The Canadian Journal for the Scholarship of Teaching and Learning

Volume 15 | Issue 2 Article 3

Summer 08-31-2024

The Impact of Two-Stage Testing and Other Tutorial Co-Learning Activities on Cohort Cohesion and Learning Outcomes in a Large-Enrollment Undergraduate Course

Sajeni Mahalingam
McMaster University, mahals4@mcmaster.ca
Ali Moinuddin
McMaster University, moinudda@mcmaster.ca
Veronica G. Rodriguez Moncalvo
McMaster University, rodrigvg@mcmaster.ca

https://doi.org/10.5206/cjsotlrcacea.2024.2.14931

Recommended Citation
Mahalingam, S., Moinuddin, A., & Rodriguez Moncalvo, V. G. (2024). The impact of two-stage testing and other tutorial co-learning activities on cohort cohesion and learning outcomes in a large-enrollment undergraduate course. *The Canadian Journal for the Scholarship of Teaching and Learning, 15*(2). https://doi.org/10.5206/cjsotlrcacea.2024.2.14931

The Impact of Two-Stage Testing and Other Tutorial Co-Learning Activities on Cohort Cohesion and Learning Outcomes in a Large-Enrollment Undergraduate Course

Abstract
Large enrollments in undergraduate courses pose several teaching and learning challenges that impact students' learning experience and performance. Implementing co-learning activities in tutorials of large courses can help mitigate these challenges and improve the learning environment. One type of collaborative learning activity that has become increasingly popular is two-stage testing but there are limitations to how two-stage testing has been conducted. We undertook a study to elucidate whether our modified two-stage testing protocol and other co-learning activities performed in tutorials can enhance the learning experiences of undergraduate students and foster a sense of community in a large-enrollment research methods course. The specific aims of our study were to: 1) Assess whether co-learning activities including two stage testing in tutorials improves learning outcomes and fosters cohort cohesion in a large-enrollment junior undergraduate science course. 2) Evaluate the impact of our modified two-stage testing approach on student learning and long-term retention. To assess cohort cohesion students were asked to complete a survey and were invited to participate in focus groups. Results indicated that tutorials did foster cohort cohesion among students in the tutorial. The tutorial activities helped scale down the course size and connect with their peers. We tested our modified two-stage testing protocol by administering a two-stage test (an individual test consisting of short-answer questions followed by a group test that was comprised of a subset of the individual test questions that was completed during tutorials). Approximately three months after the individual test, a retention test was administered. Student grades were significantly higher in group tests compared to the individual tests. Interestingly, students on average scored 6.2% higher on the retention test questions that were from the group test, compared to questions that were only on the individual test. These results support the idea that group tests help improve student retention. Students reported tutorials and two-stage testing to be a positive learning experience.

Le nombre élevé d'inscriptions dans les cours de premier cycle pose plusieurs problèmes d'enseignement et d'apprentissage qui ont un impact sur l'expérience d'apprentissage et les performances des étudiants et des étudiantes. La mise en œuvre d'activités de co-apprentissage dans les travaux dirigés des grands cours peut contribuer à atténuer ces difficultés et à améliorer l'environnement d'apprentissage. Un type d'activité d'apprentissage collaboratif de plus en plus populaire est le test en deux étapes, mais il y a des limites à la manière dont le test en deux étapes a été mené. Nous avons entrepris une étude pour déterminer si notre protocole modifié de test en deux étapes et d'autres activités de co-apprentissage réalisées dans le cadre de travaux dirigés peuvent améliorer les expériences d'apprentissage des étudiants et des étudiantes de premier cycle et favoriser un sentiment de communauté dans le cadre d'un cours de méthodes de recherche à grand effectif. Les objectifs spécifiques de notre étude étaient les suivants : 1) évaluer si les activités de co-apprentissage, y compris les tests en deux étapes dans les travaux dirigés, améliorent les résultats d'apprentissage et favorisent la cohésion de la cohorte dans un cours de sciences de premier cycle à grand effectif; et 2) évaluer l'impact de notre approche modifiée des tests en deux étapes sur l'apprentissage des étudiants et des étudiantes et la rétention à long terme. Pour évaluer la cohésion de la cohorte, les étudiants et les étudiantes ont été invités à répondre à un sondage et à participer à des groupes de discussion. Les résultats indiquent que les travaux dirigés ont favorisé la cohésion de la cohorte parmi les étudiants et les étudiantes participant aux travaux dirigés. Les activités de tutorat ont permis de réduire la taille du cours et d'établir des liens avec les pairs. Nous avons testé notre protocole modifié de test en deux étapes en administrant un test en deux étapes (un test individuel composé de questions à réponse courte suivi d'un test de groupe composé d'un sous-ensemble de questions du test individuel qui a été réalisé pendant les travaux dirigés). Environ trois mois après le test individuel, un test de rétention a été administré. Les notes des étudiants et des étudiantes étaient significativement plus élevées dans les tests de groupe que dans les tests individuels. Il est intéressant de noter que les étudiants et les étudiantes ont obtenu en moyenne des résultats de 6,2 % supérieurs aux questions du test de maintien des acquis qui faisaient partie du test de groupe, par rapport aux questions qui figuraient seulement dans le test individuel. Ces résultats confirment l'idée que les tests de groupe contribuent à améliorer la rétention des étudiants et des étudiantes. Les étudiants et les étudiantes ont déclaré que les travaux dirigés et les tests en deux étapes constituaient une expérience d'apprentissage positive.

Keywords
co-learning, undergraduate, cohesion, tutorials, two-stage testing; co-apprentissage, premier cycle, cohésion, travaux dirigés, tests en deux étapes

Author's Note
Authors contributed equally to this work.

The research was funded by MacPherson Institute at McMaster University as part of Dr. Mahalingam's Leadership in Teaching and Learning Fellowship. The authors have no competing interests to declare that are relevant to the content of this article.

We are grateful to the MacPherson Institute for funding this research, Dr. Alise de Bie for guidance with qualitative data analysis, Jayson Subaskaran for transcribing focus group discussions, Fariha Chowdhury for formatting the flow chart into a publication quality figure, and Dr. Kimberley Dej for mentorship throughout the project.

Today, junior undergraduate courses often have large enrollments, resulting in limited social interactions amongst peers in a class and poor cohort cohesion (Iipinge, 2018; Mulryan-Kyne, 2010). In these large classes, instructors mainly rely on traditional lectures and individual assessments with multiple choice questions at the expense of more meaningful learning experiences that involve collaboration among students, critical thinking, and deeper learning (Hornsby & Osman, 2014). Research indicates that the lecture-based teaching approach is far from ideal as it benefits a handful of students with a particular-learning style, leading to increased absenteeism, reduced motivation, lack of engagement, and increased mental health issues, in particular in large introductory courses (Hornsby & Osman, 2014; Westervelt, 2016).

