Research Article

JMBE Journal of Microbiology & Biology Education

Collaborative Two-Stage Exams Benefit Students in a Biology Laboratory Course

Clara L. Meaders[a] and Yalila Vega[b]

[a]Department of Cell and Developmental Biology, School of Biological Sciences, University of California, San Diego, La Jolla, California, USA

[b]School of Biological Sciences, University of California, San Diego, La Jolla, California, USA

Collaborative two-stage exams provide an effective mechanism to incorporate group work into summative course assessments. We implemented these exams in an upper-level biology laboratory course over two terms, one with online exams and one with in-person exams. We compared student exam performance and perceptions of two-stage exams and group work across terms and demographic groups. Quantitative analyses revealed that across three exams per term, students in groups outperformed students who took the exams individually, and on average the group exam benefited all students, in particular students from groups recognized as persons historically excluded from science because of their ethnicity or race (PEERs). Student responses to both closed and open-ended questions indicated overall positive perceptions of both two-stage exams and group work. We found no significant differences in student perceptions based on PEER student status, gender, or the number of exams helped by group exams, but we found differences related to term and group exam approaches. These findings build upon the literature supporting student learning and perceptions from two-stage exams and provide novel insights for a role of group work in decreasing inequities in course assessments.

KEYWORDS two-stage exam, collaborative exam, group work, collaborative learning

INTRODUCTION

One of the core competencies outlined by AAAS's *Vision and Change* report is the ability for students to communicate and collaborate with other disciplines (1). These skills require team participation and emphasize the collaborative nature of biology. To impart this competency and support increased student learning outcomes, university instructors are increasingly using collaborative learning in the classroom (2). In laboratory courses, this is often demonstrated by students working in pairs or groups on experiments and problem sets. Collaborative exams provide a mechanism to retain this group work on assessments.

Two-stage collaborative exams include an individual as well as a group portion in which students have the opportunity to work together on subsections of the exam. Students receive significantly higher scores on exams when taken as a group compared to when exams are taken individually (3). However, reports are mixed regarding the impacts of collaborative testing on long-term retention of material, with some studies reporting higher retention (4, 5) and others reporting increases in performance only and not retention (6). More recently, Cooke et al. (7) showed that collaborative exams using open-ended questions improved student retention of course content. Regardless, these types of exams have been shown to increase student performance for all students, in particular for lower-performing students (8). Lower-performing students typically benefit from active learning and more highly structured courses (9). Active learning is a recommended strategy for decreasing equity gaps in the classroom, and two-stage collaborative exams may be an assessment tool that aligns with these instructional goals.

Student feedback regarding collaborative exams is generally positive, with the majority of students often reporting that this type of exam is good or helpful for their learning (7, 10), that the exams promote peer collaboration and communication (11), and that the exams reduce test anxiety (12). In addition to supporting learning and collaboration, two-stage exams are a recommended method for promoting academic integrity during online exams, as students may be deterred from cheating if they will be collaborating with peers on a portion of the exam (13).

While the literature supports the effectiveness of in-person collaborative exams for student learning and exam experiences, little is known about the effectiveness of two-stage collaborative

Editor Deborah K. Anderson, St. Norbert College
Address correspondence to Department of Cell and Developmental Biology, School of Biological Sciences, University of California, San Diego, La Jolla, California, USA. E-mail: cmeaders@ucsd.edu.
The authors declare no conflict of interest.
Received: 22 August 2022, Accepted: 9 November 2022, Published: 1 December 2022

exams on reducing equity gaps on exams in the biology classroom or how students experience these exams in online compared to in-person environments. In this study, we set out to explore the following questions: (i) How does student performance on the individual and group components of two-stage exams compare in an upper-level biology laboratory course offered online and in-person? (ii) Are two-stage exams a strategy that can be used to address grade equity gaps, in particular for PEER students, i.e., students historically excluded from science because of their ethnicity or race (14)? (iii) How do biology students perceive two-stage exams in terms of impacting their learning?

METHODS

Study population

We implemented two-stage exams in an upper-level biology laboratory course at a research-intensive university in the western United States that enrolled 49 students in winter 2022 and 49 students in spring 2022. Of these students, 42 students from winter quarter and 39 students from spring quarter consented to participate (overall participation rate, 83%). The population demographics were 79% not underrepresented and 21% underrepresented in science, technology, engineering, and mathematics (STEM); 59% of students identified as women (cis and trans) and 40% identified as men (cis and trans).

Exam structure

The students completed three exams over the 10-week quarter. During the winter quarter the course was online for the first 4 weeks (due to the ongoing 2019 coronavirus disease [COVID-19] pandemic) and in person for the remaining 6 weeks. All three exams took place online. During the spring quarter, the course was entirely in person.

