testing was the opportunity for the instructor to identify concepts that are not well answered by the students in order to re-visit these topics in class to improve student understanding.

The outcome of the collaborative testing was measured by using a student's paired t-test to compare the scores when the students answered the test individually and when they answered as a group. The traditional mid-term and final test scores were compared to the previous year's cohort to see if there was any significant change in the final average of the class. Finally, the students were surveyed at the end of the term and asked about their experience with group testing modules for their overall learning.

METHODS

Course structure

Microbiology is a second-year core course for both biology and biomedical science majors at Ryerson University. It runs in the fall semester (3rd semester) and includes a laboratory component. The course was taught by the same individual for several years and typically contained a single section of 250–275 students in the lecture portion. The students had been exposed to group-learning activities such as personal response (clicker) questions in other classes and participated in laboratories in small groups (often pairs); however, none had ever participated in group testing.

Traditional testing for this course consisted of one mid-term and one final for the lecture material (60%) and individual quizzes and lab reports for the laboratory component (40%). The traditional mid-term and final exam format consisted of 50% multiple choice-based questions and 50% short answer questions.

Design of in-class assessments.

To evaluate the effect of collaborative testing on content retention, 265 students in Microbiology (BLG 151) were given group-based tests in addition to their traditional testing evaluations. To evaluate whether group testing had any effect on the scores on the traditional testing components, the group was compared with the same class the year previous. The demographics in both

years were consistent for age, gender, academic level and grade point average. In year 1, the students followed traditional testing while in year 2, on three separate occasions, the students (265) also participated in three group tests. The collaborative group testing consisted of very short multiple-choice tests (10 questions). The tests were answered individually and then the students were asked to form groups of 4 to complete the same test together as a group. Only one scantron per group was provided so that the group had to come to a consensus on the answer. Individual test sheets with the questions were provided for the individual part but for the group assessment, the questions were projected one at a time as PowerPoint slides. This kept all the groups at the same point in the test and encouraged some discussion even between groups since they were all trying to solve the same question at the same time. Figure 1 shows a flow chart of the research design.

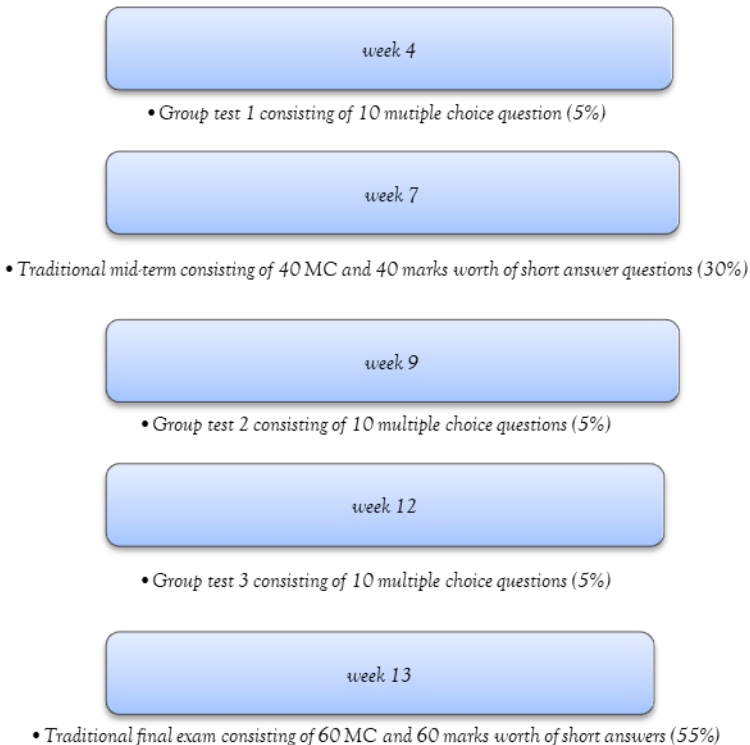


FIGURE 1. *Research design flow chart*

FIGURE NOTES. Students took the traditional tests individually. The group tests were completely individually first and then completed a second time as a group. The percentages represent the amount that assessment contributed to the final lecture grade. The students the previous year only had the traditional tests in week 7 and 13 worth 35% and 65% respectively.

The group tests were designed as short concept-based multiple-choice tests that would generate discussion. Samples of some of the questions are supplied in the Appendix.

Traditional testing was completed as paper copies as previously done and consisted of both multiple-choice questions and short answer questions as previously described.

Analysis of in-class assessments.