Implementation of small-size tutorials in courses that have large enrollments represents one way by which instructors can scale down the large class learning experience while providing students with the opportunity to actively participate and collaborate. Depending on the discipline, tutorial groups may vary in size from very few (e.g., one to four, also known as the 'Oxford' tutorial model) to as many as 45 students (referred here to as 'small group tutorials'; Balwant & Doon, 2021). Small group tutorials are more cost-effective than small class sizes and provide students with an opportunity to apply concepts learned in traditional lectures, clarify misunderstandings, and test their ideas in a safe environment due to a sense of 'equality' with respect to level of knowledge among peers (Robillard et al., 2011) and where the traditional hierarchy in teacher-student relationships is not as prevalent (Balwant & Doon, 2021). Many universities have adopted this small group learning strategy, and studies have shown that tutorial attendance had a strong positive impact on final grades and final exam scores (Boulatoff & Cyrus, 2022). However, literature examining the effectiveness of tutorials as an educational intervention is relatively scarce and, to our knowledge, there is a lack of research regarding the potential for tutorials to help build cohort cohesion through collaborative learning activities.

Collaborative learning (co-learning) through frequent small group interactions constitutes a student-centered educational strategy that can lead to cohort cohesion and the formation of learning communities (Norris & Barnett, 1994; Saunders et al., 2010), where participants acquire, use, and share their knowledge (Brown & Duguid, 2001). Several studies have shown that a collaborative social environment improves learning and contributes to student success (Chickering & Gamson, 1987; Duane & Satre, 2014; Johnson & Johnson, 1994; Williams, 2007). Furthermore, it has been reported that learning communities in colleges and universities increase retention, satisfaction, success, intellectual development and decrease attrition rates (Minkler, 2002; Teitel 1997), potentially indirectly via an enhancement of student feelings of relatedness (Beachboard et al., 2011), sense of security to express themselves (Maher, 2005), and exchange of ideas (Saltiel & Russo, 2001). Thus, large enrollment courses would benefit from adopting activities and assessments that involve collaboration, as these would directly and indirectly improve academic student success, learning experiences, and cohort cohesiveness.

One type of collaborative learning activity that has become increasingly popular is two-stage testing. Two-stage testing is an assessment that requires students to write an individual test without receiving feedback immediately followed by a collaborative test. The group test normally consists of the same or a subset of questions from the individual test, often of multiple-choice format. When used as a summative assessment, both the individual (typically 85-90%) and group components (typically 15-10%) of the test count towards the student's overall score. The impact of two-stage testing and its impact on cognitive and sociopsychological aspects have been studied with mixed results. On the one hand, accumulating evidence suggests that two-stage testing can promote team development (Michaelsen et al., 1999; Michaelsen and Sweet, 2008), facilitate learning (Bremert et al., 2020; Caldecott & Emmioglu, 2015; Gilbride, 2021; Giuliodori et al., 2008; Knierim et al., 2015; Levy et al., 2018; Rempel et al., 2021), enhance retention (Eastwood et al., 2020; Gilley & Clarkston, 2014; Knierim et al., 2015), and improve the student learning experience by reducing stress and anxiety that is normally associated with tests (Bremert, et al., 2020; Caldecott & Emmioglu, 2015; Giuliodori et al., 2008; Levy et al., 2018; Rempel, et al., 2021). On the other hand, a few studies have suggested that collaborative testing does not enhance learning or retention (Leight et al., 2012; Sandahl, 2010; Walker et al., 2023).

There are still many gaps in the collaborative testing literature and drawbacks to how two-stage testing was conducted in previous studies. For instance, in many studies, the group phase of collaborative testing occurred among students that had not previously worked together outside of the testing protocol (Gilbride, 2021; Giuliodori et al., 2008). Collaborative testing requires students to trust one another, share their knowledge and effectively communicate with one another to generate their answers. The discrepancies in the potential cognitive and/or social benefits previously observed may be due to differences in the trust and cohesion amongst students during the group test. To our knowledge research focused on cohesion amongst students that participate in two-stage testing is lacking. Furthermore, when the group test and retention test were administered can also impact the findings of a study. In most studies, the group test was administered immediately after the individual test, which can lead to collaborative inhibition (Gilbride, 2020; Giuliodori et al., 2008; Takahashi & Saito, 2004). Similarly, the retention test was usually administered relatively close to the original two-stage test (one to eight weeks) and while students were still enrolled in that course. There is limited research on the long-term effects of two-stage testing, with the exception of a more recent paper by Eastwood et al. (2020) that conducted the retention test eight months later and found that two-stage testing enhanced retention. These discrepancies in retention may also be because most studies used multiple-choice questions (Leight et al., 2012; Gilley & Clarkston, 2014), and there is a long-term cognitive disadvantage of multiple-choice relative to short answer questions (Cooke et al., 2019; Kang et al., 2007). Our two-stage testing protocol builds on previous studies and aims to improve the student experience, learning outcome and long-term retention.

We undertook a study to elucidate whether our modified two-stage testing protocol and other co-learning activities performed in tutorials can enhance the learning experiences of undergraduate students and foster a sense of community in our large-enrollment research methods course. This second-year mandatory course, Research Methods in Life Sciences, is offered as part of the Life Sciences program offered by the School of Interdisciplinary Science at McMaster University and normally has an enrollment of approximately 500 students. The course consists of two weekly components: large lectures and small group tutorials. Tutorials are capped to 25 students and have been implemented to scale down the large class learning experience by providing students with the opportunity to interact with their peers in a collaborative learning environment and to enhance cohort cohesion. In tutorials, students apply concepts learned during lectures, while learning from peers through different group activities, including case studies, group discussions, peer editing, and more recently two-stage testing. The design of the small group tutorials with co-learning activities is based on a structured cooperative approach rooted in social constructivist theory, which posits that knowledge is the product of interaction, interpretation, and understanding (Vygotsky, 1962). Social constructivist theory focuses on learners as active co-constructors of knowledge and promotes assessment as an active process of acknowledging shared understanding (Adams, 2006). Thus, implementation of collaborative learning activities in tutorials has the potential to foster cohort cohesion and can lead to the formation of learning communities (Saunders et al., 2010; Norris & Barnett, 1994).

In order to address our research question, we examined the impact of small group tutorial co-learning activities including a two-stage test on student performance, long-term learning and cohort cohesion in our introductory research methods course using a mixed methods approach, more specifically concurrent triangulation. We conducted a survey and focus groups to measure cohort cohesion amongst students in the tutorials. We also implemented a novel two-stage testing and retention test protocol. Specifically, we conducted the two-stage test that consisted of application based short-answer questions two and half months after the start of the term, with the collaborative test happening a few days after the individual test. We introduced these modifications to give students an opportunity to build cohesion within their groups and to allow for self-reflection after the individual test in order to minimize collaborative inhibition. In addition, we conducted a retention test three months after the group test to better assess long-term learning. Thus, the specific aims of our study were to: 1) assess whether co-learning activities including two stage testing in tutorials improves learning outcomes and fosters cohort cohesion in a large-enrollment junior undergraduate science course and 2) evaluate the impact of our modified two-stage testing approach on student learning and long-term retention.

Method

Ethics and Consent

This study was approved by McMaster University's Research Ethics Board (Project #2364). All data included in this study are from students that voluntarily consented for their responses and/or their test grades (individual and group tests) to be collected and included in the study. Consent for inclusion of grades and/or responses did not impact students' standing in the course in any way.

