STUDENT PERFORMANCE ON AND PERCEPTIONS OF COLLABORATIVE TWO-STAGE EXAMS IN A MATERIAL AND ENERGY BALANCES COURSE

Justin F. Shaffer
Colorado School of Mines • Golden, CO 80401

INTRODUCTION

Active learning is a style of teaching in which students actively participate and engage in the learning process during class through a variety of activities including group work, problem solving exercises, the use of student response systems, and working through case studies, among others.[1] Evidence shows that the use of active learning yields benefits for student learning that surpass traditional lecture in a variety of science, technology, engineering, and math (STEM) courses.[1,2] Due to this, active learning has been widely promoted in college and university STEM courses, and there have been calls to action to promote and use these pedagogical methods.[3,4] The benefits of active learning have also been documented for chemical engineering courses, including the material and energy balances (MEB) course.[5-7] Active learning is often used in a class session by having students work in groups to complete no- or low-stakes formative assessment exercises, such as answering "clicker" questions in class, solving open ended problems, and generating concept maps, among others.[8] However, high-stakes summative assessment items, such as exams (which could be worth up to 90% of a student's grade or more), are usually taken individually and not in groups. While collaboration, group work, and peer instruction have been shown to be strongly linked to learning,[9,10] these practices are not typically used during individual exams. One of the goals, therefore, of active learning would be for students to develop strong conceptual and technical skills working in their groups, which then could be applied to individual assignments such as exams.

To provide a more direct connection between the benefits of collaborative active learning exercises and high-stakes exams, two-stage collaborative exams have been developed and assessed in college STEM courses. In a two-stage collaborative exam, "students are given an opportunity to improve their understanding of a topic by first taking a test alone and then taking the test, or a portion of the test, again while interacting with a peer group."[11] Logistically, a two-stage exam involves allotting time for students to take the exam on their own, which they complete and then turn in to the instructor. Next, students form groups (usually of three or four students), and the entire exam (or a portion of it, or a modified portion of it using isomorphic questions, for example) is passed out to the groups. Since the students have already seen the exam during the individual phase, less time is allotted for the group component. Students then complete the group component of the exam and turn in a single copy. Grading of the individual and group components of the two stage-exam can vary, usually with 25% of the grade for the group component being an upper bound.[12,13]

Two-stage exams build on the situated framework of learning wherein students work within a group or community to acquire knowledge and learn.[14] They have been assessed in a variety of college STEM courses, with most studies finding them to be strongly favored by students and to provide immediate and potentially long-term benefits to student performance.[15] For example, an analysis of multiple geology and oceanography courses found that student performance on the group component of two-stage exams was significantly higher than individual performance, and that the majority of group scores were higher than any individual in the group,

Justin F. Shaffer *is a Teaching Associate Professor in Chemical and Biological Engineering at the Colorado School of Mines. He teaches material and energy balances, introductory thermodynamics, introductory biology, and anatomy and physiology. His research focuses on the efficacy of components of high structure courses and engineering students' attitudes towards biology.*

suggesting that students worked together to improve their answers during the group component.[16] Similar results were found in several types of STEM courses including mechanical engineering design,[17] introductory physics (mechanics),[18] computer science,[19] introductory geology,[20,21] introductory biology,[22] biochemistry,[23] earth and ocean science,[24] physiology,[23,25-27] and introductory sociology.[28] While students benefit in the moment by taking the group component of the exam immediately after the individual component, studies have been conducted to determine the impact of two-stage exams on longer term retention of knowledge. Many studies have found positive effects of two-stage exams on longer-term retention;[19,23,26,29,30] however, others have found mixed or no impacts.[11,22,31,32] In addition to performance benefits, students demonstrate strong positive attitudes towards the use of two-stage exams.[12,13,17] Based on the literature, it is clear that students across a wide range of STEM disciplines gain immediate benefits from two-stage exams and are in favor of these assessments.