Both individual and group tests were marked electronically. Individual marks were used as the final grade for the test if that mark was higher than the group mark. However, if the group mark was higher than the individual mark, then the final mark was obtained by weighting each test as 50% of the mark. This promoted the benefit of working as a group but didn't penalize individuals for group decisions. Traditional tests were marked individually, and the results were compared to the previous year when group-testing had not been carried out.

Design of post-test questionnaire

At the end of the term, students were asked to complete an on-line questionnaire to evaluate their perceptions of the group-testing format. Four of the questions addressed mostly formatting issues, another 5 questions addressed how they felt about the group tests (they were asked to indicate their level of agreement with one of the following answers; yes, somewhat, a little bit, or no), while the last 2 questions allowed students to write about their likes and dislikes about group tests. The survey was conducted after the end of the term and no incentive was provided to students who responded. Sixty-eight students provided feedback, which represented 26% of the class.

Statistical analysis

A student t-test was performed on the mean scores to determine if the group scores were significantly different than the individual scores and whether group testing improved overall course scores where $p < 0.05$ was considered significant.

RESULTS

Students' individual performance versus group performance on group tests

All students were awarded a mark for both their individual tests and their group tests. Overall, the majority of the students (>99%) scored higher on their group tests than on their individual tests. The individual scores and group scores of the entire group (see Figures 2, 3 and 4) were compared and the average mark on the group tests was 10.9%, 14.5% and 20.9% higher than the individual score average for tests 1, 2 and 3 respectively.

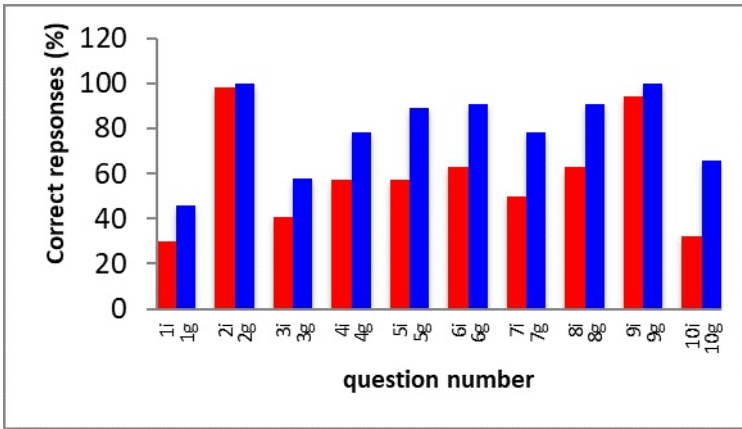


FIGURE 2. *Percent of correct responses for group test 1*

FIGURE NOTES. The group marks were 10.9% better than the individual marks. There were 10 questions. i = individual mark. g = group mark.

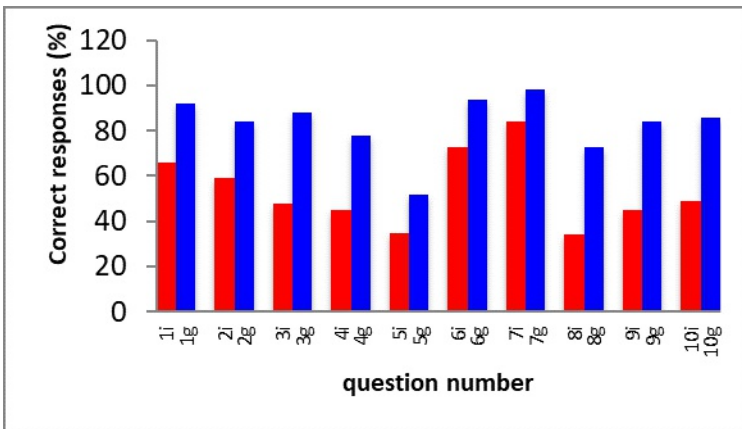


FIGURE 3. *Percent of correct responses for group test 2*

FIGURE NOTES. The group marks were 14.5% better than the individual marks. There were 10 questions. i = individual mark. g = group mark.

To examine what academic level of student would benefit most from group testing, the students' scores were divided into 3 categories based on their individual score out of 10 (see Table 1). Group A were those students that scored 5 or lower on the individual test, Group B were those students that scored 6 or 7 on the individual test, and Group C were those students that scored 8 to 10 on the individual test. The difference between the individual mark and the group mark was the highest in the lowest achieving academic

group, with Group A scoring 32–38% better in the group testing situation. Group B showed a 6.7–25.1% increase and Group C had a 5.5–5.3 % increase in the group test mark over the individual mark.