Student Population and Course Structure

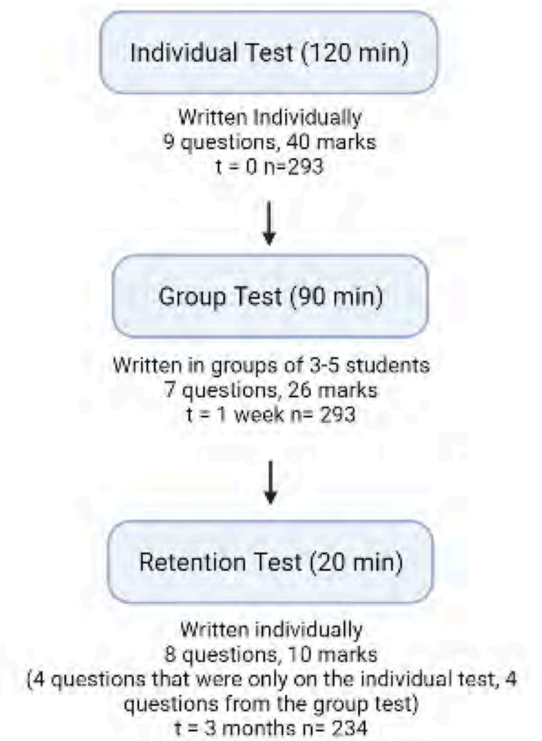
The study was conducted with students enrolled Research Methods in Life Sciences in Fall of 2019. Research Methods in Life Sciences is a Level II Life Sciences course offered by the School of Interdisciplinary Science from the Faculty of Science at McMaster University. The majority of the students complete this course during the second year of their undergraduate degree. The course consisted of two weekly one-hour lectures instructed by Dr. Rodriguez Moncalvo, and weekly two-hour tutorials instructed by undergraduate teaching assistants. In 2019, there were 17 tutorial sections and each one was led by a different teaching assistant. The teaching assistants were provided with a slide deck and a tutorial guide to ensure that all tutorials were led in a similar fashion. Furthermore, there were weekly meetings to train teaching assistants and to model how to run the tutorials. The final grade comprised the following three components: 1) individual lecture-based assessments that were worth 48%, 2) collaborative activities completed in tutorial that were worth 32%, and 3) a two-stage test that was worth 20%, with 90% of the two-stage test grade being from the individual test and 10% from the collaborative test.

Study Design

Testing

Our testing methodology consisted of three different tests: a two-stage test, consisting of an individual test and a group test completed a week apart, followed by an individual retention test completed three months after completion of the individual test (Figure 1).

Figure 1
Testing Methodology Included in the Research Study



Note. The overall testing procedure consisted of three tests: 1) an individual test, 2) a group test, and 3) an individual retention test. T represents the time that passed since the individual test was administered.

Two-stage Testing

All students completed the first test individually on November 9th, 2019. The test was completed on campus, using McMaster University's online learning management platform (Avenue to Learn), and invigilated by teaching assistants. The test consisted of nine application based short-answer questions and students were given 120 minutes to complete the individual test (Figure 1). The questions were graded by the teaching assistants using an answer key provided by the instructor. The individual grade accounted for 90% of each student's two-stage test grade.

During the following week, students were asked to answer a subset of the same questions included in the individual test in groups of three to five students in tutorial (Figure 1). Students completed the group test with peers with whom they had completed other co-learning activities during the course. The group test consisted of seven application based short-answer questions and students were given 90 minutes to complete the group test. The collaborative group test accounted for 10% of the student's two-stage test grade, with all students in a group receiving the same score. Students were provided with an opportunity to view their individual and groups tests after both tests were graded, which included feedback on where they lost marks. 293 students consented to releasing their individual and group test grades.

Individual Retention Test

The individual retention test was administered to students enrolled in a subsequent second-year course, Topics in Life Sciences, during the winter semester of 2020. The retention test was administered during the 6th week in tutorials of Topics in Life Sciences, approximately 3 months following the initial testing (Figure 1). Students were given 20 minutes to complete the test, which consisted of two 'pools' of questions: 1) questions that were included only in the individual test (pool 1) and 2) questions that were included in the group test (pool 2). The retention test had a total of eight questions worth 10 marks in total. Each pool had four questions that were worth five marks in total. Overall, the level of difficulty of both pools was similar. Nevertheless, the questions were not identified to students or clustered on the test as being from either pool or from the individual or group test. Students who completed the retention test did not receive any incentive for completing it. They were given the option to participate in the study and asked their permission to use their answers for this study. 234 students who completed the retention test and took the research methods course consented to releasing their formative responses.

Survey on Two-stage Testing

Following the group test, students completed an online survey administered through the LimeSurvey platform regarding their experience with two-stage testing. This survey was adapted from the study conducted by Giuliodori et al. (2008). The survey consisted of 13 Likert-style questions; students could choose to either strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each. 348 students that completed the collaborative group test and the survey consented to release their survey responses.

Cohort Cohesion

Survey on Cohort Cohesion

Students engaged in tutorial activities in accordance with the program's approach to providing more cohort interactions within the junior large-enrollment courses. Tutorials act to scale down these large enrollment courses into more collaborative interactions, which cannot be feasibly achieved in a lecture hall setting. During the final tutorial of Research Methods in Life Sciences, students were presented with an online exit survey and given approximately 20 minutes to complete it. The survey consisted of Likert-scaled survey questions, adapted from Rovai's (2002) study, to determine the students' feelings towards collaborative learning and community in the program. 290 students completed the online exit survey and consented to release their responses.

Focus Groups on Cohort Cohesion

Near the end of the fall term, students were invited to participate in focus groups and share their opinions about the co-learning activities in tutorial, the tutorial environment, and sense of belonging to their program cohort. The following 10 prompting questions were used in the focus groups, and they were designed to address whether the tutorials helped facilitate stronger cohort bonds among program peers:

1. Please state the program and level that you're enrolled in.
2. Did you find the tutorial to be a positive learning environment? Yes/No? Explain.
3. Did the co-learning activities (group activities) in the tutorials help you with your Understanding/learning of concepts? Yes/No? Explain.
4. What was your favourite tutorial activity and why?
5. Did the tutorial activities make the course more enjoyable? Yes/No? Explain.
6. Have you made friends in tutorial? Yes/No? If yes, do you think you will continue to have a relationship with anyone in your tutorial?
7. Do you feel like you are a part of the Life Sciences cohort? Yes/No? Explain.
8. Has this course provided you with a sense of belonging to your program? If yes, how?
9. How did/didn't the group test activity lead to stronger cohort connections?
10. Do you have recommendations (changes/additions to the tutorials) that would help create a sense of community and increase cohort cohesion?

The focus groups lasted approximately 30-45 minutes per group. Medium-sized groups of approximately four students were preferred; however, due to scheduling and student availability, group size varied from one to five students per session. There were six focus group sessions in total (number of students in each = 5, 3, 3, 2, 2, 1). Research students facilitated the focus groups and took high-level notes of student responses for further qualitative analysis. The focus groups were not recorded in order to make students more comfortable with sharing their experience and perspectives. A total of 16 students participated in the focus groups. Students were provided physical copies of the question prompts if they wished to write a response instead of sharing it aloud with the focus group.

Data Analysis

Test Grades

All data are presented as mean ± standard error of means (SEM). GraphPad Prism 9 was used for all figure creation and statistical analysis, $p < 0.05$ was considered significant. Learning was measured as the change in students' performance based on differences between their individual and group test grades. In order to calculate grade changes, test grades were initially collected with student numbers to pair the individual and group test, but once they were paired the student numbers were deleted before analyzing the data. The difference between individual and group test grades was calculated for each student as percent difference in grades and analyzed statistically using a paired t-test.