Exams were implemented during laboratory sections, with students having 60 min to complete the individual portion and 30 min to complete the group portion. Students worked in their laboratory groups (4 to 5 people), which were stable during both quarters and consisted of two pairs of students who worked collaboratively on experiments and shared a lab bay. The individual exam comprised 75% of the final exam score, and the group exam comprised 25% of the final score. The group exam typically included three questions, two from the individual exam and one new question. Students were informed that if their final cumulative score was less than the percentage correct on their individual exam, their individual exam would be used for their exam grade. As such, the group exam could only help students, incentivizing group discussion and collaboration.

Student groups were formed by the instructor during the first week of the quarter, after students filled out a brief precourse survey where they provided information regarding their prior experience (e.g., "Have you taken a molecular biology course prior to this lab course?" and "What is your lab work experience?"). Groups were formed such that all groups had a maximum of two students without molecular biology experience, and students with lab experience were distributed evenly across groups.

Ethics statement

This study was approved by IRB protocols 170886 and 804993.

Data collection

We collected three types of data for this study: exam performance, survey data, and student demographic information.

Exam performance. Student performance on the two-stage exams was calculated by scoring each component of the exam (individual, group, and final exam performance) as a percentage of 100. If a student received a higher percentage on the individual exam compared to the combination of their individual and group scores, their individual exam was used as their final exam score.

Survey. To assess student perceptions of two-stage exams and group work, we implemented a survey after the second group exam in week 7 of the 10-week quarter. Survey items included Likert-type items derived from survey methods of Grzimek et al. (15) on attitudes toward group work, specifically student views on impacts on learning and grades and general attitudes toward group work, and also the survey methods of Shaffer (16) on perceptions of two-stage exams and open-ended questions. This survey was open for 1 week, and students received extra credit for their participation. Students were given the option of filling out the survey and declining to submit their responses for research purposes.

Demographic data. Student gender and identification with racial and ethnic groups traditionally underrepresented in STEM were obtained from the Registrar. The race and ethnic groups identified as persons excluded because of their ethnicity or race included African American or Black, Hispanic (Chicano/Latino), and Native American/Alaska Native.

Data analysis

Descriptive statistics were generated in JASP and R. Data visualization was conducted in R using the ggplot2 (17) and Likert (18) packages.

We conducted an exploratory factor analysis (EFA) in JASP. Separate EFAs were conducted for items related to two-stage exam perceptions and general group work perceptions. Factors were extracted using parallel analysis, and an oblique rotation was used to determine the final factor structure. Scree plot analyses were examined to support the final total factor structure. Based on the plots, we considered one factor for two-stage exam perceptions and one factor for general group work perceptions.

We asked students an open-ended question: "My perception of two-stage exams. . ." Seventy three of the 81 students who filled out the survey responded to this question. We conducted a

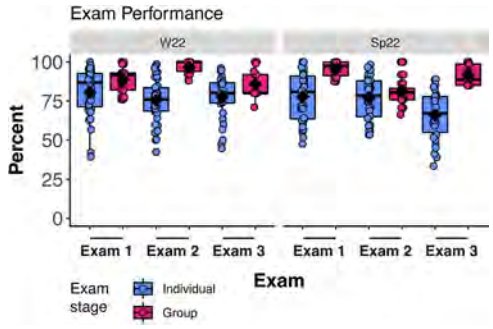

FIG 1. Overall exam performance as shown in boxplots of student performance on the three course exams, normalized out of 100%, disaggregated by quarter (W22 = winter 2022; Sp22 = spring 2022). Boxes represent interquartile ranges, black lines represent median scores, and black diamonds represent mean scores. Each circle represents the exam score for one student.

thematic analysis to identity codes present in the student responses. We iteratively coded a random subsample of 20 student responses using preliminary codes until reaching intercoder reliability (19). The minimum percent agreement for each code was 85%, and the average percent agreement across codes was 92.3%. Cohen's Kappa was 0.80, indicating strong intercoder reliability (20).

RESULTS

Students performed highest on the group exam portion of the two-stage exam

In both online (winter quarter 2022) and in-person (spring quarter 2022) exam settings, students consistently received higher grades on the group portion of the exam compared to the individual exam (Fig. 1, Table 1). An analysis of variance revealed that there were significant differences ($F = 24.17$, $P < 2e{-}16$), and Tukey-adjusted pairwise tests revealed that for all three exams, group exam scores were significantly higher than individual exam scores (Table 2). The group exam helped 47 students on all of their exams, 24 students on two exams, and 8 students on one exam, and 2 students received higher scores on their individual exam than any of their group exams. Consequently, cumulative scores were higher than scores students would have received if they were graded solely on their individual exam performance.

Group exams may decrease equity gaps on exams

Students who were identified as PEERs by the university received lower exam grades on the individual exams for the first two exams, with students from non-PEER groups receiving scores on average 14.7 percentage points higher on exam 1, 11.2 percentage points higher on exam 2, and 5.2 points higher on exam 3 (Table 3). After applying a Bonferroni correction, this difference was only significant for the second exam. There were no significant differences for students from these groups in grades on the group exam, indicating that students across all groups benefited from the exam. The combination of individual and group exam scores slightly decreased the equity gaps in final exam scores. However, final student course grades indicated that there were remaining equity gaps in the course, with PEER students receiving an average final grade of 90.28 ± 6 (mean ± standard deviation [SD]) and students who were non-PEERS receiving an average final grade of 93.7 ± 5.4. Welch's *t* test revealed that these differences were significant ($P = 0.04$), with a moderate effect size (Cohen's $D = 0.6$).