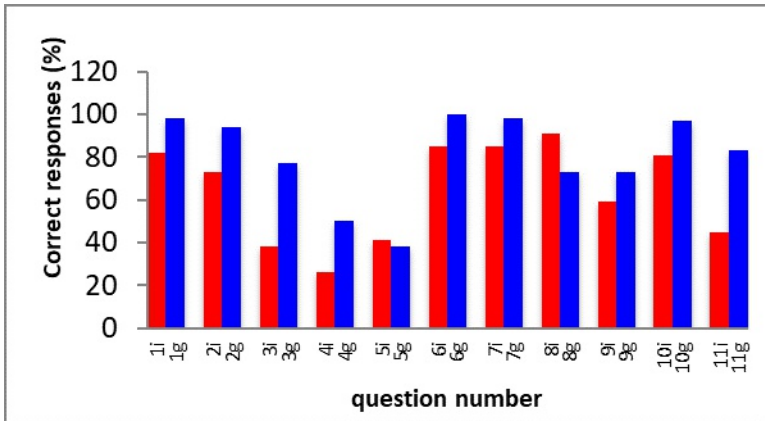


FIGURE 4. *Percent of correct responses for group test 3*

FIGURE NOTES. The group marks were 20.9% better than the individual marks. There were 11 questions. i = individual mark. g = group mark. This was the only test where a question (#5) was answered more poorly as a group than individually.

The effect of group testing on subsequent traditional testing

The marks on the traditional tests after participating in the group tests were compared to previous year mid-term marks where students had not participated in group testing (see Table 2). The questions on the individual and group part of group testing were identical (see the Appendix for examples of questions). However, the traditional tests contained a lot more questions than the group tests (40 multiple choice instead of 10) with the same material being tested including but not only the same 10 multiple choice questions used on the group tests. The class averages were 65.6% and 65.0% for the groups with and without the group tests, respectively. These averages were not significantly different from one another. The final exam mark averages were 64.9% and 63.5% with the final overall class average of 70% and 67.5% respectively for the class that participated in group testing versus the class that did not participate in group testing. The scores on the traditional tests do not fluctuate much from year to year so if extra group testing was effective we expected to see a difference in the scores on the traditional tests with this group of students, which we did not, although the students felt that they were better prepared (see below).

TABLE 1. The average score in the individual and group portions of the tests based on academic performance

Group ¹	Number of students	Test 1 average individual score (%)	Test 1 average group score (%)	Number of students	Test 2 average individual score (%)	Test 2 average group score (%)	Number of students	Test 3 average individual score (%)	Test 3 average group score (%)
A*	101 (39.1%)	41	76	131 (51.2%)	38	76	79 (31.1%)	42	74
B*	99 (38.4%)	65	81	84 (32.8%)	65	72	95 (37.4%)	65	80
C*	58 (22.5%)	82	87	41 (16%)	85	97	80 (31.5%)	83	98
Total	258	59	80	256	54	78	254	60	84

TABLE NOTES. 1. Group A had individual scores 5 or less, Group B had individual scores between 5.5 and 7.5 inclusive, Group C had individual scores of 8 and above.

TABLE 2. *Mid-term and final exam scores¹*

Group	Mid-term average Year 1	Mid-term average Year 2	Final average Year 1	Final average Year 2
A	NA ²	59.2	NA	NA
B	NA	66.7	NA	NA
C	NA	74.9	NA	NA
Total	65.0	65.6	63.5	64.9

TABLE NOTES. 1. Group A had individual scores 5 or less, Group B had individual scores between 5.5 and 7.5 inclusive, Group C had individual scores of 8 and above. 2. Not applicable.

Students' response to group testing

At the end of the course an on-line survey to evaluate students' perception of and participation in group testing activity was conducted. Participation was voluntary and no mark incentive was given. Sixty-five students responded. The questionnaire addressed several facets of group testing exercise, including the format, the usefulness of the tests, and the students' overall attitude towards the group-testing format. Of the respondents, 93% participated in all three tests and, overall, 97% liked the group portion of the tests (see Table 3). The multiple-choice format was liked by 88% with 94% saying that the number of questions asked was appropriate for the time allotted. Interestingly, only 54% stayed in the same group for all three tests, however, the reason why they moved to another group was not surveyed. Anecdotal comments from students about their groups suggests that students switched groups or recruited a new member into their previous group for very obvious reasons such as "someone was sick so we adopted a new member" or "I looked for a better group".

The response to questions that addressed the usefulness of the tests indicated that 87% found the tests helpful for learning the material and reviewing the concepts, 85% said that the tests helped to test their knowledge and 84% found the tests helpful for keeping up with the lecture material. This last sentiment was also echoed in the comments the students gave about what they liked the most about group testing (see Table 4). The students indicated that they liked both the test content and the group collaboration component as a means to enhance their understanding of course concepts they would not have understood as well on their own. Group collaboration also came up when the students were asked about what they liked the least about testing. In almost all answers to these questions, the students commented on how difficult it was to collaborate within the group if not all group members contributed equally (see Table 5).