In order to determine whether all students had benefitted from the group test and if they had done so equally, the individual grades were separated into grade ranges and these were then compared based on the students' percent difference in grades using a one-way analysis of variance (ANOVA) and a Tukey's multiple comparisons post-hoc analysis.

The retention tests were completely anonymous. The grades from the retention test of each student were divided into questions that were only on the individual test (pool 1) and questions that were on the group test (pool 2). The grades received by each student on these two pools of questions were then compared using a paired t-test.

Two-stage Testing and Cohort Surveys

The two-stage testing survey and cohort survey consisted of Likert-style questions; students could choose to strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree with each. For the analysis, strongly agree and agree were binned together as agree, neither agree nor disagree was counted as neutral, and disagree and strongly disagree were binned together as disagree. The percentage of students in each of these categories was calculated (agree, neutral, and disagree).

Focus Groups

Focus group discussions were transcribed via high-level notes to analyze the participants' opinions and perceptions of the fall undergraduate course included in our study, focusing on cohesion and use of collaborative activities in tutorials. Transcribed conversations were manually coded by the second author. Once the coding was completed, the data were reviewed collectively by the author team. Themes were developed inductively, using recurring codes from the transcripts to posit general feelings and perceptions expressed by the participants. Three overarching themes were identified. A thematic qualitative tree was created using code words/phrases under each of the three thematic concept stems (Figure 4).

Concurrent Triangulation Design

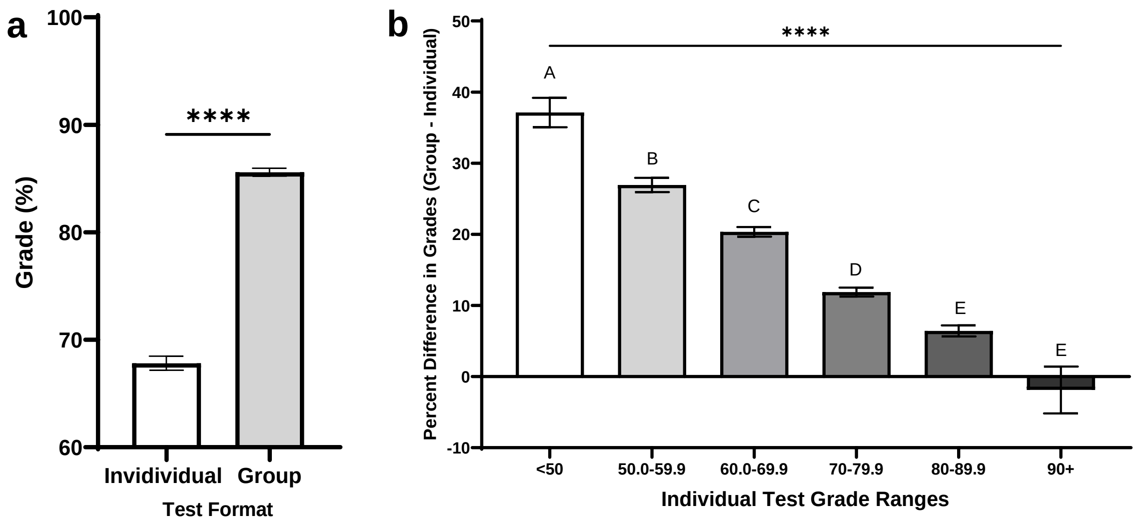
As a more holistic research approach, we used concurrent triangulation design including the data from test grades, surveys and focus groups to further address our research question and specific aims. In brief, after conducting independent analysis of each data set, the outcomes from the different data sources were compared side-by-side to draw our general conclusions that were supported by the different sources.

Results

Two-Stage Testing: Individual and Group Test Grades

Examination of the effect of the two-stage testing approach on students' grades showed that, overall, almost all the students scored higher in the group test compared to their individual test, with the exception of three students that scored over 90% on the individual test. The average individual test grade was 67.81 ± 0.66% and the average group test grade was 85.59 ± 0.38% (Figure 2a), resulting in the group test grade average that was ~17.5% higher than the average of the individual grades ($p < 0.0001$). Further analysis after grouping the individual grades into different ranges indicates that students that scored lower on the individual test had a greater difference between their individual grades and group test grades compared to students that scored higher on the individual test ($p < 0.0001$; Figure 2b).

Figure 2
Effects of Two-Stage Testing on Individual Student Performance



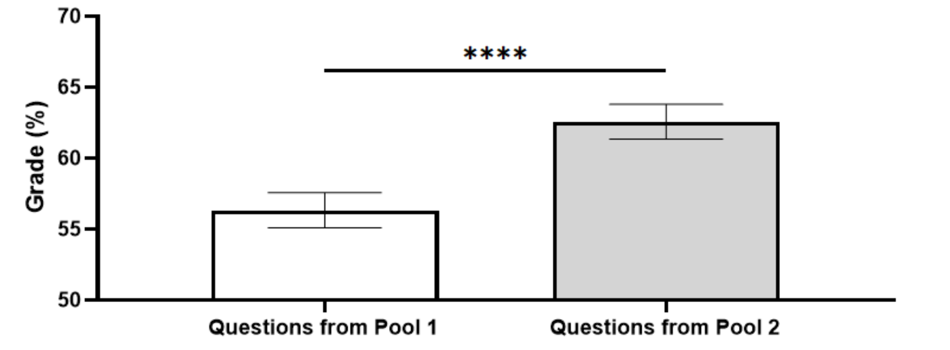
Note. a) Averages of individual and group test grades. Group test grades were higher than individual test grades ($N = 293$, $p<0.0001$). ****Significant difference between groups. b) Difference in test grades (group test grade (%) – individual test grade (%)) by individual test grade ranges (%). The difference in individual and group test grades by grade ranges were higher in students that scored lower on the individual test ($N = 293$, $p<0.0001$). Different uppercase letters represent significant pairwise differences between grade ranges.

Retention Test Grades

In addition to examining the short-term effect of two-stage testing on individual student performance, we evaluated the long-term effectiveness of two-stage testing on student success. For this purpose, we examined students' performance in a retention test completed in the winter term that consisted of questions found only on the individual test (pool 1), and questions of similar difficulty found in the group test (pool 2). Our analysis of the retention test revealed that students performed ~6.3% better on questions that were on the group test (average grade of pool 2 questions = 62.56 ± 1.22%) than on questions that were only on the individual test (average of pool 1 questions = 56.23 ± 1.23%) ($p<0.0001$; Figure 3).

Figure 3

Individual Retention Test Grades Separated into Questions That Were Only on the Individual Test (Pool 1) and Questions That Were on the Group Test (Pool 2)



Note. Students scored higher on questions that were from pool 2 compared to questions that were from pool 1 (N = 234, p <0.0001). ****Significant difference between groups.

Two-Stage Testing Survey

The survey adapted from Giuliodori et al. (2008) consisted of 13 Likert-style survey questions that consisted of statements that support the use of two-stage testing to enhance learning. A summary of the survey questions and analysis is shown in Table 1. Our analyses indicated that most students felt that two-stage testing improved their understanding of concepts and provided a positive learning experience. For example, more than 90% of the students agreed with the following two statements: "Every group member contributed to the learning process" and "This testing methodology provided the opportunity to discuss incorrect answers and fill in knowledge gaps."

