The Canadian Journal for the Scholarship of Teaching and Learning

Volume 14 | Issue 2 Article 10

Fall 10-15-2023

The Influence of Two Stage Collaborative Testing on Peer Relationships: A Study of First Year University Student Perceptions

Brian Rempel
University of Alberta, brempel@ualberta.ca
Elizabeth McGinitie
University of Alberta, elizabeth.mcginitie@ualberta.ca
Maria Dirks
University of Alberta, mdirks@ualberta.ca

https://doi.org/10.5206/cjsotlrcacea.2023.2.14252

The Influence of Two-Stage Collaborative Testing on Peer Relationships: A Study of First Year University Student Perceptions

Abstract
Two-stage testing is a form of collaborative assessment that creates an active learning environment during test taking. In two-stage testing, students first complete an exam individually, and then complete a subset of the same questions as part of a learning team with the ultimate exam score being a weighted average of the individual and team portions. In the second (team-based) part of the exam, students are encouraged to discuss solutions until a consensus among team members is achieved, thus actively engaging students with course material and each other during the exam. A short open-ended survey was administered to students at the end of the semester, and the responses coded by thematic analysis, with themes generated using inductive coding based on the principles of grounded theory. The most important conclusion was that students overwhelmingly preferred two-stage tests for the development of positive peer relationships in class. The most common themes that emerged from student responses involved positive feelings from forced interaction with their peers, the benefits of meeting and socializing with other students, sharing of knowledge with others, and solidarity or positive affect towards the process of working as part of a team. Finally, students also expressed an overall preference for two-stage exams when compared to solely individual, one-stage exams.

Les examens en deux étapes sont une forme d'évaluation collaborative qui crée un environnement d'apprentissage actif pendant les examens. Dans le cadre d'un examen en deux étapes, les étudiants et les étudiantes passent d'abord un examen individuel, ensuite ils répondent à un sous-ensemble des mêmes questions dans le cadre d'une équipe d'apprentissage, le résultat final étant une moyenne pondérée de la partie individuelle et de la partie en équipe. Dans la deuxième partie de l'examen (réalisée en équipe), les étudiants et les étudiantes sont encouragés à discuter des solutions jusqu'à ce que tous les membres de l'équipe en arrivent à un consensus, ce qui permet aux étudiants et aux étudiantes de s'engager à la matière du cours et de participer entre eux durant l'examen. Un court sondage à questions ouvertes a été mené auprès des étudiants et des étudiantes à la fin du semestre et les réponses ont été codées par analyse thématique, les thèmes ayant été générés par un codage inductif basé sur les principes de la théorie ancrée. La conclusion la plus importante a été que les étudiants et les étudiantes ont massivement préféré les examens en deux étapes pour le développement de relations positives entre pairs, en classe. Les thèmes les plus communs qui ont été évoqués dans les réponses des étudiants et des étudiantes ont impliqué des sentiments positifs survenus suite à l'interaction forcée avec leurs pairs, les avantages de la rencontre et de la socialisation avec d'autres étudiants et étudiantes, le fait de partager les connaissances avec les autres et la solidarité ou l'affect positif à l'égard du processus de travail en équipe. Pour finir, les étudiants et les étudiantes ont également déclaré avoir une préférence générale pour les examens en deux étapes par rapport aux examens strictement individuels à une étape.

Keywords
collaborative learning, testing and assessment, peer relationships, first-year undergraduate; apprentissage collaboratif, examens et évaluations, relations entre pairs, première année du premier cycle

Cover Page Footnote
The authors wish to thank the students who agreed to participate as research subjects. We also thank Dr. James Kariuki for helpful comments and discussions, Dr. Glen Hvenegaard for assistance in survey design, and Dr. Shauna Wilton for guidance in thematic analysis.

This research paper/Rapport de recherche is available in The Canadian Journal for the Scholarship of Teaching and Learning: https://doi.org/10.5206/cjsotlrcacea.2023.2.14252

A sense of belonging and feeling socially supported are integral components to the success of students enrolled in post-secondary institutions. In particular, students who receive this social support generally experience lower levels of depression, loneliness and anxiety, are typically more engaged and motivated in their studies, achieve higher grades, and are more likely to graduate (Anistranski and Brown, 2021; Baumeister and Leary, 1995; Slaten et al., 2016). For first-year students especially, a sense of belonging and feeling socially supported are essential for a successful transition and adjustment to university life (Pittman and Richmond, 2008; Rayle & Chung, 2007; van der Zanden et al., 2018).

Factors that contribute to a student's sense of belonging at university include formation of positive student-instructor relationships, formation of positive peer relationships, feeling a sense of classroom connectedness, and participation in extracurricular activities (Freeman et al., 2007; Frisby & Martin, 2010; LaBelle & Johnson, 2020; Sollitto et al., 2013; Swenson et al., 2008). Of these, formation of positive peer relationships is particularly important as peer-peer interactions tend to impact all aspects of a student's university experience. Positive peer relationships also substantially contribute to a student's sense of social acceptance and support at university (Bronkema & Bowman, 2019; Buote et al., 2007; Goguen et al., 2010). The formation of positive peer relationships at university can prove difficult for many students. For example, students who experience social anxiety or shyness in social settings often have a difficult time connecting and actively engaging with peers (Brook and Willoughby, 2015; Shell and Absher, 2019). As well, marginalized students are more likely to experience some degree of social isolation at university due to uncertainties regarding welcoming and social inclusion from nonmarginalized peers (Baker and Robnett, 2012; Fischer, 2007; Khalis et al., 2018; Vaccaro and Newman, 2017).

One intervention to facilitate the formation of positive peer relationships among university students is the incorporation of in-class collaborative, active learning opportunities for students that involve meaningful peer discussions (Frisby and Martin, 2010; LaBelle and Johnson, 2020). Collaborative, active learning activities provide students with the opportunity to interact with their peers and discuss or debate course concepts in a small group or team setting. These activities ideally occur in a structured environment and are moderated by a course instructor and teaching assistants to encourage civil behaviour and promote positive, respectful student discourse. Previous research has shown that such in-class collaborative, active learning activities can promote solidarity and more intimate, familiar communication among students. Furthermore, in-class collaborative, active learning has been shown to facilitate the creation of positive peer relationships and friendships among students (Rasco et al., 2020; Sato, 2013; Summers et al., 2009). Collaborative learning activities may be particularly impactful for students who otherwise struggle with connecting and communicating with their peers, and the role of active student engagement has a clearly demonstrated positive impact on course outcomes for students who otherwise struggle in the course (Freeman et al., 2014).