TABLE 3. *Student perceptions on collaborative testing*

Statement in questionnaire	Student response
1. How many in-class group tests did you participate in?	3 - 93%, 2 - 6 %, 1 - 1 %
2. Did you find the number of questions asked on each test to be appropriate for the time allowed?	94% said yes, 6% said too few
3. Did you like the multiple-choice format?	88% said yes
4. What percentage of your group remained the same for all three tests?	54% stayed in the same group
5. Did you find the group tests helpful for keeping up with the lecture material?	47% said yes 37% said somewhat
6. Did you find the group tests helpful for learning the material?	48% said yes 39% said somewhat
7. Did you find the group tests helpful for reviewing the concepts taught in lecture?	54% said yes 33% said somewhat
8. Did you find the group tests helpful for testing your knowledge on the lecture material?	58% said yes 27% said somewhat
9. Did you like the group portion of the test?	97% said yes

DISCUSSION

Our study used group-tests in a second year Microbiology science majors course. It was a large class of over 250 students with a traditional teaching format consisting of lecture style classes accompanied by weekly labs for groups of 24. The students had not previously participated in group-tests.

To determine whether group testing improved student performance, we compared both individual test scores with group test scores on the group testing activity and mid-term and final exam averages in the traditional testing format from 2 subsequent years. Our study confirmed previous findings that students performed better in groups than individually in the group tests (Corrigh et al., 2003). The discussions the students had in the groups improved their final group test marks by 10.9% in test 1, 14.5% in test 2 and 20.9% in test 3. A similar increase in overall performance on group tests has been seen by others (Bloom, 2009; Corrigh et al., 2003; Rivaz et al., 2015). Since almost all of the group scores were higher than the top individual score in that group it suggested that the students pooled their knowledge as has been suggested by others (Gilley & Clarkston, 2014; Giuliodori et al., 2008). The myth that group tests would result in everyone attaining 100% on the tests because of the teamwork was not found suggesting that the concept based multiple choice questions presented material that was still difficult enough that even when working together debate could still lead to incorrect answers.

TABLE 4. *Comments from students on what they liked the best about the group test experience.*

Comments about content	
1)	I liked how tricky some of the questions were, made us think.
2)	It helps you keep up with course material, so that you are not cramming before the final. I also feel like it prepares you for the final because similar questions appeared.
3)	It was a taste tester for the exam and it forced us to study what was taught so far.
4)	Keeping up with lecture material throughout the semester.
5)	It helped me keep up to date with studying and making notes.
6)	The fact that made me study notes through the year
7)	The fact that it made sure I didn't fall behind, so when I am studying for the finals I have already made notes on it from before.
8)	The fact that it forces you to study the material periodically so when come time for the final you're not swamped with a lot of material you hadn't looked at in weeks.
9)	The quizzes made you review and I was better prepared when it came to studying for the final.
10)	The fact that it keeps you updated with the lecture, so when it was the final exam, I practically knew everything. However, 3 per semester is fine do not do more like chemistry where it is torture rather than helpful.
11)	It tells the students where they stand in terms of preparation.
12)	It forced me to keep up with the content, rather than cramming prior to an exam.
13)	How they helped with the tricky questions on the exams. Helped us to know what to expect for finals.
14)	Split up to allow a good review of material as the semester progressed.
15)	Learning the right answer
16)	The portion that I liked is the fact that there was a second chance. I learned so much after making the mistakes I made. Although it would average out with my individual mark, I was still able to boost up my mark, even if just by a little. I also liked that because of the group test, I kept up with the course in general. Without these, I know I would be studying last minute so this group test idea was very beneficial.

TABLE 4 (continued).