Table 1

Findings from the Two-stage Testing Survey (N = 348 students).

Survey Question	Percentage of Students		
	Agree	Neutral	Disagree
1. The purpose of and rationale behind the educational process was fully explained.	81.90%	11.21%	6.90%
2. The process was not too lengthy or complex in its format.	74.14%	14.37%	11.49%
3. The peer discussions improved my level of confidence on the answers.	77.87%	13.22%	8.91%
4. Every group member contributed to the learning process.	92.82%	4.60%	2.59%
5. The level of peer (group) discussions was very high.	93.97%	4.60%	1.44%
6. The group discussions enhanced my understanding of the concepts.	85.06%	10.92%	4.02%
7. My level of involvement during discussions was high.	93.39%	6.32%	0.29%

Survey Question	Percentage of Students		
	Agree	Neutral	Disagree
8. This testing methodology provided a more constructive classroom environment.	83.91%	13.22%	2.87%
9. This testing methodology provided the opportunity to discuss incorrect answers and fill in knowledge gaps.	92.24%	5.46%	2.30%
10. This testing methodology was educationally attractive due to the novelty of this style and format.	67.24%	20.11%	12.64%
11. This testing methodology was less stressful than traditional testing methods.	81.03%	9.77%	9.20%
12. Learning from my peers was a positive learning experience for me.	91.95%	6.61%	1.44%
13. Teaching my peers was a positive learning experience for me.	85.34%	11.49%	3.16%

Note. Survey questions were adapted from "Collaborative Group Testing Benefits High- and Low-Performing Students," by M. J. Giuliodori, H. L. Lujan, & S. E. DiCarlo, 2008, *Advances in Physiology Education*, *32*(4), p. 277 (https://doi.org/10.1152/advan.00101.2007). Copyright 2008 by the American Physiological Society. Reprinted with permission.

Cohort Cohesion Survey

The survey adapted from Rovai (2002) consisted of 25-Likert style questions that consisted of statements related to students' feelings of connectedness to their peers and the overall program. A summary of the survey questions and analysis is shown in Table 2. On average, about two-thirds of the students found that the tutorial activities contributed to cohort cohesion amongst students in their program. For example, about 67% of the students agreed that they felt connected to others in the tutorial and trust others in the tutorial.

Table 2
Findings from the Cohort Cohesion Survey (N = 290 students).

Survey Question	Percentage of Students		
	Agree	Neutral	Disagree
1) I feel that students in this tutorial care about each other.	65.52%	26.90%	7.59%
2) I feel that I am encouraged to ask questions.	83.79%	13.45%	2.76%
3) I feel connected to others in this tutorial.	64.48%	9.31%	26.21%
4) I feel that it is hard to get help when I have a question.	9.31%	13.10%	77.59%
5) I do not feel a spirit of community.	16.90%	30.34%	52.76%
6) I feel that I receive timely feedback.	73.79%	15.86%	10.34%
7) I feel that this tutorial is like a family.	31.03%	34.83%	34.14%

Survey Question	Percentage of Students		
	Agree	Neutral	Disagree
8) I feel uneasy exposing gaps in my understanding.	23.1%	28.28%	48.62%
9) I feel isolated in this course.	12.07%	21.03%	66.90%
10) I feel reluctant to speak openly.	15.86%	24.48%	59.66%
11) I trust others in this tutorial.	66.55%	23.79%	9.66%
12) I feel that the tutorials result in only modest learning.	35.7%	41.03%	23.79%
13) I feel that I can rely on others in tutorial.	63.10%	22.31%	14.48%
14) I feel that other students in tutorial don't help me learn.	12.41%	21.38%	66.21%
15) I feel that members in this tutorial depend on me.	33.10%	43.10%	23.79%
16) I feel that I am given ample opportunities to learn.	71.72%	22.07%	6.21%
17) I feel uncertain about others in this tutorial.	16.90%	37.24%	45.86%
18) I feel that my educational needs are not being met.	14.48%	23.10%	62.41%
19) I feel confident that others will support me.	58.28%	30.69%	11.03%
20) I feel that the tutorials do not promote a desire to learn.	18.97%	24.48%	56.55%
21) The tutorial interactions helped guide my extracurricular interests.	41.38%	35.86%	22.76%
22) Some of the members of this tutorial are individuals I will keep in touch with as fellow Life Sciences students.	66.55%	17.93%	15.52%
23) The tutorials facilitated a sense of belonging to the greater Life Sciences Program.	55.86%	26.90%	17.24%
24) The tutorials helped to scale the large program size down to a more interactive environment among cohort peers.	71.72%	20.34%	7.93%
25) The tutorials helped me feel connected to other Life Sciences students.	70.34%	17.59%	12.07%

Note. Survey questions 1-20 were adapted from "Development of an Instrument to Measure Classroom Community," by A. P. Rovai, 2002, *The Internet and Higher Education*, *5*(3), p. 209 (https://doi.org/10.1016/s1096-7516(02)00102-1). Copyright 2002 by Elsevier Science Incorporated. Reprinted with permission.

Focus Groups

Three main themes were developed from the focus group data: individual success, characteristics of the tutorial that facilitated co-learning, and recommendations for future program development (Figure 4).

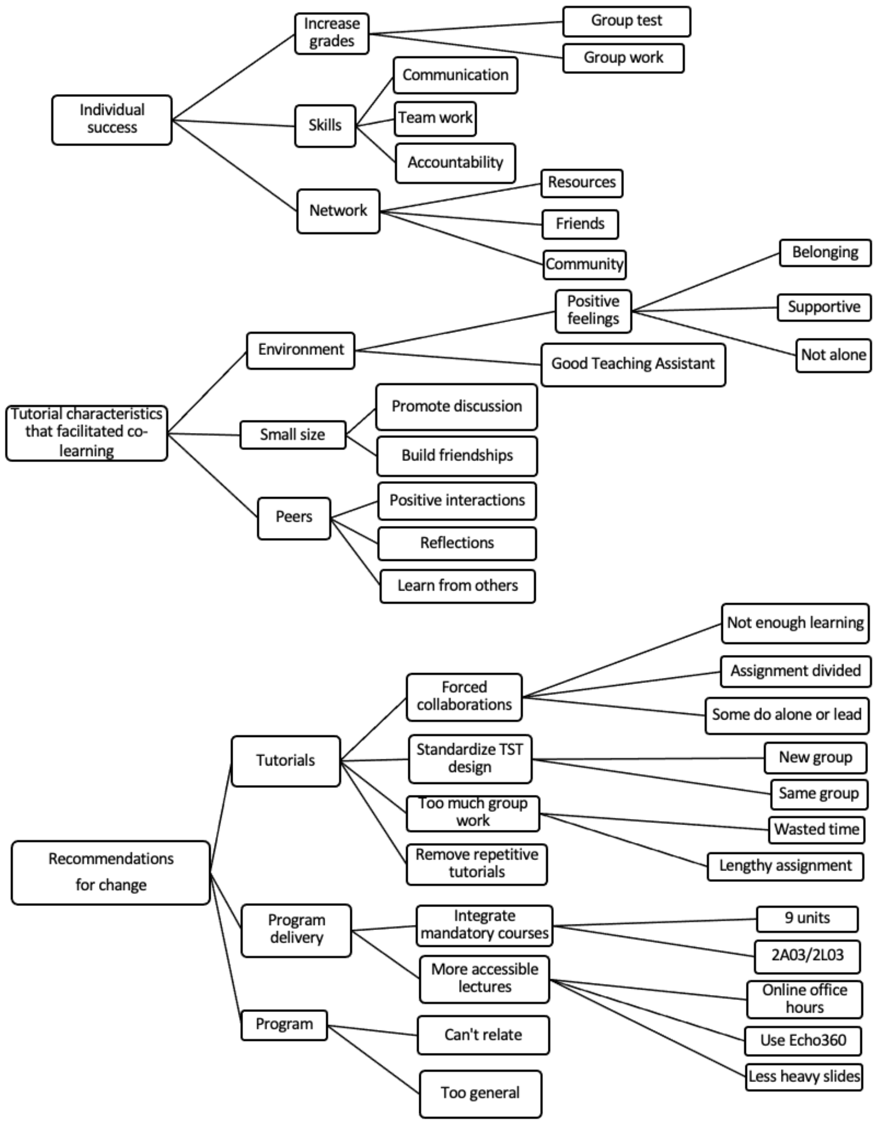
Many students prioritized and spoke about individual success and learning. They were specifically interested in improving their own grade, developing skills, and networking. They enjoyed the co-learning approach used in tutorials, as it helped them increase their grades through two-stage testing and other group work. For example, one student stated that the group work helped them improve their grade by having peers to edit work and to correct their misunderstandings. Students felt that these group activities helped improve their communication and teamwork skills. Students also considered the tutorial as a place to network and make connections. Students stated they would probably keep in touch with students from their tutorial for program related information and that they would be happy to work with them in the future.