To assess student knowledge, individually-completed summative written exams are a common element in many university courses. Some of the documented problems with individual summative exams include the fact that they have been identified as triggers for student anxiety, may not accurately capture student knowledge, and can present inequitable barriers for students with language or learning challenges (Cantwell et al., 2017; Lusk & Conklin, 2003). Two-stage testing, a form of collaborative testing that falls under the broad umbrella of collaborative, active learning, has emerged as an alternative method for assessing student knowledge and could address some of the drawbacks of summative written exams (Meseke et al., 2008; Mitchell & Melton, 2003; Rieger & Heiner, 2014; Slusser & Erickson, 2006; Zipp, 2007). In many forms of two-stage

testing, students first individually complete and submit the written exam. Then in the second stage, students complete some or all of the same exam questions with a small group of their peers and submit either a single copy or individual copies of the team exam. The role of the latter part of the exam is to encourage collaborative discussion among students, facilitate active learning through test taking, and create a peer-led learning opportunity for students during the exam itself. There have been reports on the use of two-stage exams in a variety of disciplines, including nursing (Burgess & Medina-Smuck, 2018; Caboral-Stevens & Fox, 2020; Green et al., 2018), dentistry (Jones, 2019; Keselyak et al., 2009), chemistry (Kloepper, 2015; Rempel et al., 2021), biology (Green et al., 2016; Khong and Tanner, 2021), business (Kruger, 2019; Schmulian and Coetzee, 2019), and statistics (Eastridge and Benson, 2020; Kapitanoff, 2018). There has also been a great deal of excellent research demonstrating the benefits of two-stage exams, including increased knowledge retention, improved understanding of course material due to immediate feedback, higher student engagement in the learning process, decreased exam anxiety, enhanced communication skills and positive peer relationships (Breedlove et al., 2004; Smith et al., 2009; Vazquez-Garcia, 2018; Vogler and Robinson, 2016).

Recently, we adopted two-stage exams in some first- and second-year courses. As part of this process, we also initiated a research project to examine student perceptions and responses to the use of two-stage exams in our courses and context. In particular, we are interested in learning what students, as the primary participants in the process, think about two-stage exams. Generally speaking, qualitative research on student perceptions of two-stage exams is largely absent from the literature. In an effort to address this need, we recently published our findings on student perceptions of two-stage exams as they relate to exam anxiety. Briefly, our study thematically-coded student statements about potential impacts of two-stage exams on self-perceived exam anxiety. Student responses were generally positive, and key thematic reasons for the perception that two-stage exams helped reduce exam anxiety were identified (Rempel et al., 2021). The same students were also asked to separately describe how two-stage exams compared to one-stage exams when considering their comfort and ability to express their understanding of course content, and how both exams impacted their relationships with their peers.

In the present work, data is analyzed from the question of how students perceive two-stage versus one-stage exams in terms of their impacts on peer relationships. Our research study focused on whether two-stage exams produced a positive or negative impact on students' relationships with their peers when compared to their prior experience with one-stage exams in other courses, as well as what factors influenced students' conclusions. The objectives for this paper are threefold:

1) Describe a method for implementation of two-stage exams that was successful from the instructor perspective and generated very positive student responses.
2) Provide evidence from student perceptions that the active learning component of the two-stage exam fosters the development of positive peer relationships.
3) Identify and describe the major themes that students considered when describing the impact of the two-stage exam on their peer relationships.

Our results should provide insight to post-secondary instructors wanting to incorporate more active learning during the evaluative portion of their course, both an effective method, but also some of the key themes their students could express in response to two-stage exams.

Method

The research described herein was conducted in the Augustana Faculty of the University of Alberta in the Fall semesters of 2018 and 2019. Augustana is a small, rural, primarily residential undergraduate Liberal Arts and Sciences campus of the University of Alberta which is itself a large, research-intensive institution. All research subjects were students registered in one of three class sections in the first semester of a typical 1st-year General Chemistry course. Each section was taught by one of the authors, either Instructor #1 (Fall 2018 and 2019) or Instructor #2 (Fall 2018). All three course sections used in-class time for a mixture of lecture and active-learning (problem-solving) activities. Each section covered the same content material in the same order. All three sections were 90-minutes long and held near the middle of the day (between 12:00 PM and 2:00 PM) twice weekly.

Ethics approval for this research was obtained from the University of Alberta Research Ethics Office, and all students provided informed consent to participate. As discussed in the Limitations and Future Directions section below, no specific demographic information was collected on the research participants. However, some general statements can be made based on the overall makeup of the student population at Augustana and both instructors' observations. First, most students were first-time university students who had recently finished their secondary (high school) education and were enrolled in the first year of their university education. Second, the approximate proportion of female to male students was 3:2, which reflects the self-declared gender identity for incoming Augustana students. Third, the vast majority of students were registered in a B.Sc. program with the most commonly-declared major being Biology. Lastly, the proportion of international to domestic students was roughly 1:4, with the majority of international students coming from countries in which English is not the primary language. Neither instructor noted a difference in class performances between the three class sections as judged by final grades.

Structure of Two-Stage Quizzes and Exams

In-class quizzes and midterm exams in all three class sections used the two-stage format with only very slight differences. In all sections, the quiz was administered one or two weeks prior to the midterm exam both to familiarize students with the format, but also to model for students the type of questions they could expect on the exam. The material on which students were examined was the same across all sections. On the day of the assessment, the assessment was first individually completed by students. After students submitted their paper, they would gather with a team of other students to collaboratively work on the team copy which contained a subset of the same questions. The individual:group portions of the assessments were weighted at approximately 80:20 in all three sections, meaning that a students' final score was still mostly based on their individual work. Additionally, if a student's individual score was higher than their blended score, then only the individual score was used when calculating final course grades. There were only slight differences in two-stage assessment implementation between the two course instructors, as can be seen in Table 1.

Table 1

Details of the Structure and Implementation of Two-Stage Assessments in the Three Class Sections

Section (Instructor)	Timing Between Quiz and Exam	Time Provided in Minutes (Individual / Team)	Relative Weighting (Individual / Team)	% Weight of Final Grade	Research Participants / Total Student Enrollment	Team Structure
#1 Fall 2018 (Instructor #1)	2 weeks	35 / 15 (Quizzes) 65 / 25 (Exams)	70 / 30 (Quizzes) 85 / 15 (Exams)	3.5% each (Quizzes) 15% each (Exams)	49 / 55 (89%)	Formed for the assessment only
#2 Fall 2018 (Instructor #2)	1 week	30 / 30 (Quizzes) 60 / 30 (Exams)	N/A[a] (Quizzes) 80 / 20 (Exams)	2.5% each (Quizzes) 18% each (Exams)	39 / 46 (85%)	Permanent throughout the term
#3 Fall 2019 (Instructor #2)	1 week	30 / 30 (Quizzes) 60 / 30 (Exams)	N/A[a] (Quizzes) 80 / 20 (Exams)	2.5% each (Quizzes) 18% each (Exams)	41 / 60 (68%)	Permanent throughout the term

Note. [a] Quizzes were scored on a participation basis, with completion earning a score of 100% for purposes of calculating Final Grades.