While two-stage exams have been widely and successfully used in other STEM disciplines, it is unknown what impact two-stage exams would have on the learning and attitudes of chemical engineering students. As active learning has been shown to promote student learning in chemical engineering courses,[5-7] it is likely that two-stage exams would also yield positive benefits. The goal of this study was to describe and evaluate the use of two-stage exams in an MEB course.
The research questions in this study were:

1. How does student performance on the individual and group components of two-stage exams compare in an MEB course?

2. How do chemical engineering students perceive two-stage exams in terms of impacting their learning?

METHODS

Study Population

Undergraduate chemical engineering students (n = 34) enrolled in one section of a material and energy balances course (MEB) in Summer 2019 were invited to participate in this study. Of these students, 30 (88.2%) consented to being in this study. The population demographics were 63.3% female and 36.7% male, 46.7% white, 20.0% Asian, 13.3% Hispanic, 10.0% international, 6.7% multiple races, and 3.3% African American. This study was determined to be exempt by the Colorado School of Mines Human Subjects Research Committee.

This summer MEB course was taught in an accelerated six-week timeframe as compared to the standard 16-week semester. The course met for three two-hour and 50-minute sessions a week. Topics included process variables, material balances for systems with and without reactions, single-phase systems, multiple-phase systems, first law energy balances for systems with and without reactions, and transient systems. Student performance was evaluated primarily by weekly homework sets (submitted by groups of students), weekly in-class written two-stage exams (see more information below), and a final exam (taken individually). The course was taught by the author primarily with active learning and a moderate structure format,[33,34] which means the course included one or more graded review assignments per week in addition to using active learning in class. Specifically, the course included optional pre-class reading guides,[35] the use of iClickers (Macmillan Learning, New York, NY) for formative in-class assessments, and graded weekly homework assignments.

Two-stage Exam Format

Students completed weekly two-stage exams at the start of class each Wednesday that assessed the topics from the prior calendar week of class (Monday through Friday). The entire exam session took about 60 minutes and was conducted as follows. Students first took the exam individually for 30 minutes. Each exam consisted of two open-ended questions that covered material from the prior week of class. Each question was intended to take 15 minutes to complete and required one or more numerical answers. Students were allowed to use a calculator and were provided with the necessary physical property data. After 30 minutes, students submitted their individual exams and then rearranged their desks into groups of three to five students (more information on how groups were formed can be found below). Students were given approximately two minutes to discuss the exam and then a single copy of the identical exam was distributed to each group. Students then had 20 minutes to complete the group exam. After 20 minutes had passed, students then turned in the single group exam, and then the exam solution was projected. Approximately five to seven minutes were spent going over the exam solutions, and students were able to ask clarifying questions during this time. The exams were graded with identical rubrics, and overall exam scores were calculated with 80% weight for the individual component and 20% weight for the group component.

Student groups were formed by the instructor as follows. Students completed a brief survey during the first week of class where they provided information on what days they preferred to complete homework and study and any other considerations they wanted to share about working in groups. Gender and GPA data were obtained from the Registrar's office. Custom software in MATLAB was used to generate groups of three or four students based on 1) study/homework day preference, 2) gender, and 3) GPA so as to provide equitable and fair group assignments.[36] The algorithm formed groups based on students' study day preferences (all with the same day preference in a group), ensured gender balance (no groups with all males or all females), and ensured GPA balance (each group had students with a mixture of high, middle, or low GPA and GPAs were binned into tertiles for grouping purposes). From the 34 students in the course, a total of ten groups were formed: seven groups of three students, two groups of four students, and one group of five students (this group was formed due to one student withdrawing early in the course and combining two groups of three). In addition to working in their groups on the two-stage exams, students also were required to turn in their weekly homework in the same groups and were encouraged to work in these groups to solve problems during class time.

Data Collection

Two types of data were collected for this study: 1) performance data and 2) survey data. Student performance on the two-stage exams was collected, and each component of the exam (individual and group) was scored as a percentage out of 100. Individual and group portions of the exams were graded identically using the same rubric. Additionally, overall student performance and final exam performance were collected as a percentage out of 100. The final exam was taken individually but mirrored the same general format as the individual and group exams (open-ended problems). During the last week of the course, students were asked to complete an online survey on Canvas (worth one point towards their grade) that asked questions regarding their experience in the course, including that with two-stage exams. The survey was designed de novo for this study and included Likert-type statements where students rated their agreement with the statements on four levels (strongly agree, agree, disagree, and strongly disagree). Twenty-six students (86.7 %) completed the survey.