Comments about collaboration	
1)	Discussion
2)	Consulting with each other
3)	The group discussions
4)	Learning from one another through discussion
5)	Can communicate with other groups
6)	Group discussion, so I get to understand where I went wrong in my thinking.
7)	I like the group discussion. I also like how the tests pushed us to review the lecture material.
8)	The discussion part where we could come together and come up with the best answer. Also, the bonus mark question on test 3
9)	Discussing about the questions from different points of view.
10)	We were able to discuss the answers and it did help to pull up my mark.
11)	Got to discuss answers and debate what was right
12)	Got to discuss everyone's answers and explain
13)	I liked being able to talk to my group members to discuss the questions that were being asked, to see how everyone's input impacted our final decision.
14)	I liked how we all communicated as a group and worked together to figure out the answers to the questions.
15)	I liked the aspect of discussing the questions with others to help in solving the problems. As the test questions may be similar to what would be on the exam, getting feedback from others on how to solve the problem was helpful
16)	I enjoyed that the group test sparked discussions and debates during and after the group tests because it introduced me to my peers' reasoning and perspective on course material.
17)	Having the chance to compare you ideas with three different group members and understanding the concept in more details.
18)	The length and the format were the strong point of the group test
19)	Teamwork they knew what I knew
20)	I liked how students were able to form groups after and answer the questions with the knowledge of the entire group to arrive to an agreeable answer.
21)	I liked collaborating with others. It allows me to figure out where I went wrong on my individual test
22)	I liked how I got to interact with different classmates. I also liked seeing how other people think and come to the answer. Everybody had a unique thinking process.
23)	The group testing aspect
24)	It was a way to find new friends and also get help from peers if you knew you got questions wrong.
25)	I liked the fact that we got to discuss answers with group members
26)	The possibility of overall marks increasing if the group test was better than the individual test
27)	I like how we did the tests individually and then together as a group.
28)	Get to consider other answers to the problem that were overlooked during the individual portion
29)	That it first tested our own knowledge and then we were able to discuss it.
30)	I liked being able to talk with my classmates and ask them about questions I wasn't sure about because it helped me learn from my mistakes.
31)	Being able to work together to solve the questions and knowing how each person arrived at the answer.
32)	I enjoyed being able to contribute all of our ideas to collectively come up with one answer. My group was attentive to what everyone had to say and it made the entire process much easier and effective when determining which was the correct answer and our group marks every time were awesome!

TABLE 5. *Comments from students about what they liked the least about the group test experience.*

Comments about collaborating	
1)	Not everyone in the group prepared for the test.
2)	Group members who are completely unprepared and knew nothing.
3)	Some members did not contribute to the discussion and said, "I don't know" to every question.
4)	The fact that the question were up for debate a lot of the time which meant that my mark suffered compared to an exam format
5)	When no one knew the right answer
6)	Arguing
7)	The indecisiveness between group members at times
8)	Too many people in a group, it should be about 2-3 people
9)	I didn't like that there wasn't enough time to discuss with the members in the group.
10)	Occasionally there were minor disagreements as to what we thought was the right answer, but that's normal in any group setting.
11)	Not too much discussing of why that answer was picked. The correct answer was chosen based on what most people answered in the individual test
12)	Although the group test helped spark academic discussions amongst my group members and I, the group test sometimes made me feel more uncertain about my individual answers. This caused my group members and I to become indecisive with some of our answers. However, this helped us all develop good deductive skills to come up with a final answer.
13)	The fact that some students would get carried by the people who knew what they were doing
14)	Your mark relied on answers other people thought were right (didn't always agree)
15)	Some questions were tricky for our group, and even though we studied the contents, the lecture slides didn't have enough information to help us come to a correct answer.
16)	It is kind of unfair that in some groups all the members are prepared so every member contributes and they get the advantage of the group mark but in some groups most members of group are not prepared then only one person is contributing so it basically becomes like individual test. There should be a fair way of grouping.

This outcome identified concepts that could be re-visited in class lectures. Others have observed the same distribution of marks on group scores (Giuliodori et al., 2008).

Comparing the traditional test performance of this class versus the class in the same course the year before when group testing had not been used showed no difference in the average mid-term or final exam average. Overall, there was a 2.5% increase in the class average at the end of the semester, which could be attributed to the group test marks that were worth 15% of the final grade in the course. Although the differences could be due to other factors, none could be identified. Both groups contained students with the same mean age, percent male to female ratio, and program level. Moreover, the large class sizes (>250) helped to minimize any academic difference that might exist between the two groups. Regardless of the overall increase in the traditional test averages, they were not found to be significantly different ($p > 0.05$). This suggests that lessons learned in the group tests may not contribute significantly to the retention of the material, an issue that has also been identified by others (Giuliodori et al., 2008).

However, although the mathematical averages were not significantly different, the attitude of the students as vetted through the survey tool suggested they felt the group tests did help their understanding. Leight et al. (2012) also reported that his group testing did not raise the retention of material of his students and he proposed it was because the cohort of students he studied had already been exposed to numerous interventions to improve learning. He hypothesized that students in a lecture-based class may have more untapped potential for learning improvement. In our case, the students had not been exposed to previous collaborative learning activities, yet they did not display an increased understanding of the material in their traditional tests, which does not support his hypothesis. Others have also found that there was no significant impact of group testing on future tests examining the retention of material (Shindler, 2004; Zimbardo et al., 2003).