Students who required exam accommodations through Accessibility Resources were given the option to complete the assessments individually without completing the group portion. In such a scenario, student grades were derived entirely from the individual portion of the assessment. This option was provided to ensure that these students were not inadvertently forced to disclose their need for exam accommodations to their peers. Alternatively, these students were also given the option to complete the assessments as a two-stage exercise with their accommodations activated for the individual portion of the assessment before they joined their peers for the group portion. Students requiring extended examination time needed to schedule the individual portion of the assessments such that they would be finished the individual portion at the same time as their peers.

Survey Design and Implementation

An open-ended survey with four questions (see Table 2) was designed to prompt student responses when comparing two-stage exams to one-stage exams with a specific focus on three elements: 1) expression of knowledge, 2) exam anxiety, and 3) peer relationships. The final question asked students to provide their overall opinion on their preference for two-stage versus one-stage exams. The initial wording for each survey question was vetted by a colleague with extensive experience in survey design, and the final questions' wording (see Table 2) crafted based on that expert feedback.

Table 2

Two-Stage Exam Survey Questions Administered to Student Participants

Question	Description
Question 1 (expression of knowledge)[a]	Which was better for letting you express your mastery of the knowledge/conceptual understanding of the material being assessed: two-stage or one-stage exams? Please explain the rationale for your answer.
Question 2 (exam anxiety)[b]	Which exam format better reduced your exam anxiety: one-stage or two-stage exams? Please explain the rationale for your answer.
Question 3 (peer relationships)	Please explain to us how the different exam formats (one-stage and two-stage) impacted your relationships with your classmates. Did they strengthen, deepen or affect in any way your relationships? Did one do this better than the other?
Question 4 (format preference)	Which exam format (one-stage or two-stage) did you prefer for the assessment of your learning of the course material?

Note. [a] Not analyzed in the present work. [b]See (Rempel et al., 2021) for complete analysis.

Paper surveys were administered in the first fifteen minutes of class during the final two weeks of the semester by a member of the research team who was not the instructor for a given class section. The student participants had no prior connection with the research team member who described the project and collected responses. Students were given verbal instructions to compare their experience of one-stage exams from other classes or prior experiences to their experience with two-stage exams in General Chemistry. All surveys were collected anonymously with no identifying or demographic information, and were analyzed at the conclusion of the data collection period.

Thematic Coding of Data

Student responses were transcribed into Google Docs on a secure server and a random subset of responses were selected. These responses were independently read by the research team members, and preliminary thematic codes were prepared. Thematic codes (Bengtsson, 2016; Hsieh and Shannon, 2005) were developed using inductive coding based on the principles of grounded theory (Urquhart, 2013) in which no theoretical framework was applied beforehand, and instead the codes emerged from student responses. The team then discussed how to code each response, and the preliminary codes were arrived at by consensus. Affective student responses that were not definitively positive or negative were discussed at length by the research team (which included a student researcher) until consensus was reached. In some instances, multiple affective codes were applied to student responses containing more than one remark. These codes were used by each research team member to independently analyze a second subset of data. A second round of discussion led to simplification and clarification of what types of responses belonged under individual codes. Using this second set of codes, themes were identified and related codes grouped under each relevant theme. Finally, inter-coder agreement was established by having a final subset of survey responses examined and coded separately by each author. A final round of discussion produced agreement on the types of responses that belonged under each code. Inter-coder

agreement was measured by calculating the % agreement (96%) between all three members of the research team.

Lastly, all of the student surveys were analyzed using the final set of codes generated (see Table 3 in the Results section). Each author examined the entire data set and initial results were recorded separately, meaning each response was examined three separate times. Codes were tallied using Google sheets, and were formally assigned to a given response if at least two of the three team members had applied it. Note that while all four questions on the survey were coded at the same time, only the results from Question 3 (on peer relationships) and Question 4 (on overall preference) are analyzed here. Results from the analysis on Question 2 (on exam anxiety) have already been reported elsewhere (Rempel et al., 2021), and the results of analysis on Question 1 (on expression of knowledge) are forthcoming.

Data Analysis

There was no prior reason to assume that the data from any of the three class sections should be different, so the decision was made to pool the results from all three class sections. To quickly verify that no class data set stood out from the others given the numbers in the data set, the percentage of each class with a given thematically-coded response was compared for each code (see Supporting Information). On the basis of rudimentary visual analysis, it was noted that none of the three classes had strikingly different values or abnormal trends in responses. Based on this simple examination, separate class section data sets were pooled for all subsequent analysis. A detailed statistical analysis to rigorously prove the validity of this assumption is beyond the scope of the present work, given the qualitative nature of the research project and modest number of respondents.

Results and Discussion

Instructor Concerns

As discussed above, there is a substantial body of evidence for the usefulness of collaborative testing in university education. When implementing two-stage exams, both instructors had a few considerations and concerns in mind. Firstly, it was important that student grades should not decrease as a result of the team part of the two-stage assessment. This was implemented by separately calculating and taking the highest of each students' individual exam score weighted at 100% of the exam score, or with the team portion weighted at between 15-30% of the exam score. This grade guarantee was intended to reassure students unfamiliar with two-stage exams that there was no potential for the (presumably) novel team portion of the exam to lower their grades. Given how students often identify grades as highly motivating, the aim was to reduce students' concerns with grades when engaging with the collaborative portion of the exam. Both instructors also acknowledged it was possible that students might have different experiences with two-stage exams. By guaranteeing that student grades could not be negatively impacted by the two-stage exam, the instructors hoped to reduce student anxiety about factors such as gender, ethnicity and race, language proficiency, learning challenges, neurodiversity, etc. from having any potential negative impact on exam performance.

One concern shared by both instructors prior to implementation of two-stage exams was the potential for grade inflation to occur because students' grades would not decrease if they scored

lower on the team portion of the exam. Additionally, if such grade inflation occurred, then the ability to discriminate between high-performing and low-performing students might be impaired. In practice, these concerns proved unfounded. An examination of the relative weighting of the team assessments showed that grade inflation was unlikely. On average, the team portion of the midterm exam was ~20% higher than the average individual score, a value deemed typical based on conversations with colleagues at other institutions. This ~20% increase for the two midterm exams, which are collectively worth approximately 32% of the final grade, produced an overall average grade increase of only 0.96%. This relatively modest increase in student grades ended up being less than the typical division between letter grades, such that overall class GPA did not change substantially. In practice, while direct comparisons of class GPAs between successive years (before and after implementation of two-stage exams) have too many confounding factors to be directly useful, both course instructors did not experience any difficulty in assigning letter grades that adequately differentiated lower- and higher-performing students.