Students overall perceived two-stage exams as helpful and enjoyable

We asked students five Likert-type questions regarding their perceptions of two-stage exams. Across both quarters, students agreed or strongly agreed that the group exam helped them understand the material more clearly than if they had not had the group portion of the exam, with 100% of students agreeing with

TABLE 1
Descriptive statistics for exam performance

	Winter 2022 (*N* = 42)			Spring 2022 (*N* = 39)	
Exam	**Exam type**	**Mean %**	**SD**	**Mean %**	**SD**
Exam 1[a]	Individual	80.7	14.8	77.8	15.4
	Group	88.8	8.5	95.9	4.4
	Final	83.3	12.5	82.4	11.7
Exam 2[b]	Individual	75.4	14.0	76.4	14.2
	Group	96.6	4.6	81.0	8.7
	Final	80.7	10.8	78.5	12.5
Exam 3[c]	Individual	77.7	14.0	64.3	17.4
	Group	85.9	8.2	93.3	19.1
	Final	80.1	11.9	71.7	15.8

[a]One person did not take exam 1 in winter 2022, and two people did not take exam 1 in spring 2022.
[b]Two people did not take exam 2 in spring 2022.
[c]Ten people did not take exam 3 in winter 2022, and nine people did not take exam 3 in spring 2022.

TABLE 2
Pairwise comparisons of exam performance

Comparison	Tukey-adjusted *P* value
Exam 1: individual vs group	0.0000000
Exam 2: individual vs group	0.0000000
Exam 3: individual vs group	0.0000000

this statement in the online winter quarter exams and 82% of students agreeing with this statement during the in-person spring quarter exams (Fig. 2). Overall, 90% of students reported enjoying the group portions of exams. The majority of students agreed that this type of assessment should be used in other biology courses and felt that their group members contributed equally to the exam (Fig. 2).

Exploratory factor analysis revealed that four out of five items loaded onto one factor for two-stage exam perceptions (Table 4). One item, "students in my group unfairly benefited from the group part of the exam," was reverse coded prior to analysis. This item did not load onto the factor for two-stage exam perceptions and was removed from subsequent analyses. Each of the items that loaded onto the factor were related to positive perceptions of two-stage exams. This factor explained 49.6% of variation in student responses and had high internal consistency reliability (Cronbach's alpha, >0.8).

Students overall agreed or strongly agreed that group work was beneficial for them (Fig. 3). Exploratory factor analysis revealed that all 11 items loaded onto one factor for attitudes toward group work (Table 5). This factor explained 51.5% of variation in student responses and had high internal consistency reliability (Cronbach's alpha, >0.8).

We generated a summary score for each student for their perceptions toward two-stage exams and group work (Table 6). Overall, students felt positively regarding each factor. There were no significant differences in average student scores for either factor disaggregated by students' PEER status or gender (see Appendix SA in the supplemental material).

Most students worked collaboratively on the group portion of the exam

Groups may vary in how they approach group exams, with some approaches fostering more discussion than others. We asked students to select which of four options best fit their

TABLE 3
Exam performance disaggregated by PEER student status[a]

Comparison	PEER category	No. of students		Mean score (%)	SD	*P* value
		Valid	Missing			
Exam 1						
Individuals	Not PEERs	63	1	82.3	12.4	0.009
	PEERs	16	1	67.6	19.2	
Groups	Not PEERs	63	1	92.7	7.5	NS
	PEERs	16	1	90.4	8.1	
Final	Not PEERs	63	1	85.2	10.0	0.008
	PEERs	16	1	73.4	15.0	
Exam 2						
Individuals	Not PEERs	62	2	78.5	13.1	0.005*
	PEERs	17	0	67.3	13.2	
Groups	Not PEERs	62	2	89.1	10.5	NS
	PEERs	17	0	90.4	9.9	
Final	Not PEERs	62	2	81.7	11.0	0.008
	PEERs	17	0	73.2	11.0	
Exam 3						
Individuals	Not PEERs	46	18	73.6	15.7	NS
	PEERs	15	2	68.4	16.1	
Groups	Not PEERs	46	18	91.9	9.7	NS
	PEERs	14	3	88.9	9.7	
Final	Not PEERs	46	18	78.0	12.5	NS
	PEERs	15	2	73.3	12.6	

[a]Valid represents the number of students from each category. Students could drop one exam, and the number of missing students represents the number of students who did not take the indicated exam. *P* values were calculated by conducting Welch's *t* test between non-PEER and PEER students; the asterisk indicates $P < 0.05$.