To investigate whether the invention benefitted some students more than others we divided the student achievement scores into 3 categories on each occasion; those that scored poorly, those in the mid-range and those that scored well. Our data indicates the students with lowest marks in their individual tests benefitted the most from group testing, in other words, students that scored 5 or lower on the individual part increased their mark 32–38 % with the group part of the testing. On the other hand, high scoring students (scoring 8 or higher) only increased their grades by 5.3–5.5%. Overall, the marks obtained by the students on traditional mid-terms and finals correlated better with their individual mark than with their group mark suggesting that without the combined group knowledge and discussion the students did not retain more knowledge. The benefit of collaboration for

lower achieving students has also been seen by others (Giuliodori et al., 2008; Wiggs, 2011) but, in our case, the absence of an increase in the overall class average suggests that the additional benefit seen with group testing does not necessarily result in an overall higher grade in the course as has been seen by others (Sandahl, 2010).

At the end of the semester a survey to evaluate students' perception of the group testing exercises was available on-line for students to participate in. The overwhelming majority reported favorable views of group testing (see Table 3). They felt the group test improved their understanding and helped them to learn the lecture material (see Table 4). To measure this improvement, we used traditional testing methods and compared this year's results with last year's. We assumed that improved understanding would be measurable and would show a significant increase in class average. However, this was not the case. Perhaps students' understanding learned through non-traditional testing interventions cannot be represented by traditional grading metrics. As shown by their responses on the survey, it seems that regardless of the grades that the students received, they were happy with the group tests and very open to continuing group testing exercises. The comments obtained from the students suggested that they felt that group testing helped them learn the material regardless of the outcome measured by grades (Meseke et al., 2010). Any activity that allows students to be exposed to material more than once (group test and then traditional test) is likely to increase their retention of the material. How many times does a student need to be tested on the same material before grades rise significantly is a question that is difficult to answer. This study looked at results from three group tests, but since the material covered in each test was different, it cannot be used to determine whether more frequent exposure to material helps to improve retention.

However, others have found that despite the lack of empirical data to show an increase in retention of the material, collaborative testing may still create a positive learning environment worth implementing (Sandahl, 2009) and may still be useful for student engagement and contribute to future team and professional skills. It was noted that while the individual marks of high-achieving groups were similar for all three tests (82-85%), the group mark increased over the semester (87→97→98%). This may suggest that collaboration among the students improved the more times that they met even if the retention of the material by an individual student did not. On the traditional tests they did not have that opportunity to collaborate and therefore performed, as they would have, as individuals.

As a sideline, the benefit of group testing exercises can be extended to include information that an instructor could take away from the intervention. For example, questions that are poorly answered both in the individual and group portions of the test gave the instructor the ability to identify concepts that

were difficult for the students. These concepts can then be re-visited in a class lecture to clarify the students' perceptions. The re-exposure to the self-identified difficult material could greatly improve the understanding and, hopefully, the retention of that content for the student. Furthermore, group testing may also be very beneficial in a large class where it is difficult for the instructor to interact with each student individually and therefore scheduled student discussions can augment students understanding.

Students were also asked to comment on what they liked least about group testing exercises and the overwhelming most common comments were directed at the group dynamics and not the outcome of the group test. It appeared that students were most frustrated when other students did not contribute or participate in the group discussions. The students were allowed to change groups from test to test and it appears that only 54% of the students that answered the survey had exactly the same group of students for all three tests. For most cases, only one member of the group was replaced (data not shown). The reason for some students changing groups is not clear from our survey but in a few cases, students changed groups to increase their chances of getting into a more capable group whether that be for participation purposes or for marks (see Table 5, comment 16). These students however appeared to represent a small percentage of the class (<6%).

Overall, collaborative testing in our study supports previous evidence that learning as a group may not increase retention of the material but that students liked the format and generally thought the intervention helped them to understand the material better. In general, group testing was viewed as less stressful yet some students indicated they experienced some stress when others in the group had not prepared themselves to participate in discussion about the material. A side benefit of the exercise was the identification of concepts that students found hard, which allowed them to be re-taught in a later class. Lastly, the students did appear to enjoy the inherent networking feature of group tests. Although retention of material for the course in question was not evident, the intervention made the students feel that they were understanding and learning the material better. Our takeaway hypothesis is that a positive attitudinal response from the students about group testing will cause these students to be better able to apply their knowledge in future courses. A follow-up study proposes to follow these students to monitor their achievement in the subsequent years of their program.