Although there was not a meaningful impact on student grades, collaborative testing in general has been shown to produce learning gains for all students (Jang et al., 2017; Mahoney & Harris-Reeves, 2019), with especially positively impacts on learning outcomes for typically lower-performing students (Bruno et al., 2017; Giuliodori et al., 2008). Therefore, two-stage exam implementation was designed to minimize grade inflation while harnessing some of the other cognitive and social benefits of collaborative testing.

On the other end of the grade distribution, there is another potential concern that typically-high-performing students might choose not to participate with the team portion of the exam and adopt a "zero-sum game" mentality towards trying to maintain a higher relative standing in class. This instructor concern originated from informal conversations with past students who held the misconception that grades are assigned from a pre-defined ranking system (curve). Fortunately, in our analysis of student responses from all three of Questions 1-3, this theme was entirely absent. Informal conversations with current students and direct observation of group dynamics during the team portion of the two-stage exam also provided no evidence that high-performing students refused to participate in order to cause a relative decrease in the class average.

Both instructors were initially concerned with the compatibility of the two-stage testing format for students requiring exam accommodations. However, no negative student feedback was obtained from survey responses or private conversations with students related to exam accommodations. In fact, the two-stage exam format was easily adapted and implemented for students requiring exam accommodations provided instructors and students worked closely with Accessibility Resources staff during assessment scheduling.

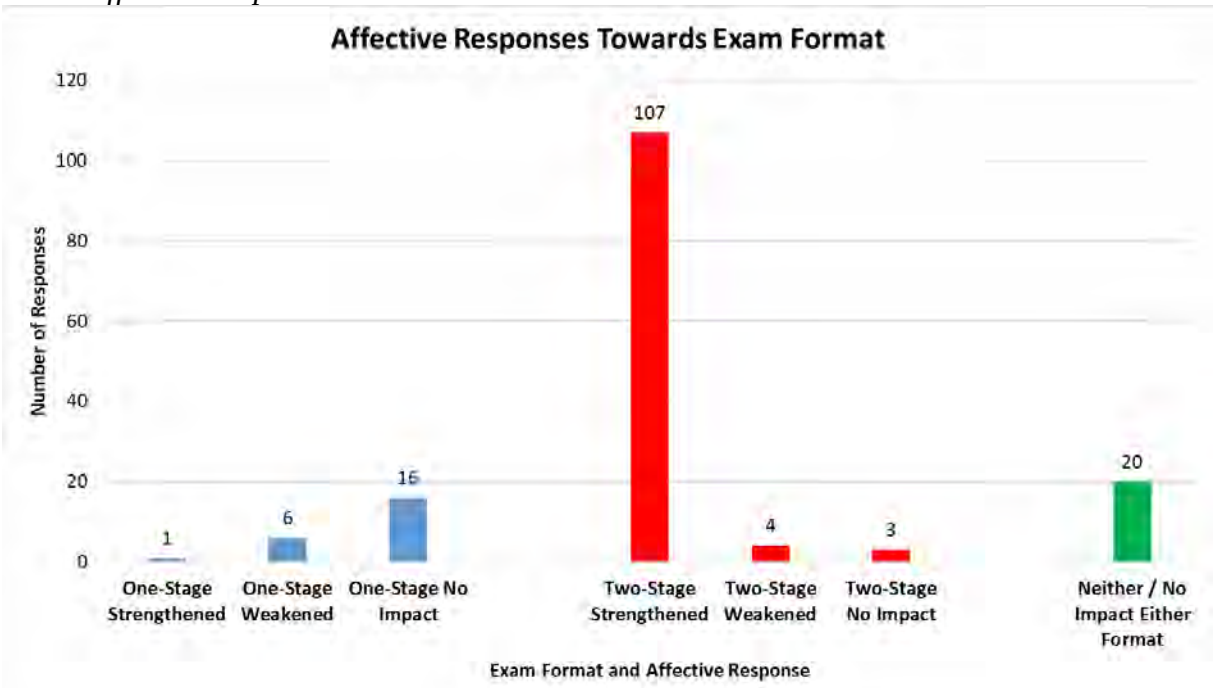
Coding of Data

The data was simultaneously coded using two broad categories: 1) affective responses towards the exam format, and 2) thematically-coded explanations. Therefore, a statement could simultaneously have two codes for both how the student felt about a given assessment, and their rationale explaining that feeling. For example: "Two-stage exams did help strengthen my relationship with classmates. It helped me to meet and interact with a few new people at the beginning of the year." This was coded as having a positive affect for two-stage exams, and with the theme 'meeting others' and the code 'meeting new people', as summarized below. Note that the above and all following quotations were taken directly from student responses and are presented without editing.

Figure 1 shows a summary of student responses in terms of affective statements towards each assessment format. Student statements were coded as having positive, negative, or neutral affect towards each exam format (one-stage or two-stage). The purpose of this layer of coding was to figure out if responses were being driven by positive or negative attitudes towards one of the two exam formats. Note that the total number of responses (n = 157) is actually greater than the number of participants (n = 129). This discrepancy arises because in some cases, student responses contained more than one statement that could be coded. For example: "A two-stage exam generated discussion during the exam, where students could debate which answer is correct. A one-stage exam isolates students from others, until after the exam." This answer received codes for having both a positive affective statement towards two-stage exams (first sentence), as well as a negative affective statement towards one-stage exams (second sentence). In fact, some responses could contain seemingly contradictory statements, for example: "[The two-stage exam] definitely strengthened my relationship with some people as I talked with them and realized if they think and study the way I do and on the other hand it make me distant from others as some people really didn't care for the course and they were the one that didn't know anything in the group exam." This was coded as having both positive and negative affective statements towards two-stage exams.