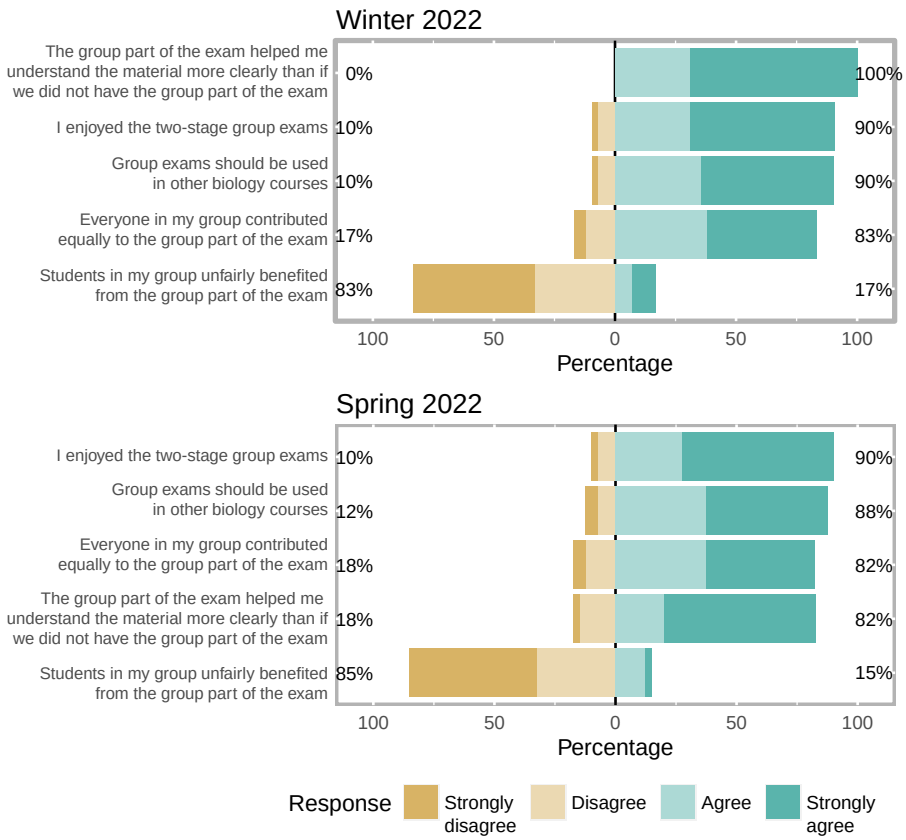

FIG 2. Student perceptions of two-stage exams, as shown by the percentages of students who reported that they strongly agreed, agreed, disagreed, or strongly disagreed with items assessing their perceptions of two-stage exams. Items are ordered from top to bottom by the percentage of students reporting that they either strongly agreed or agreed.

group's approach. The majority of students ($N = 68$) reported working collaboratively, "discussing each question until all members agreed on an answer or an explanation." Six students reported they "took a vote and if unanimous moved on, otherwise discussed the questions until all members agreed on an answer"; three students "took a vote and used the majority to determine the answer"; and four students "used the answers from the one person in the group who knew the most biology."

Qualitative feedback identified aspects of the exam that students perceived positively

When asked about their perceptions of two-stage exams, 60 students reported statements that included only positive perceptions, 4 students reported negative perceptions, and 6 students reported a mix of both. We found no significant differences in student perceptions across gender, PEER student status,

TABLE 4
Factor loadings for perceptions of two-stage exams, with summary statistics

Items and summary statistics	Factor loadings	Mean	SD	Result
Items				
Group exams should be used in other biology courses	0.865	3.383	0.784	
I enjoyed the two-stage group exams	0.827	3.494	0.744	
The group part of the exam helped me understand the material more clearly than if we did not understand the group part of the exam	0.731	3.580	0.668	
Everyone in my group contributed equally to the group part of the exam	9.646	3.247	0.845	
Summary statistic				
Sums of squared loadings				2.479
Proportion of variance				49.6%
Construct reliability (alpha)				0.84