REFERENCES

- Bloom, D. (2009). Collaborative test taking: Benefits for learning and retention. *College Teaching*, 57(4), 216-220. <https://doi.org/10.1080/87567550903218646>
- Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom*. 1991

ASHE-ERIC *Higher Education Reports*. ERIC Clearinghouse on Higher Education. <https://doi.org/ED340272>

Brame, C. J., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *Cell Biology Education*, 14(2). <https://doi.org/10.1187/cbe.14-11-0208>

Breedlove, W., Burkett, T., & Winfield, I. (2004). Collaborative testing and test anxiety. *Journal of Scholarship of Teaching and Learning*, 4(2), 33–42.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content Is improved by collaborative-group testing. *AJP: Advances in Physiology Education*, 27(3), 102–108. <https://doi.org/10.1152/advan.00041.2002>

Freeman, S., Eddy, S. L., McDonough, M., 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*, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>

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. <https://doi.org/10.1017/CBO9781107415324.004>

Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high- and low-performing students. *Advances in Physiology Education*, 32(4), 274–278. <https://doi.org/10.1152/advan.00101.2007>

Karatzoglou, A., & Weimer, M. (2011). Collaborative preference learning. In Fürnkranz J., Hüllermeier E. (eds) *Preference Learning* Springer, Berlin, Heidelberg. (pp. 409–427). https://doi.org/10.1007/978-3-642-14125-6_19

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. In *Procedia - Social and Behavioral Sciences* (Vol. 31, pp. 486–490). <https://doi.org/10.1016/j.sbspro.2011.12.091>

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 Sciences Education*, 11(4), 392–401. <https://doi.org/10.1187/cbe.12-04-0048>

Lord, T. R. (2001). 101 reasons for using cooperative learning in biology teaching. *The American Biology Teacher*, 63(1), 30–38. [https://doi.org/10.1662/0002-7685\(2001\)063\[0030:RFUCL\]2.0.CO;2](https://doi.org/10.1662/0002-7685(2001)063[0030:RFUCL]2.0.CO;2)

Lujan, H. L., & DiCarlo, S. E. (2006). Too much teaching, not enough learning: What is the solution? *Advances in Physiology Education*, 30(1), 17–22. <https://doi.org/10.1152/advan.00061.2005>

Meseke, C. a, Nafziger, R., & Meseke, J. K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *The Journal of Chiropractic Education*, 24(1), 19–29.

Rivaz, M., Momennasab, M., & Shokrollahi, P. (2015). Effect of collaborative testing on learning and retention of course content in nursing students. *Journal of Advances in Medical Education & Professionalism*, 3(4), 178–182.

Sandahl, S. S. (2009). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives*, 31(3), 142–147. <https://doi.org/10.1043/1536-5026-31.3.142>

Shindler, J. V. (2004). Greater than the sum of the parts? Examining the soundness of collaborative exams in teacher education courses. *Innovative Higher Education*, 28(4), 273–283. <https://doi.org/10.1023/B:IHIE.0000018910.08228.39>

Smith, A. C., Stewart, R., Shields, P., Hayes-Klosteridis, J., Robinson, P., & Yuan, R. (2005).

Introductory biology courses: A framework to support active learning in large enrollment introductory science courses. *Cell Biology Education*, 4(2), 143–156. <https://doi.org/10.1187/cbe.04-08-0048>

Srougi, M. C., Miller, H. B., Witherow, D. S., & Carson, S. (2013). Assessment of a novel group-centered testing schema in an upper-level undergraduate molecular biotechnology course. *Biochemistry and Molecular Biology Education*. <https://doi.org/10.1002/bmb.20701>

Vázquez-García, M. (2018). Collaborative-group testing improves learning and knowledge retention of human physiology topics in second-year medical students. *Advances in Physiology Education*. 42(2), 232–239. <https://doi.org/10.1152/advan.00113.2017>

Wiggs, C. M. (2011). Collaborative testing: Assessing teamwork and critical thinking behaviors in baccalaureate nursing students. *Nurse Education Today*, 31(3), 279–282. <https://doi.org/10.1016/j.nedt.2010.10.027>

Zimbardo, P. G., Butler, L. D., & Wolfe, V. A. (2003). Cooperative college examinations: More gain, less pain when students share information and grades. *The Journal of Experimental Education*, 71(2), 101–125. <https://doi.org/10.1080/00220970309602059>

APPENDIX

TABLE S1. *A sample of the types of multiple-choice questions used for the group*