Figure 1

Student Affective Responses to Exam Format

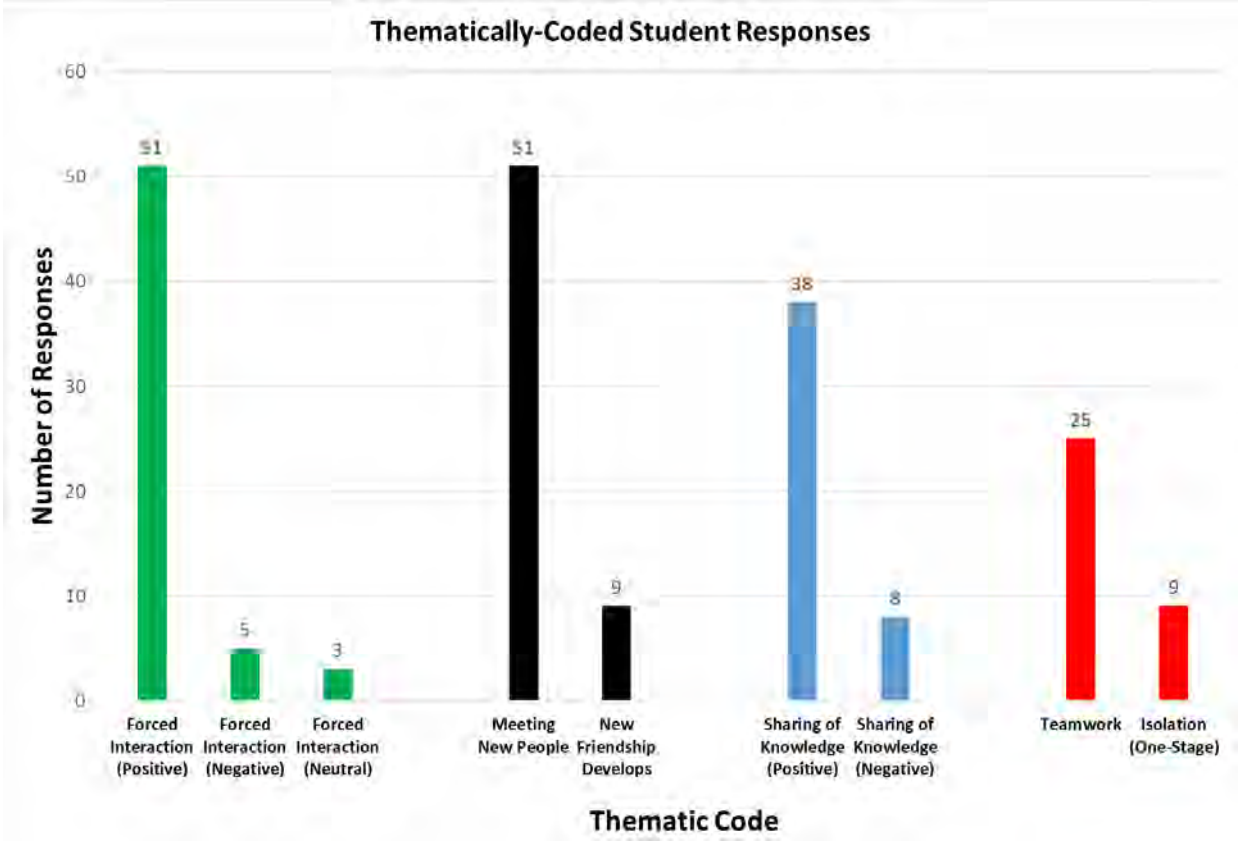


Note. Graphical summary of affective student responses (n = 129) in response to the question, "Please explain to us how the different exam formats (one-stage and two-stage) impacted your relationships with your classmates. Did they strengthen, deepen or affect in any way your relationships? Did one do this better than the other?"

Figure 2 shows a summary of the total numbers of thematically-coded explanations for why students were selecting one exam format or the other in terms of perceived impact on peer relationships. As discussed for the affective codes, the total number of codes applied (n = 199) was greater than the number of responses received (n = 129) because responses could have one more than one statement representing more than one theme. For example: "Two-stage exams allowed me to bond more often with people that I do not usually talk to and it also allowed me to view how other students tackle problems different than me." This statement received thematic codes for both 'meeting new people' as well as 'sharing of knowledge (positive)' in addition to an affective code for 'two-stage strengthened'.

Figure 2

Impact of Different Exam Formats on Students' Relationships



Note. Graphical summary of thematically-coded student explanations (n = 129) in response to the question, "Please explain to us how the different exam formats (one-stage and two-stage) impacted your relationships with your classmates. Did they strengthen, deepen or affect in any way your relationships? Did one do this better than the other?"

For a detailed breakdown of the nine explanatory codes that were developed, their associated theme, and some representative responses, see Table 3.

Table 3

Description of the Thematic Codes Developed by the Research Team with Representative Student Responses

Code	Description	Theme	Representative Responses
Forced Interaction (Positive)	Positive statement about being forced to interact with peers.	Forced Interaction	"…you worked with people you may not have talked to before." "…allowed me to meet more people…"
Forced Interaction (Negative)	Negative statement about being forced to interact with peers.		"People with anxiety have trouble being thrown into new social groups." "…some people were unpleasant to work with."
Forced interaction (Neutral)	Acknowledged being forced to interact with peers without clearly expressed positive or negative feeling.		"What little interaction that occurred during the two-stage between individuals was only for the duration of the exam and nothing more." "I talked with my classmates a bit more but otherwise I don't really think it made a difference."
Meeting New People	Met peers in class on a casual level.	Meeting Others	"It helped me meet and interact with a few new people at the beginning of the year." "…you worked with people you may not have talked to before."
New Friendship Develops	Meeting in class led to social contact outside of class.		"The two-stage help me to be free to ask questions to my classmates even outside of the class and strengthen my relationship with them." "…the two-stage exam helped introduce me to peers that would turn out to be helpful when it comes to studying later on in the course."

Code	Description	Theme	Representative Responses
Sharing of Knowledge (Positive)	Collaboration with peers was a positive experience.	Sharing of Knowledge	"The two-stage helped grow our relationships as we had to learn each others strengths and weaknesses in order to help improve the learning of the whole group." "Two-stage exams helped me….hear feedback which I think is very beneficial."
Sharing of Knowledge (Negative)	Collaboration with peers was a negative experience.		"But when we would disagree on an answer, it created a bit of tension." "In other cases, the relationships were weakened because there was a discrepancy of who had the correct answer."
Teamwork	Success or failure depended on teamwork.	Solidarity	"…all be in the same boat about struggling with a question or all completely getting the question." "…we all were working towards a common goal."
Isolation (One-Stage)	Writing individual exams produced feelings of isolation.		"One stage exams typically…increase the competition, dislike, and distrust present between students." "One stage separated group members as a more competitive "right or wrong" mentality arises."

Impact of Assessment Format on Peer Relationships

As shown in Figure 1, an overwhelming majority of students perceived that two-stage exams produced more of a strengthening rather than weakening of peer relationships. As well, students also perceived that two-stage exams had a much larger overall impact than one-stage exams on peer relationships in general. This latter point is perhaps unsurprising, given that two-stage exams have a directly collaborative, social component in the team portion of the exam whereas one-stage exams do not. The most common response when students directly discussed one-stage exams indicated they did not think one-stage exams had any impact on their peer relationships which is easy to understand given their essentially solitary nature. The results presented herein are consistent with the idea that the majority of students perceive the collaborative element as more positive than negative. However, for a notable minority of students, the perception that negative group dynamics produce an overall unfavourable impression of two-stage exams was instead true. That observation suggests the role of the instructor is crucial in the collaborative exam process to limit student incivility and foster positive group dynamics. Also note that despite the fact Class Section #1 had teams formed only for the duration of the exam whereas Class Sections #2 and #3 had teams that were permanent throughout the term, the same trends in student responses were observed across all three sections. In the absence of any other intervention, this result suggests that two-stage exams are beneficial to students for forming positive peer relationships regardless of the specific classroom environment.