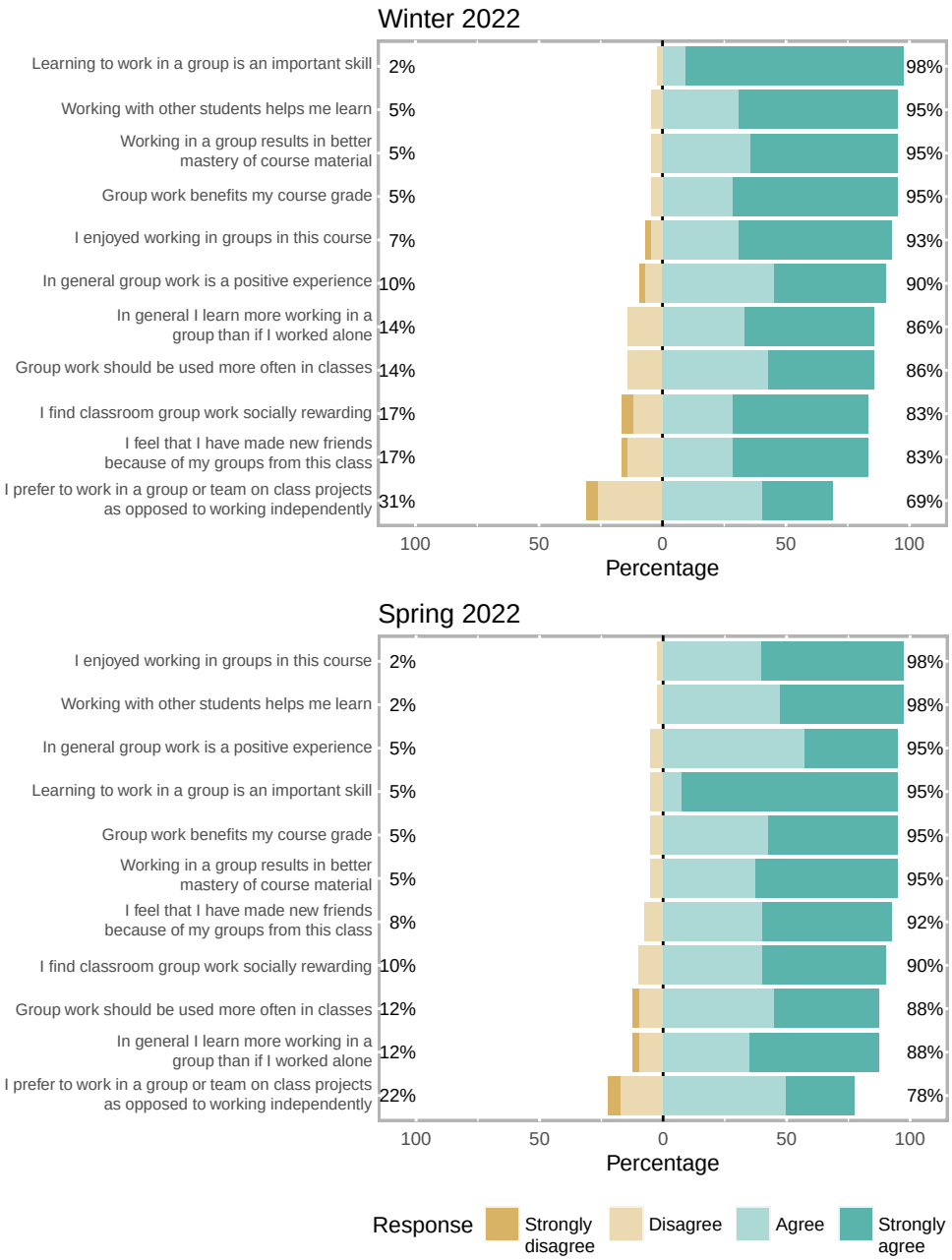

FIG 3. Student perceptions of group work, as shown by the percentage of students who reported that they strongly agreed, agreed, disagreed, or strongly disagreed with items assessing their perceptions of group work. Items are ordered from top to bottom by the percentage of students reporting that they either strongly agreed or agreed.

or by group exam grade gains (see Appendix SB). However, students during the in-person quarter were more likely to report mixed perceptions [X^2 (2, $N = 70$) = 7.9, $P = 0.02$]. Additionally, there was a significant relationship between group approaches and student perceptions [X^2 (6, $N = 70$) = 16.67, $P = 0.001$].

One student who had a positive experience responded "*My perception of the two-stage exams is that it really helps. During the group portion, when going over questions, I get insight on how they answered the questions themselves. It also helps me see how they approach and answer the questions. Lastly, I feel that I can reflect on what I did wrong which helps me to learn better.*" This statement represented the majority of student perceptions, with 49 students making statements about the utility of the exam structure and 53 students making statements about the impacts on their learning. Students frequently remarked about the ability to learn from their mistakes ($N = 24$), to see how others approached problems ($N = 28$), and about the increased understanding due to discussions ($N = 21$) (Table 7).

One student remarked on the importance of supporting positive group dynamics: "*Multi-stage exams can be great when teams work together well. However, if a team does not work together well, it can be more harmful than beneficial.*" Establishing trust

TABLE 5
Factor loadings for perceptions of group work, with summary statistics

Items and summary statistics	Factor loading	Mean	SD	Result
Items				
I enjoyed working in groups in this course	0.814	3.543	0.633	
In general I learn more working in a group than if I worked alone	0.81	3.383	0.751	
Working in a group results in better mastery of course material	0.785	3.531	0.593	
Working with other students helps me learn	0.779	3.543	0.571	
Group work should be used more often in classes	0.771	3.284	0.729	
I find classroom group work socially rewarding	0.72	3.370	0.782	
In general group work is a positive experience	0.713	3.333	0.652	
I prefer to work in a group or team on class projects as opposed to working independently	0.712	2.963	0.843	
Group work benefits my course grade	0.689	3.556	0.592	
I feel that I have made new friends because of my groups from this class	0.545	3.420	0.722	
Learning to work in a group is an important skill	0.476	3.840	0.46	
Summary statistic				
Sums of squared loadings				5.669
Proportion of variance				51.5%
Construct reliability (alpha)				0.916

within a group can help students focus on the discussions, as another student shared that "*I enjoyed the two-stage exams since after working with my group since the beginning of the year, I was able to trust their knowledge and opinions when going over questions*."