Students also talked about characteristics of the tutorial that helped establish cohort cohesion and co-learning within tutorials. Students appreciated the scaled down tutorial size from the large lecture. For example, one student felt it was hard to participate in the large lecture, so they enjoyed the tutorial setting. Students also appreciated the positive learning environment that was created by the teaching assistants. A couple of students specifically stated that they were able to participate in tutorials without being judged and that their teaching assistant was approachable. This helped students form friendships and made students more comfortable to participate in collaborative activities, which promoted discussions that led to students learning from one another. The perception of belonging in the program was a bilateral conversation, as students never felt alone, but they did not feel connected to the general program.

Finally, students provided constructive feedback about the tutorial, course, and program. They recommended changes that they felt would improve the student learning experience. Most of these recommendations were not related to co-learning. However, some students did mention that there were too many forced collaborative activities in tutorial. Furthermore, they felt that some students would lead and divide the assignments instead of working together. Therefore, they felt that some activities could have been individual activities instead. A few students also specifically spoke about the formation of two-stage testing groups; however, there were mixed opinions regarding how the groups should be formed. Some students liked that they worked in the same group that they had been working with for the whole semester, whereas others felt that a new group would have been more beneficial.

The implementation of small-group tutorials in large-enrollment courses has the potential to improve the student learning experience. Positive cohort experiences have been shown to provide benefits to students including improved academic performance (Barnett & Muse, 1993, Dyson & Hanley, 2002). Thus, we wanted to examine whether tutorials help build cohort cohesion by providing a safe interactive environment; where students can be part of a community, be confident to share ideas, and be valued. We found that implementation of tutorials can help build cohort cohesion and enhance co-learning. Furthermore, our modified two-stage testing approach within tutorials led to improved learning and long-term retention. It was also well received by students.

Figure 4
Thematic Analysis of Focus Group Data (N = 16 students)



Discussion

The findings of our study suggest that the tutorials help scale down the size of the course by giving students the opportunity to interact more closely with peers throughout the semester via other group activities beyond two-stage testing. The tutorials indirectly support student learning and improve their experience. Overall, the majority of students that filled in the survey appreciated having tutorials where they are able to work on different group activities with their peers. In particular, 70% or more students that filled in the cohort survey agreed that tutorials provide ample and encouraging opportunities to learn, get help, and receive timely feedback during group discussions (Table 2). Furthermore, the students in the focus groups also felt that the tutorials helped enhance their individual performance and success (Figure 4). They felt that the group test and group work in tutorials helped them do better in course, while also helping them develop important skills such as communication and team-work skills. They also really appreciated the small size, positive learning environment and the opportunity to learn from one another. Together, the focus group and survey data support the social constructivist theory that sharing knowledge amongst peers can enhance learning (Amineh & Asl, 2015; Kim, 2001).

The positive outcomes of co-learning activities such as group testing also help students recognize the value of collaborative learning while providing an overall learning experience that is more motivating than stressful (Lusk & Conklin, 2003; Mitchell & Melton, 2003; Keselyak et al., 2009). The use of collaborative testing in the undergraduate course included in our study was well received and valued by students enrolled in our course (Table 1). In particular, most of the students that completed the survey thought that this testing methodology was less stressful than more traditional ways of testing, increased their confidence on the answers given, and provided a more constructive, positive learning experience (Table 1). In previous studies, students have reported benefits from co-learning activities beyond individual performance, such as enhanced communication skills and improved problem-solving skills (Mitchell & Melton, 2003). In addition, two-stage testing and other group activities are known to have positive effects on student relationships and team development (Michaelsen et al., 1999; Sandahl, 2010) and can help students prepare for team-oriented careers.

Two-stage testing is a powerful assessment tool known to have multiple benefits to students including increased individual learning and performance (Gilley & Clarkson, 2014; Giuliodori et al., 2008; Knierim et al., 2015; Levy, et al., 2018; Mitchell & Melton, 2003; Rempel et al., 2021; Roediger & Karpicke, 2006). Social constructivist theory promotes assessments as an active process and two-stage testing is one educational strategy that provides strong evidence that assessments for learning can enhance the learning experience and student performance (Giuliodori et al., 2008; Kapitanoff, 2009; Koles et al., 2010; Cantwell et al., 2017; Eastwood et al., 2020). The findings of our study further support this idea. We found that the two-staging testing strategy used in the research methods undergraduate course effectively improved student short-term learning as measured by overall test grades. Our findings showed that, on average, group test grades were higher than individual test grades (Figure 1a). Furthermore, most of the students that completed the survey thought that the process provided an opportunity to correct misunderstandings, to fill in their knowledge gaps, and to better understand course material (Table 1). During collaborative testing, students are expected to discuss and agree upon the best answer to each question. In this process, students are not only able to learn alternative approaches and answers, but they also receive instant feedback on their previous individual answers from their peers. Together, these outcomes make this assessment a highly valuable learning experience.