Another key result is that students' positive affect towards two-stage exams is not simply a reflection of increased grades. Indeed, the team portions of the exams still produced some cases in which teams failed the collaborative portion. The fact that the distribution of exam scores on the collaborative portion still looked typical, with both high- and low-performing teams, argues against an interpretation that students felt the two-stage exam improved their relationships with their peers simply because it boosted grades. The actual impact on student grades was less than 1% on average, as described above and communicated to students at the beginning of term.

The thematically-coded explanations (Figure 2) help provide a better understanding about why students believed that two-stage exams produced a positive effect on their peer relationships. The types of student responses could be grouped into four major themes: forced interaction, meeting others, sharing of knowledge, and solidarity. Within each theme, codes were applied that either classified that concept as positive, negative, or neutral (forced interaction, sharing of knowledge, and solidarity) or a thematic concept that had different degrees of positive (meeting others).

The theme of forced interaction emerged for students who described how the collaborative portion of the exam created both the opportunity and need to introduce themselves and talk to other students in the class. In many cases, students favourably identified these introductions as something they did not believe would happen otherwise. The minority of students who identified forced interaction as being viewed negatively typically mentioned either concerns about anxiety or a difficult teammate. For example: "…there were simply some people who thought they were always right even if wrong so when choosing groups they would be avoided at times."

However, it is also useful to note that not all students who find social introductions difficult viewed the forced interaction negatively. For example: "I am generally a shy individual, so I didn't as much, I wish I did, but when [Course Instructor] placed us in groups it really helped me to form connections I would personally not initiate." This statement shows that at least some students who don't find social introductions easy held positive feelings towards the forced interactions with

others. Nonetheless, it is useful to note that while for the majority of students the collaborative learning environment was viewed positively, there still existed a minority of students who perceived that negative group dynamics were more important.

The second theme that emerged is making friends, which is linked to forced interaction but encompasses comments that suggested social contact beyond simple interaction. In some cases, what students described were positive social interactions more substantial than only sharing names and a common task, and occasionally even led to social contact outside of class. The two codes developed reflected the more common (n = 51) student experience of meeting peers in class and having a chance to visit and work with them, as well as the less common (n = 9) case in which students met peers and eventually used that opportunity to see their peers outside of class (either socially or for studying).

The third theme surrounded the sharing of knowledge during the collaborative portion of the assessment. For the majority of students, this sharing was viewed as a positive interaction with their peers, with comments that two-stage exams let them "…connect with other students….and let me see the exam through someone else's eyes." There were also some student comments that indicated they viewed the sharing of knowledge in a negative way, typically coupled with remarks about poor communication within the group or personality conflicts. Some remarks were nuanced, and acknowledged that both positive and negative affects could arise from the sharing of knowledge. For example: "In some cases, the relationships were strengthened, when we were able to help one another. In other cases, the relationships were weakened because there was a discrepancy of who had the correct answer." This describes both the benefits of sharing knowledge, but also the tensions that can sometimes arise when collaborating.

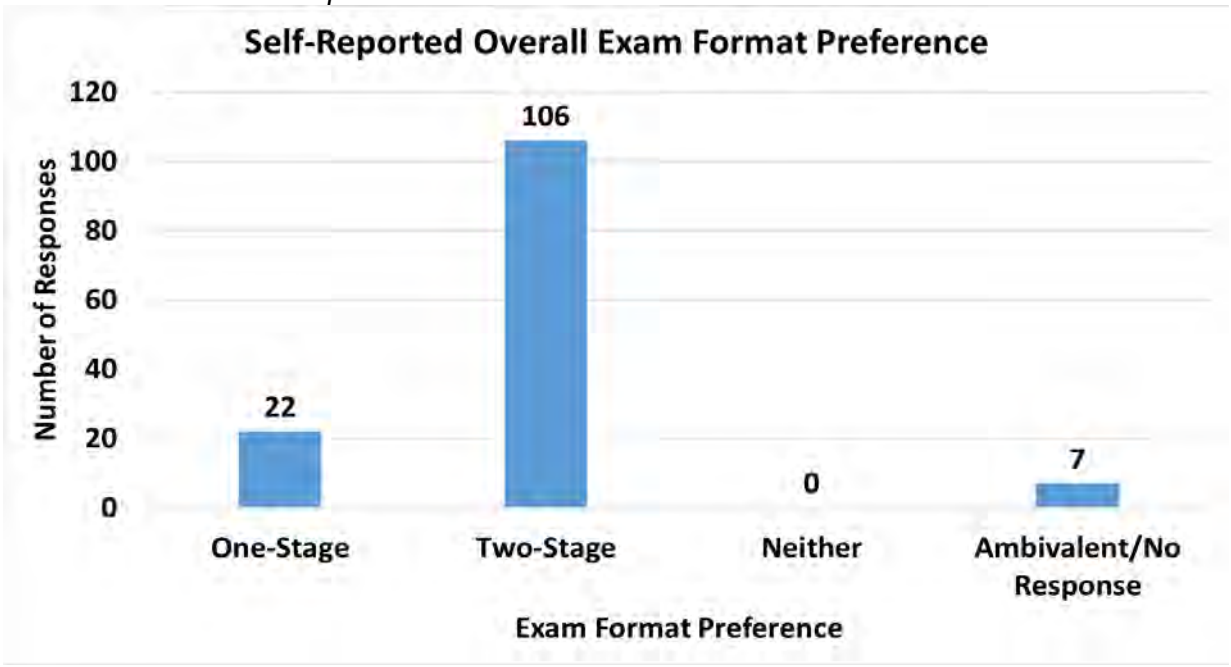
Finally, and closely related to the sharing of knowledge in a collaborative setting, is the theme of solidarity. Students remarked positively on the impact of the team portion of the two-stage exam as increasing cohesion, with comments such as "The two-staged exam strengthened the trust I had in my team because we completed the exam together" and "I found my teamwork skills got better." It is useful to note that many of the responses thematically described as solidarity and specifically coded under teamwork (for two-stage exams) were, like the above quote, focused less on the outcome (i.e., getting the right or wrong answer) but instead concentrated on the benefits of the process of working collaboratively. Reinforcing this viewpoint, some students (n = 9) also specifically commented on how the process of completing exams only individually increased negative feelings towards their peers ("One stage makes it feel more like a competition against other students") or increased feelings of being alone or isolated, as can be seen in the following quotation: "A one-stage exam isolates students from others, until after the exam."

Overall Assessment Preference

Question 4 of the survey was designed to provide direct feedback to the course instructors on student preferences for exam format. The overall preferences data can be seen in Figure 3. Note that, as described previously, there were some student responses that were coded with more than one code when ambivalent statements were included in the responses in addition to simple statements of preference meaning the total number of responses (n = 135) is slightly greater than the number of participants (n = 129).

Figure 3

Student Exam Format Preference



Note. Graphical summary of the numbers of stated preferences (n = 129) in response to the question, "Which exam format (one-stage or two-stage) did you prefer for the assessment of your learning of the course material?"