A subset of students ($N = 6$) shared a mix of positive and negative perceptions. These students ranged in gender and PEER identities, but all had benefited from group exams, were from the in-person term, and had reported that their groups engaged in discussion. One student responded "*It sometimes made me feel incompetent if other group members knew how to tackle a problem and I did not, but was useful in understanding how to approach those problems*." This type of sentiment was common for students with mixed perceptions, with another student sharing, "*I liked getting new perspectives from my group members as to how they answered a question. However, this also gave me a bit of extra test anxiety when I realized that they did a problem correctly and I probably got it wrong*." These students often shared concrete examples of benefits for their learning but reported feelings of stress, anxiety, or discomfort upon realizing they had made errors in the individual portion of the exam.

TABLE 6
Summary scores[a]

Factor	Summary score	
	Mean	SD
Two-stage exams	3.426	0.629
Attitudes towards group work	3.433	0.497

[a]Summary scores were calculated by taking the average score for an item (on a scale from 1 to 4) for each factor.

Four students, ranging in identities, shared only negative perceptions. Two of the four students reported that their groups did not engage in discussion, with one student sharing "*I feel as though the group portion of the exam is stressful, socially taxing, and I am more likely to relent to the more dominate voice than fight for my own opinion and waste time. I think that group exam work should be given MORE time than the individual portion because of the aspect of discussion that occurs in a group that DOES NOT occur when I am taking an exam by myself*." This response emphasizes the importance of aligning the time expectations for an exam with the overall goal of promoting consensus-building discussions. Finally, negative perceptions may be mitigated by exam policies. A student shared that "*It was sometimes frustrating to know that one of the answers could be wrong because someone else decided it was the best answer, but that was mitigated by the fact that you couldn't get a lower score on the exam total compared to your individual scores*."

DISCUSSION

In this study, we found that students taking either online or in-person group exams in a biology course performed significantly higher on the group portion of collaborative two-stage exams (Fig. 1, Tables 1 and 2). These results are consistent with previous literature detailing positive achievement gains for students in biology (21) and other disciplines (16, 22–25) and invite future directions for study.

Improvement for high-achieving and low-achieving students has been well-documented, but to our knowledge this is the first study to explore the potential of two-stage exams to address

TABLE 7
Student perceptions of two-stage exams[a]

Overall category	Detailed code	Description of code	No. of times mentioned
Positive experience (total statements: 161)	Useful or helpful experience	Statements that the structure had utility	49
	Impactful for learning	Able to see or learn from mistakes; identified areas to work on	24
		Able to see how others approached problems	28
		Able to understand concepts	21
	Better for grade	Students expressing impacts on exam scores	8
	Reduce stress	Relieving test anxiety	8
	Increase community	Increasing sense of belonging and/or group camaraderie	3
	Affect: like, love, or appreciate it	Explicitly positive statements	20
Negative experience (total statements: 10)	Tiring	Feeling tired after multiple exams	2
	Rushed	Feeling rushed for time	3
	Increase stress	Feeling increased stress	2
	Other negative	Comparing to others, size of the groups, exam logistics (not enough copies of the group exam)	3
Other		Statements that did not fit in the other codes	4

[a]Seventy-three students provided responses regarding their perceptions. Codes were categorized as positive or negative experiences.

equity gaps that exist for students along racial and ethnic axes. Equity gaps in course performance exist in many courses, with socioeconomic status and minority status associated with course achievement (26). We did not have a control term without group exams, but the equity gaps identified during the individual exams indicated that these gaps likely were present in prior terms. In our course, equity gaps existed for PEER students in STEM, with non-PEER students receiving scores on average 14.7% higher on exam 1, 11.2% higher on exam 3, and 5.2% higher on exam 3 (Table 3). The group exam helped all students (Table 3) with their exam performance, indicating that the collaborative learning had benefits for students regardless of background. Previous studies have explored the impacts of explaining reasoning on helping both higher- and lower- achieving students (25). Exams are increasingly being viewed as learning tools, an acknowledgment that during a term students are still engaged with the active process of learning (27). Two-stage exams provide a graded incentive for students to discuss and engage in the learning process together. Further, with the grading structure (75% individual, 25% group), this resulted in benefits in exam scores for all students and a decrease in (but not elimination of) equity gaps for students' final scores. Future work should explore the causes of these remaining equity gaps for exams 1 and 2. Notably, there were no significant differences in exam scores for students during either the individual or group components of exam 3. This was likely due to another aspect of the exam structure, in which students could drop their lowest exam score; high-achieving students who were satisfied with their performance on the first two exams could opt out of the third exam.