Similar to previous studies (Gilley & Clarkson, 2014; Giuliodori et al., 2008), our findings indicate that students that obtain lower individual scores benefit the most from group testing (Figure 1b). A study by Mahoney & Harris-Reeves (2019) found that only lower to middle achieving students performed better on group tests compared to individual tests. Nevertheless, our findings suggest that reciprocal collaboration between low- and high-achieving students must have occurred as group scores were higher than individual scores within a group. This is in agreement with some of the survey responses given by our students (Table 1). In particular, approximately 93% of students that completed the two-stage testing survey agreed that each member contributed to group discussions to a great extent, and that these discussions were highly productive. Furthermore, the requisite of first completing the individual test to participate in the group test and the fact that both the individual and the group scores are used as part of the test grade likely helped ensure some level of individual accountability towards better preparation and participation in the co-learning process (Michaelsen & Sweet, 2008). Furthermore, Mahoney & Harris-Reeves (2019) found that all students benefited from collaboration on questions that were designed for higher order thinking. The fact that our tests consisted of application based short-answer questions may explain why all students performed better on the group tests.

It is important to note that our study involved group testing that occurred during tutorial time and among students that had formed tutorial groups at the start of the course. Since the group test took place closer to the end of the term, the students knew their group peers well. This likely helped facilitate open and more productive discussions during the group test and avoid group conflict/poor group dynamics during the test, a previously reported disadvantage of group testing (Eastwood et al., 2020; Levy et al., 2018), which can be particularly prominent among junior students due to lack of experience.

Another unique aspect of our study design was that students did not write the group test immediately after the individual test but three to five days afterwards, providing students with the opportunity to check resources and notes prior to the group testing. This may have alleviated potential anxiety and minimized collaborative inhibition by allowing students to reflect on their answers while still holding them responsible for their own learning. Together, these factors likely contributed to the overall positive effects of the two-stage testing approach on student performance.

Another advantage of collaborative testing is an increase in knowledge retention and long-term learning (Cortright et al., 2003; Eastwood et al., 2020; Rivaz et al., 2015). Of note, increased overall test scores as a result of participation in group testing should not be interpreted as increases in knowledge retention. To better assess this, a retention test should be performed sometime after group testing. The inclusion of a retention test three months after the initial testing allowed us to more accurately measure the impact of collaborative testing on individual student long-term learning and retention. Our analyses from the retention test revealed that students remembered content that was on the group test better than content of similar difficulty that was only on the individual test, suggesting that discussing concepts with peers helps students remember concepts longer. Moreover, our tests consisted of short-answer questions, which is known to be better for consolidation of material and retention compared to multiple-choice questions (Cooke et al., 2019). Therefore, we recommend using short-answer questions when using two-stage testing.

It is important to note that one potential limitation of our study design is that the improved performance observed in the retention test could have been caused by repeated exposure to group test questions (i.e., two times as supposed to one time; Gilley & Clarkson, 2014). In fact, previous research has demonstrated the benefit of repeated initial testing (test-enhanced learning; Karpicke & Roediger, 2008; Roediger & Karpicke, 2006; Wheeler & Roediger, 1992). However, other studies have shown that the process of elaborative retrieval that occurs during group testing and not just exposure to the same material contributes to better retention (Cranney et al., 2009; Eastwood, et al., 2020), consistent with the 'retrieval hypothesis'. Future research on two-stage testing should control for repeat exposure in order to be able to make more definitive conclusions about the retention benefits of two-stage testing.

Another limitation to our overall design is the lack of feedback on the other collaborative learning activities used in tutorials. Almost 72% of the students felt that they were given ample opportunities to learn (Table 2), but we are unable to pinpoint which collaborative activities worked better in tutorials. In fact, some students in the focus groups felt that there were too many collaborative activities in tutorials (Figure 4). Unfortunately, the sample size for focus groups was also very small and that made it difficult to be confident in the focus group data. Furthermore, the focus groups were not recorded in order to allow students to feel more comfortable, but that may have led to a loss of some information.

Overall, tutorials and the collaborative activities were well received by students; however, improvements could be made to increase cohesion within tutorials and amongst students within the program. For instance, only about half of the students felt a spirit of community within the tutorials (Table 2). Future research should focus on improving cohort cohesion in tutorials by adding team building exercises and evaluating the impact on student learning.

Conclusion

Increases in program enrollment coupled with resource limitations have pushed higher education institutions to massification via lecture-based learning (Hornsby and Osman, 2014). Our findings suggest that implementation of collaborative activities in tutorials in large-enrollment courses can help overcome superficial learning that may result from large lectures while helping to alleviate some of the financial constraints associated with having a small student to teacher ratio. In particular, tutorials can help improve the student experience by fostering cohort cohesion and enhancing learning. Furthermore, implementing our modified two-stage testing protocol in tutorials with pre-formed groups can enhance the benefits of two-stage testing by minimizing collaborative inhibition, reducing test anxiety, and promoting long-term learning.

References

Adams, P. (2006). Exploring social constructivism: Theories and practicalities. *International Journal of Primary, Elementary and Early Years Education*, *34*(3), 243–257. https://doi.org/10.1080/03004270600898893

Amineh, R., & Asl, H. (2015). Review of constructivism and social constructivism. *Journal of Social Sciences, Literature, and Languages*, *1*, 9–16.

Balwant, P. T., & Doon, R. (2021). Alternatives to the conventional 'oxford' tutorial model: A scoping review. *International Journal of Educational Technology in Higher Education*, *18*, 29. https://doi.org/10.1186/s41239-021-00265-y

Barnett, B. G., & Muse, I. D. (1993). Cohort groups in educational administration: Promises and challenges. *Journal of School Leadership*, *3*(4), 400–415. https://doi.org/10.1177/105268469300300405

Beachboard, M. R., Beachboard, J. C., Li, W., & Adkison, S. R. (2011). Cohorts and relatedness: Self-determination theory as an explanation of how learning communities affect educational outcomes. *Research in Higher Education*, *52*(8), 853–874. https://doi.org/10.1007/s11162-011-9221-8

Boulatoff, C., & Cyrus, T. L. (2022). Improving student outcomes in large introductory courses. *International Review of Economics Education*, *41*, 100247. https://doi.org/10.1016/j.iree.2022.100247

Bremert, H., Stoff, A., & Boesdorfer, S. B. (2020). Collaborative assessments: Learning science and collaborative skills during summative testing. *The Science Teacher*, *87*(9), 32–37. https://doi.org/10.1080/00368555.2020.12293541

Brown, J. S., & Duguid, P. (2001). Knowledge and organization: A social-practice perspective. *Organization Science*, *12*(2), 198–213. https://doi.org/10.1287/orsc.12.2.198.10116

Caldecott, M., & Emmioglu, E. (2015). Do group exams support English as an additional language student learning? *Multicultural Learning and Teaching*, *12*(1), 27–48. https://doi.org/10.1515/mlt-2014-0013

Cantwell, E. R., Sousou, J., Jadotte, Y. T., Pierce, J., & Akioyamen, L. E. (2017). PROTOCOL: Collaborative testing for improving student learning outcomes and test-taking performance in higher education: A systematic review. *Campbell Systematic Reviews*, *13*(1), 1–18. https://doi.org/10.1002/cl2.186

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, *39*(7), 3–7.