As can be seen in Figure 3, there is an overwhelming student preference for two-stage exams. Students clearly expressed that they preferred the two-stage exam format over the more typical one-stage format.

A slight majority of students also provided a short explanation for their choice. In such cases, the explanation was often a variation on one or more of the themes that appeared in their answers to one of Questions 1-3. It was decided not to analyze these explanations in Question 4 for two reasons. Firstly, the question had not prompted students to provide a rationale for their answers, meaning that many students (understandably) did not. Therefore, the explanations given in Question 4 were not necessarily representative of what all groups of students thought. Second, the authenticity of these responses was suspect given the placement of this question in the survey. The fact that almost all of the 66 respondents provided a rationale that echoed one of the themes they mentioned in Questions 1-3 suggests that students' thinking may have been influenced by the prompts to discuss the subjects in Questions 1-3. Based on these factors, it was decided to ignore all explanations in the responses to Question 4 and only consider simple expressions of assessment preference.

Limitations and Future Directions

The work described above is reflective of student perceptions rather than based on an outside measure of how two-stage exams objectively impact student relationships. Furthermore, the magnitude of any potential effect is impossible to judge from the present work. Among the myriad out-of-class activities that typically accompany post-secondary life, it is unlikely that the exam format used in one class is a major factor for a typical students' social development.

Nonetheless, the present results still show that students believe the two-stage exam format is a better choice for facilitating development of peer relationships both inside and, in some cases, outside of class.

Another limitation is the focus of student remarks on two-stage exams. Because two-stage exams were the only assessment format used in the class sections studied, students may have focused their responses more on the two-stage exam. Whether that increased frequency is because of the probable novelty of the collaborative assessment format, the fact that students overwhelmingly preferred the two-stage exams, or because two-stage exams were the only assessment format encountered in the course cannot be determined from the present data.

Finally, there is no demographic information available on the respondents other than general observations. Specifically, there is no information on whether or not students of (for example) different genders, ethnicities, or academic backgrounds all view two-stage exams as providing the same experience and the same potential benefits and drawbacks. We are similarly unable to comment on marginalized student experiences vs nonmarginalized student experiences with two-stage exams at this time. Investigating whether there are differences in the results obtained among different student populations is the focus of a planned future study. Once we have a more complete picture of how students engage with two-stage exams and their concerns related to peer interactions, we will be better equipped to facilitate in-class discussions focused on equitable and productive teamwork.

Conclusions

The key result from this study was that students overwhelmingly preferred two-stage exams for producing a positive impact on their relationships with their peers. The major themes that emerged in student responses centered on the forced interaction with peers (generally regarded as positive), the experience of sharing knowledge with peers (generally regarded as positive), positive feelings of solidarity found in the teamwork present in two-stage exams, and facilitation of positive social interactions both within and, for a small number of students, outside of class. Lastly, students also reported an overall strong preference for two-stage exams with an individual part followed by a collaborative team portion. Therefore, while all three course sections described herein were run as a mixture of lecture and active learning, these results suggest that students believe even a relatively simple intervention such as a two-stage exam can positively impact peer relationships in class.

References

Anistranski, J.A., & Brown, B.B. (2021). A little help from their friends? How social factors relate to students' sense of belonging at a large public university. *Journal of College Student Retention: Research, Theory & Practice*. SAGE Publications. https://doi.org/10.1177/1521025120985107

Baker, C.N., & Robnett, B. (2012). Race, social support and college student retention: A case study. *Journal of College Student Development*, *53*(2), 325–335. https://doi.org/10.1353/csd.2012.0025

Baumeister, R.F. & Leary, M.R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin 117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open 2*8–14. https://doi.org/10.1016/j.npls.2016.01.001

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

Bronkema, R.H., & Bowman, N.A. (2019). Close campus friendships and college student success. *Journal of College Student Retention: Research, Theory & Practice, 21*(3), 270–285. https://doi.org/10.1177/1521025117704200

Brook, C.A., & Willoughby, T. (2015). The social ties that bind: Social anxiety and academic achievement across the university years. *Journal of Youth and Adolescence 44*(5), 1139–1152. https://doi.org/10.1007/s10964-015-0262-8

Bruno, B.C., Engels, J., Ito, G., Gillis-Davis, J., Dulai, H., Carter, G., Fletcher, C., & Böttjer-Wilson, D. (2017) Two-stage exams A powerful tool for reducing the achievement gap in undergraduate oceanography and geology classes. *Oceanography 30*(2), 198–208. https://doi.org/10.5670/oceanog.2017.241

Buote, V.M., Pancer, S.M., Pratt, M.W., Adams, G., Birine-Lefcovitch, S., Polivy, J., & Wintre, M.G. (2007) The importance of friends: Friendship and adjustment among 1st-year university students. *Journal of Adolescent Research, 22*(6), 665–689. https://doi.org/10.1177/0743558407306344

Burgess, A. & Medina-Smuck, M. (2018). Collaborative testing using quizzes as a method to improve undergraduate nursing student engagement and interaction. *Nursing Education Perspectives, 39*(3),178–179. https://doi.org/10.1097/01.NEP.0000000000000223

Caboral-Stevens, M., & Fox, D.P. (2020). The use of collaborative testing with baccalaureate nursing students. *Teaching and Learning in Nursing, 15*(1), 37–41. https://doi.org/10.1016/j.teln.2019.09.007

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/https://doi.org/10.1002/cl2.186

Eastridge, J.A., & Benson, W.L. (2020) Comparing two models of collaborative testing for teaching statistics. *Teaching of Psychology, 47*(1), 68–73. https://doi.org/10.1177/0098628319888113

Fischer, M.J. (2007). Settling into campus life: Differences by race/ethnicity in college involvement and outcomes. *The Journal of Higher Education, 78*(2), 125–156. https://doi.org/10.1353/jhe.2007.0009

Freeman, S., Eddy, S.L., McDonough, M., & 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

Freeman, T.M., Anderman, L.H., & Jensen, J.M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *The Journal of Experimental Education 75*(3), 203–220. https://doi.org/10.3200/JEXE.75.3.203-220.