Test #	Sample question
1	<p>Which of the following statements are true?</p> <p>The electron microscope is able to increase the magnification of the object but not the resolution</p> <p>The phase contrast microscope is one type of light microscope</p> <p>As the numerical aperture decreases so does the resolution</p> <p>The confocal microscope uses a laser beam to illuminate specimens and therefore this increases the magnification</p> <p>The scanning probe microscope is able to view single molecules</p> <p>1, 2, and 3</p> <p>2, 3, and 5</p> <p>2, 3, 4 and 5</p> <p>1, 3, 4 and 5</p> <p>all of them are true</p>
1	<p>Photolithoautotrophs use _____ as a carbon source, _____ as an electron source and _____ as an energy source.</p> <p>Organic carbon, organic e⁻ donor, light</p> <p>Inorganic carbon, organic e⁻ donor, light</p> <p>Organic carbon, inorganic e⁻ donor, inorganic chemicals</p> <p>Inorganic carbon, Inorganic e⁻ donor, light</p> <p>Inorganic carbon, inorganic e⁻ donor, organic e⁻ donor</p>

TABLE S1 (continued).

Test #	Sample question
2	<p>A metabolic pathway has 4 steps and a different enzyme catalyzes each step. Step _____ is likely the regulatory step. The regulatory molecule is likely to be the product produce from the _____ step. If you add an inhibitor, the pathway slows down but when you then add substrate you find the inhibition has lessened, therefore you can predict that the form of regulation is</p> <p>-----</p> <p>one, first, competitive inhibition one, fourth, allosteric inhibition four, first, competitive inhibition four, fourth, allosteric inhibition one, fourth, competitive inhibition</p>
2	<p>Consider the standard reduction potential (E_0) of the following two half reactions.</p> <p>$\text{NAD}^+ + 2\text{H}^+ + 2e^- = \text{NADH} + \text{H}^+ \quad E_0 = -0.32 \text{ volts}$</p> <p>$\frac{1}{2} \text{O}_2 + 2\text{H}^+ + 2e^- = \text{H}_2\text{O} \quad E_0 = +0.82 \text{ volts}$</p> <p>When these two reactions are paired, which molecule acts as the electron donor? The electron acceptor? And what is the overall change in E_0 (ΔE_0)?</p> <p>$\text{NAD}^+, \text{O}_2, 0.5\text{V}$ $\text{NADH}, \text{NAD}^+, 1.14\text{V}$ $\text{NADH}, \text{O}_2, 1.14\text{V}$ $\text{O}_2, \text{NAD}^+, 0.82\text{V}$ $\text{NADH}, \text{O}_2, 0.50\text{V}$</p>
3	<p>Which of the following pathways is a route for CO_2 fixation?</p> <p>Reductive acetyl-CoA pathway Reductive, or reverse, TCA cycle 3-hydroxypropionate cycle Calvin cycle Embden Meyerhoff Photosystem II Gluconeogenesis</p> <p>i, ii, iii ii, iii, iv iv, v, vi, vii, i, ii, iii, iv iv, vi</p>
3	<p>Which of these represents a correct order of proteins involved in bacterial DNA replication?</p> <p>$\text{DNA A} \rightarrow \text{DNA pol III} \rightarrow \text{primase} \rightarrow \text{ligase}$ $\text{Primase} \rightarrow \text{DNA pol III} \rightarrow \text{DNA A} \rightarrow \text{ligase}$ $\text{DNA A} \rightarrow \text{primase} \rightarrow \text{DNA pol III} \rightarrow \text{ligase}$ $\text{Primase} \rightarrow \text{DNA pol III} \rightarrow \text{ligase} \rightarrow \text{DNA A}$ $\text{Primase} \rightarrow \text{DNA A} \rightarrow \text{RNA pol} \rightarrow \text{DNA pol I}$</p>

TABLE NOTES. The questions were concept-based and included the correct and most likely answers that students give for these concepts. This promoted discussion since it was rare to have a group come to a consensus for the answer immediately.

KIMBERLEY ANN GILBRIDE is a professor in the Department of Chemistry and Biology at Ryerson University in Toronto, Canada. She has taught general biology, microbiology and molecular biology courses for over 25 years. She has introduced numerous teaching interventions into her classroom over the years and was awarded the Dean's Teaching Award for the Faculty of Science in 2013. gilbride@ryerson.ca

KIMBERLEY ANN GILBRIDE est professeure dans le Département de chimie et de biologie de l'Université Ryerson à Toronto, au Canada. Elle a enseigné plusieurs cours de biologie, de microbiologie et de biologie moléculaire, et ce, depuis plus de 25 ans. Elle a introduit plusieurs interventions pédagogiques dans ses cours à travers les années, ce qui lui a valu le Prix du Doyen de la Faculté des sciences pour l'enseignement en 2013. gilbride@ryerson.ca