Cooke, J. E., Weir, L., & Clarkston, B. (2019). Retention following two-stage collaborative exams depends on timing and student performance. *CBE—Life Sciences Education*, *18*(2), 1-8. https://doi.org/10.1187/cbe.17-07-0137

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

Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, *21*(6), 919–940. https://doi.org/10.1080/09541440802413505

Duane, B. T., & Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. *Nurse Education Today*, *34*(1), 31–34. https://doi.org/10.1016/j.nedt.2013.03.005

Dyson, L., & Hanley, B. (2002). Testing the effect of a cohort grouping model as a form of instructional grouping in teacher education. *Canadian Journal of Higher Education*, *32*(2), 27–46. https://doi.org/10.47678/cjhe.v32i2.183410

Eastwood, J. L., Kleinberg, K. A., & Rodenbaugh, D. W. (2020). Collaborative testing in medical education: Student perceptions and long-term knowledge retention. *Medical Science Educator*, *30*(2), 737–747. https://doi.org/10.1007/s40670-020-00944-x

Gilbride, K. A. (2021). The use of group tests to promote collaboration and learning: Do they work? *McGill Journal of Education*, *55*(1), 237–257. https://doi.org/10.7202/1075728ar

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/10.2505/4/jcst14_043_03_83

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

Hornsby, D. J., & Osman, R. (2014). Massification in higher education: Large classes and student learning. *Higher Education*, *67*(6), 711–719. https://doi.org/10.1007/s10734-014-9733-1

Iipinge, S. M. (2018). Challenges of large class teaching at the university: Implications for continuous staff development activities. *The Namibia CPD Journal for Educators*, 105–112. https://doi.org/10.32642/ncpdje.vi.1263

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning*. Allyn and Bacon.

Kang, S. H., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education*, *10*(1), 56–70. https://doi.org/10.1177/1469787408100195

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Keselyak, N. T., Saylor, C. D., Simmer-Beck, M., & Bray, K. K. (2009). Examining the role of collaborative assessment in a didactic dental hygiene course. *Journal of Dental Education*, *73*(8), 980–990. https://doi.org/10.1002/j.0022-0337.2009.73.8.tb04787.x

Kim, B. (2001). Social constructivism. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology*, (pp. 55–61). https://cmapspublic2.ihmc.us/rid=1N5PWL1K5-24DX4GM-380D/Kim%20Social%20constructivism.pdf

Knierim, K., Turner, H., & Davis, R. K. (2015). Two-stage exams improve student learning in an introductory geology course: Logistics, attendance, and grades. *Journal of Geoscience Education*, *63*(2), 157–164. https://doi.org/10.5408/14-051.1

Koles, P. G., Stolfi, A., Borges, N. J., Nelson, S., & Parmelee, D. X. (2010). The impact of team-based learning on medical students' academic performance. *Academic Medicine: Journal of the Association of American Medical Colleges*, *85*(11), 1739–1745. https://doi.org/10.1097/ACM.0b013e3181f52bed

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

Levy, D., Svoronos, T., & Klinger, M. (2018). Two-stage examinations: Can examinations be more formative experiences? *Active Learning in Higher Education*, *24*(2), 79–94. https://doi.org/10.1177/1469787418801668

Lusk, M., & Conklin, L. (2003). Collaborative testing to promote learning. *Journal of Nursing Education*, *42*(3), 121–124. https://doi.org/10.3928/0148-4834-20030301-07

Maher, M. A. (2005). The evolving meaning and influence of cohort membership. *Innovative Higher Education*, *30*(3), 195–211. https://doi.org/10.1007/s10755-005-6304-5

Mahoney, J. W., & Harris-Reeves, B. (2019). The effects of collaborative testing on higher order thinking: Do the bright get brighter? *Active Learning in Higher Education*, *20*(1), 25–37. https://doi.org/10.1177/1469787417723243

Michaelsen, L. K., Black, R. H., & Fink, L. D. (1999). Problems with learning groups: An ounce of prevention. *Journal of Legal Studies Education*, *17*(2), 91–115. https://doi.org/10.1111/j.1744-1722.1999.tb00305.x

Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. *New Directions for Teaching and Learning*, *2008*(116), 7–27. https://doi.org/10.1002/tl.330

Michaelsen, L., Peterson, T. O., & Sweet, M. (2009). Building learning teams: The key to harnessing the power of small groups in management education. In S. J. Armstrong & C. V. Fukami (Eds.), *The SAGE Handbook of Management Learning, Education and* Development (pp. 325–343). https://doi.org/10.4135/9780857021038.n17

Minkler, J. E. (2002). ERIC Review: Learning communities at the community college. *Community College Review*, *30*(3), 46–63. https://doi.org/10.1177/009155210203000304

Mitchell, N., & Melton, S. (2003). Collaborative testing: An innovative approach to test taking. *Nurse Educator*, *28*(2), 95–97. https://doi.org/10.1097/00006223-200303000-00013

Mulryan-Kyne, C. (2010). Teaching large classes at college and university level: Challenges and opportunities. *Teaching in Higher Education*, *15*(2), 175–185. https://doi.org/10.1080/13562511003620001

Norris, C. J., & Barnett, B. (1994). *Cultivating a new leadership paradigm: From cohorts to communities*. Annual Meeting of the University Council of Educational Administration. https://eric.ed.gov/?id=ED387877

Rempel, B. P., Dirks, M. B., & McGinitie, E. G. (2021). Two-stage testing reduces student-perceived exam anxiety in introductory chemistry. *Journal of Chemical Education*, *98*(8), 2527–2535. https://doi.org/10.1021/acs.jchemed.1c00219

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.

Robillard, D. T., Spring, L. M., Pasquale, S. J., & Savageau, J. A. (2011, May 2). *Identifying characteristics of effective small group learning valued by medical students and facilitators* [Poster presentation]. Senior Scholars Program, UMass Chan Medical School, Worcester, MA, USA. https://doi.org/10.13028/ct6x-xc74

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rovai, A. P. (2002). Development of an instrument to measure classroom community. *The Internet and Higher Education*, *5*(3), 197–211. https://doi.org/10.1016/s1096-7516(02)00102-1

Saltiel, I. M., & Russo, C. S. (2001). *Cohort programming and learning: Improving educational experiences for adult learners (Professional practices in adult education and human resource development series)*. Krieger Publishing Company.

Sandahl, S. S. (2010). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives*, *31*(3), 142–147.

Saunders, F., Jasper, M., & Whitton, P. (Eds.) (2010). Promoting collaborative learning in engineering management education through the use of wikis. In *Engineering Education 2010: Inspiring the Next Generation of Engineers*, *EE 2010*. The Higher Education Academy Engineering Subject Centre.

Takahashi, M., & Saito, S. (2004). Does test delay eliminate collaborative inhibition? *Memory*, *12*(6), 722–731. https://doi.org/10.1080/09658210344000521

Teitel, L. (1997). Understanding and harnessing the power of the cohort model in preparing educational leaders. *Peabody Journal of Education*, *72*(2), 66–85. https://doi.org/10.1207/s15327930pje7202_4

Vygotsky, L. S. (1962). *Thought and language*. MIT Press.

Walker, J. D., & Robinson, D. H. (2022). Does two-stage testing promote long-term individual learning? *The Journal of Experimental Education*, *91*(3), 431–449. https://doi.org/10.1080/00220973.2022.2059750

Westervelt, E. (2016, April 14). *A nobel laureate's education plea: Revolutionize teaching*. NPR. https://www.npr.org/sections/ed/2016/04/14/465729968/a-nobel-laureates-education-plea-revolutionize-teaching

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*(4), 240–246. https://doi.org/10.1111/j.1467-9280.1992.tb00036.x

Williams, R. B. (2007). *Cooperative learning: A standard for high achievement*. Corwin.