Frisby, B.N., & Martin, M.M. (2010). Instructor–student and student–student rapport in the classroom. *Communication Education 59*(2), 146–164. https://doi.org/10.1080/03634520903564362

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

Goguen, L.M.S., Hiester, M.A., & Nordstrom, A.H. (2010). Associations among peer relationships, academic achievement, and persistence in college. *Journal of College Student Retention: Research, Theory & Practice 12*(3), 319–337. https://doi.org/10.2190/CS.12.3.d

Green, R., Worthey, T., & Kerven, J. (2018). Collaborative testing: An effective invitational strategy for high-stakes testing in nursing. *Journal of Nursing Education 57*(5), 291–295. https://doi.org/10.3928/01484834-20180420-07

Green, R.A., Cates, T., White, L., & Farchione, D. (2016). Do collaborative practical tests encourage student-centered active learning of gross anatomy? *Anatomical Sciences Education 9*(3), 231–237. https://doi.org/10.1002/ase.1564

Hsieh, H-F., & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Jang, H., Lasry, N., Miller, K., & Mazur, E. (2017) Collaborative exams: Cheating? Or learning? *American Journal of Physics 85*(3), 223–227. https://doi.org/10.1119/1.4974744

Jones, T.A. (2019). Effect of collaborative group testing on dental students' performance and perceived learning in an introductory comprehensive care course. *Journal of Dental Education 83*(1), 88–93. https://doi.org/10.21815/JDE.019.011

Kapitanoff, S., & Pandey, C. (2018). Collaborative testing in statistics: Group interaction, anxiety, and class performance. *Statistics Education Research Journal*, *17*(2), 51–67. https://doi.org/10.52041/serj.v17i2.158

Keselyak, N.T., Saylor, C.D., Simmer-Beck, M., & Krust Bray, 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

Khalis, A., Mikami, A.Y., & Hudec, K.L. (2018). Positive peer relationships facilitate adjustment in the transition to university for emerging adults with ADHD symptoms. *Emerging Adulthood 6*(4), 243–254. https://doi.org/10.1177/2167696817722471

Khong, M.L., & Tanner, J.A. (2021). A collaborative two-stage examination in biomedical sciences: Positive impact on feedback and peer collaboration. *Biochemistry and Molecular Biology Education 49*(1), 69–79. https://doi.org/10.1002/bmb.21392

Kloepper, K.D. (2015). Stoplight quizzes: A multilevel assessment strategy for lecture and laboratory courses. *Journal of Chemical Education 92*(3), 509–511. https://doi.org/10.1021/ed500512u

Kruger, S.J. (2019). Improving student assessment feedback in an introductory accounting course through two-stage cooperative assessment. *South African Journal of Higher Education 33*(6), 172–190. https://doi.org/10.20853/33-6-2939

LaBelle, S., & Johnson, Z.D. (2020). The relationship of student-to-student confirmation and student engagement. *Communication Research Reports 37*(5), 234–242. https://doi.org/10.1080/08824096.2020.1823826

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

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

Meseke, J.K., Nafziger, R., & Meseke, C.A. (2008). Facilitating the learning process: A pilot study of collaborative testing vs individualistic testing in the chiropractic college setting. *Journal of Manipulative and Physiological Therapeutics 31*(4), 308–312. https://doi.org/10.1016/j.jmpt.2008.03.007

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

Pittman, L.D., & Richmond, A. (2008). University belonging, friendship quality, and psychological adjustment during the transition to college. *The Journal of Experimental Education 76*(4), 343–362. https://doi.org/10.3200/JEXE.76.4.343-362

Rasco, D., Day, S.L., & Denton, K.J. (2020). Student retention: Fostering peer relationships through a brief experimental intervention. *Journal of College Student Retention: Research, Theory & Practice*. https://doi.org/10.1177/1521025120972962

Rayle, A.D., & Chung, K-Y. (2007). Revisiting first-year college students' mattering: Social support, academic stress, and the mattering experience. *Journal of College Student Retention: Research, Theory & Practice 9*(1), 21–37. https://doi.org/10.2190/X126-5606-4G36-8132

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*, 2527–2535. https://doi.org/10.1021/acs.jchemed.1c00219

Rieger, G., & Heiner, C. (2014). Examinations that support collaborative learning: The students' perspective. *Journal of College Science Teaching 43*(4). https://doi.org/10.2505/4/jcst14_043_04_41

Sato, M. (2013). Beliefs about peer interaction and peer corrective feedback: Efficacy of classroom intervention. *The Modern Language Journal 97*(3), 611–633. https://doi.org/10.1111/j.1540-4781.2013.12035.x

Schmulian, A., & Coetzee, S.A. (2019). Students' experience of team assessment with immediate feedback in a large accounting class. *Assessment & Evaluation in Higher Education 44*(4), 516–532. https://doi.org/10.1080/02602938.2018.1522295

Shell, M.D., & Absher, T.N. (2019). Effects of shyness and friendship on socioemotional adjustment during the college transition. *Personal Relationships 26*(3), 386–405. https://doi.org/10.1111/pere.12285

Slaten, C.D., Ferguson, J.K., Allen, K-A., Brodrick, D-V., & Waters, L. (2016) School belonging: A review of the history, current trends, and future directions. *The Educational and Developmental Psychologist 33*(1), 1–15. https://doi.org/10.1017/edp.2016.6

Slusser, S.R., & Erickson, R.J. (2006). Group quizzes: An extension of the collaborative learning process. *Teaching Sociology 34*(3), 249–262. https://doi.org/10.1177/0092055X0603400304

Smith, M.K., Wood, W.B., Adams, W.K., Wieman, C., Knight, J.K., & Su, T.T. (2009) Why peer discussion improves student performance on in-class concept questions. *Science* 323(5910). Washington: Amer Assoc Advancement Science: 122–124. https://doi.org/10.1126/science.1165919

Sollitto, M., Johnson, Z.D., & Myers, S.A. (2013). Students' perceptions of college classroom connectedness, assimilation, and peer relationships. *Communication Education 62*(3), 318–331. https://doi.org/10.1080/03634523.2013.788726

Summers, J.J., Bergin, D.A., & Cole, J.S. (2009). Examining the relationships among collaborative learning, autonomy support, and student incivility in undergraduate classrooms. *Learning and Individual Differences 19*(2), 293–298. https://doi.org/10.1016/j.lindif.2008.09.006

Swenson, L.M., Nordstrom, A., & Hiester, M. (2008). The role of peer relationships in adjustment to college. *Journal of College Student Development 49*(6), 551–567. https://doi.org/10.1353/csd.0.0038

Urquhart, C. (2013). *Grounded theory for qualitative research: A practical guide*. SAGE Publications. https://doi.org/10.4135/9781526402196

Vaccaro, A., & Newman, B.M. (2017). A sense of belonging through the eyes of first-year LGBPQ students. *Journal of Student Affairs Research and Practice 54*(2), 137–149. https://doi.org/10.1080/19496591.2016.1211533

van der Zanden, P.J.A.C., Denessen, E., Cillessen, A.H.N., & Meijer, P.C. (2018). Domains and predictors of first-year student success: A systematic review. *Educational Research Review 23*, 57–77. https://doi.org/10.1016/j.edurev.2018.01.001

Vazquez-Garcia, 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

Vogler, J.S., & Robinson, D.H. (2016). Team-based testing improves individual learning. *Journal of Experimental Education 84*(4), 787–803. https://doi.org/10.1080/00220973.2015.1134420

Zipp, J.F. (2007). Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology 35*(1), 62–76. https://doi.org/10.1177/0092055X0703500105