Future studies should further explore the types of group dynamics that are most productive for learning during group exams. The literature is mixed regarding whether homogenous or heterogenous groups of higher- and lower-achieving students result in larger benefits for learning (28, 29). Regardless, when students have a friend in their group, this is associated with higher student comfort (30), and when students perceive personal connections and active contributions from their group members this contributes to willingness to work together (31). We did not identify significant differences in perceptions based on student demographics or group exam grade gains, but student interviews could explore how group approaches and other factors, such as student extroversion or introversion, may impact student experiences.

Question difficulty may impact student experiences. During the group exams, students received the most challenging open-ended questions from the individual exam, i.e., questions that were from higher Bloom's taxonomy levels (e.g., analyze and apply, evaluate, create) and would benefit from group discussion and consensus building. We theorized that the group exam consequently was a higher-difficulty exam than the individual exam, but we did not we did not track variability among items or across exams. Recently, Martin (32) provided a quantitative framework for researchers interested in exploring the interactions between individual knowledge, group dynamics, and question difficulty. Applying this framework to future analyses would allow us to explore variability among questions.

In our course, students participating in two-stage exams either online or in-person overwhelmingly felt positively about their experiences (Fig. 2), but some in-person students reported mixed perceptions. With the small sample size, these results should be interpreted with caution but may indicate that in-person group exams can elicit stressors such as within- or between-group comparisons that are less prevalent during online exams. Regardless of environment, we anticipate that so long as groups

have established trust, the mode of exam delivery is less relevant and students can experience the benefits in either mode. Overall, the most common positive aspects of two-stage exams perceived by students focused on the impacts for their learning, including the opportunity to immediately reflect on their work, learn how others approached problems, and increase their understanding of the material by engaging in discussion (Table 7). A study tracking students' internal feedback during two-stage exams found that students engaged in self-regulatory feedback processes after engaging in the group portion of the exam, such as clarifying each answer, exploring different perspectives, and planning their responses (33). A benefit of the two-stage exam structure is that individual knowledge at a point in time is assessed but that the group exam provides a mechanism for the exam to promote student learning and reflection as well. We found that students more commonly commented about their learning than grades (Table 7), but future studies could use our codes to inform closed-ended surveys assessing which aspects of two-stage exams students value the most.

Students' critiques of the two-stage exam, while less common, provided valuable insights into areas to improve the structure. While eight students cited the two-stage exams as reducing stress, two students reported that they experienced increased stress during the discussions when they realized their responses on the individual exam were likely incorrect. Instructors may be able to mediate this stress by promoting growth mindsets among their students (34) and providing multiple reminders to students that while the individual exam is assessing their understanding at a particular moment, the group exam rewards their learning. Additional student critiques focused on the logistics of the exam, in particular, the importance of additional time for discussion. One student asked for more copies of the exam to be provided during the in-person group so that students did not have to crowd around one another to review the questions. Instructors implementing two-stage exams should take these points under consideration for easing student experiences and helping them maximize the group collaborations. Providing a longer time window could potentially alleviate any concerns from students, such as the one from our study who admitted to succumbing to a dominant voice due to feeling rushed during the group exam.

Personally, we enjoyed witnessing the dynamic discussions that students engaged in during the group exam. In our course, the groups of ~4 to 5 people resulted in 10 to 12 additional exams. We used questions from the individual exam in the group exam (with one new question), reducing instructor workload with question design. Overall, our study builds upon the existing literature of two-stage exams and suggests that both online and in-person exams have benefits for student exam performance and student perceptions of the exams.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENT

We thank the UCSD Center of Advancing Multidisciplinary Scholarship for Excellence in Education for their help with data deidentification and collection.

REFERENCES

1. AAAS. 2011. Vision and change in undergraduate biology education: a call to action. American Association for the Advancement of Science, Washington, DC.
2. National Research Council. 2012. Understanding and improving learning in undergraduate science and engineering discipline-based education research. National Academies Press, Washington, DC.
3. Rao SP, Collins HL, DiCarlo SE. 2002. Collaborative testing enhances student learning. Adv Physiol Educ 26:37–41. https://doi.org/10.1152/advan.00032.2001.
4. Cortright RN, Collins HL, Rodenbaugh DW, DiCarlo SE. 2003. Student retention of course content is improved by collaborative-group testing. Adv Physiol Educ 27:102–108. https://doi.org/10.1152/advan.00041.2002.
5. Eaton TT. 2009. Engaging students and evaluating learning progress using collaborative exams in introductory courses. J Geosci Educ 57:113–120. https://doi.org/10.5408/1.3544241.
6. 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 Sci Educ 11:392–401. https://doi.org/10.1187/cbe.12-04-0048.
7. Cooke JE, Weir L, Clarkston B. 2019. Retention following two-stage collaborative exams depends on timing and student performance. CBE Life Sci Educ 18:ar12. https://doi.org/10.1187/cbe.17-07-0137.
8. Giuliodori MJ, Lujan HL, Dicarlo SE. 2008. Collaborative group testing benefits high-and low-performing students. Adv Physiol Educ 32:274–278. https://doi.org/10.1152/advan.00101.2007.
9. Haak DC, HilleRisLambers J, Pitre E, Freeman S. 2011. Increased structure and active learning reduce the achievement gap in introductory biology. Science 332:1213–1216. https://doi.org/10.1126/science.1204820.
10. Rieger GW, Heiner CE. 2014. Examinations that support collaborative learning: the students' perspective. J Coll Sci Teach 43:41–47.
11. Khong ML, Tanner JA. 2021. A collaborative two-stage examination in biomedical sciences: positive impact on feedback and peer collaboration. Biochem Mol Biol Educ 49:69–79. https://doi.org/10.1002/bmb.21392.
12. Rempel BP, Dirks MB, McGinitie EG. 2021. Two-stage testing reduces student-perceived exam anxiety in introductory chemistry. J Chem Educ 98:2527–2535. https://doi.org/10.1021/acs.jchemed.1c00219.
13. Hsu JL, Goldsmith GR. 2021. Instructor strategies to alleviate stress and anxiety among college and university STEM students. CBE Life Sci Educ 20:es1. https://doi.org/10.1187/cbe.20-08-0189.
14. Asai DJ. 2020. Race matters. Cell 181:754–757. https://doi.org/10.1016/j.cell.2020.03.044.
15. Grzimek V, Kinnamon E, Marks MB. 2020. Attitudes about classroom group work: how are they impacted by students' past experiences

and major? J Educ Bus 95:439–450. https://doi.org/10.1080/08832323.2019.1699770.
16. Shaffer JF. 2020. Student performance on and perceptions of collaborative two-stage exams in a material and energy balances course. Chem Eng Educ 54:52–58.
17. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
18. Bryer J, Speerschneider K. 2016. Analysis and visualization Likert items. R package, version 1.3.5. https://github.com/jbryer/likert.
19. O'Connor C, Joffe H. 2020. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 19:1–13. https://doi.org/10.1177/1609406919899220.
20. Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174. https://doi.org/10.2307/2529310.
21. Newton G, Ritchie K, Albabish W, Gilley B, Rajakaruna R, Kulak V. 2019. Research and teaching: two-stage (collaborative) testing in science teaching: does it improve grades on short-answer questions and retention of material? J Coll Sci Teach 48:64–73.
22. Knierim K, Turner H, Davis RK. 2015. Two-stage exams improve student learning in an introductory geology course: logistics, attendance, and grades. J Geosci Educ 63:157–164. https://doi.org/10.5408/14-051.1.
23. Levy D, Svoronos T, Klinger M. 2018. Two-stage examinations: can examinations be more formative experiences? Act Learn High Educ https://doi.org/10.1177/1469787418801668.
24. Bruno BC, Engels J, Ito G, Gillis-Davis J, 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 https://doi.org/10.5670/oceanog.2017.241.
25. Zipp JF. 2007. Learning by exams: the impact of two-stage cooperative tests. Teach Sociol 35:62–76. https://doi.org/10.1177/0092055X0703500105.
26. Sirin SR. 2005. Socioeconomic status and academic achievement: a meta-analytic review of research. Rev Educ Res 75:417–453. https://doi.org/10.3102/00346543075003417.
27. Efu SI. 2019. Exams as learning tools: a comparison of traditional and collaborative assessment in higher education. Coll Teach 67:73–83. https://doi.org/10.1080/87567555.2018.1531282.
28. Donovan DA, Connell GL, Grunspan DZ. 2018. Student learning outcomes and attitudes using three methods of group formation in a nonmajors biology class. CBE Life Sci Educ 17:ar60. https://doi.org/10.1187/cbe.17-12-0283.
29. Briggs M. 2020. Comparing academically homogeneous and heterogeneous groups in an active learning physics class. J Coll Sci Teach 49:76–82.
30. Theobald EJ, Eddy SL, Grunspan DZ, Wiggins BL, Crowe AJ. 2017. Student perception of group dynamics predicts individual performance: comfort and equity matter. PLoS One 12:e0181336-17. https://doi.org/10.1371/journal.pone.0181336.
31. Premo J, Wyatt BN, Horn M, Wilson-Ashworth H. 2022. Which group dynamics matter: social predictors of student achievement in team-based undergraduate science classrooms. CBE Life Sci Educ 67:ar51. https://doi.org/10.1187/cbe.21-06-0164.
32. Martin AP. 2018. A quantitative framework for the analysis of two-stage exams. Int J Higher Educ 7:33–54. https://doi.org/10.5430/ijhe.v7n4p33.
33. Nicol D, Selvaretnam G. 2022. Making internal feedback explicit: harnessing the comparisons students make during two-stage exams. Assess Eval High Educ 47:507–522. https://doi.org/10.1080/02602938.2021.1934653.
34. Dweck CS. 2006. Mindset: the new psychology of success. Random House, New York, NY.