Data Analysis

Average scores on each individual component of the exam were compared to the average scores on each group component of the exam using t-tests. The benefit of group exams on individual student performance was calculated as the difference between a student's group exam average and individual exam average (G-I difference). The impact of the potential group benefit on individual final exam performance (as measured as their total percentage score on the final exam out of 100) was modeled (controlling for college GPA as a potentially confounding factor) using multiple linear regression in the software package R.[37] By accounting for college GPA, multiple linear regression models allow for more confident interpretation of the results.[33,38] The outcome variable for the model was final exam score, and the response variables were G-I average and college GPA. Survey data were compiled, and student responses to each question were tabulated as percentages out of 100. All data storage and handling were done using Microsoft Excel (Microsoft, Redmond, WA), and all statistical tests were done using the software package R.[37]

RESULTS

Student Performance on the Individual and Group Components of Two-stage Exams

Student performance on the individual component of each exam was compared to the group component of each exam to determine how group performance compares to individual performance. Average student performance on each of the individual and group components of the five exams (and the overall average) are shown in Figure 1. Average group performance was significantly higher than average individual performance for each exam and for the overall average comparison ($p < 0.01$ for each).

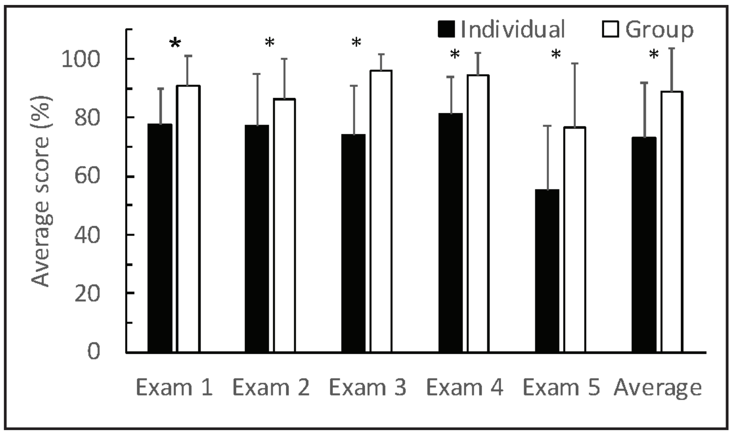

Figure 1: *Student performance on individual and group components of two-stage exams. Data are presented as average ± standard deviation. All differences were significant (*p < 0.01).*

Benefit of Group Component on Performance

While group performance was significantly higher than individual performance on each exam, it is likely that some students benefited disproportionally more than others on the group component. To investigate this, the individual average score (as a percentage out of 100) was subtracted from the group average score (as a percentage out of 100) for each student (from now on termed the G-I difference). If the G-I difference was positive for a particular student, it means that their group score was higher than their individual score. If the G-I difference value was negative for a particular student, it means that their group score was lower than their individual score. If the G-I difference was zero, then the group and individual scores were equal. As shown in Figure 2, there was a wide range of G-I differences for the students in the course, ranging from zero (no benefit from group component) to positive 8.8%. There were no negative G-I differences, indicating that no students individually outperformed their groups for the average of the five exams. The average G-I difference for all students was 3.1% ± 2.1%. Out of the 150 total exams given (five exams each for 30 students), the G-I difference was negative 13 times (8.6%), equal to zero 19 times (12.7%), and positive 118 times (78.7%), suggesting that the vast majority of the time the group exam scores were higher than individual exam scores.

To test whether the benefit that students received from the group component of the two-stage exams impacted future performance in the course, the G-I difference was compared to performance on the final exam (which was purely an individual effort). If there was no relationship between the G-I difference and final exam score, this could possibly suggest that students were learning from the group testing experience and were able to apply this to a future individual effort. As shown in Figure 3, there was a negative relationship between a student's G-I difference and their final exam score (slope = -3.9, p = 0.02) , suggesting that students who received the most benefit from the group component were not able to successfully transfer this benefit to the individual final exam. The average (± standard deviation) on the final exam was significantly lower than the average of the individual exam scores (58.7 ± 18.4 vs 78.8 ± 11.4, p < 0.0001)

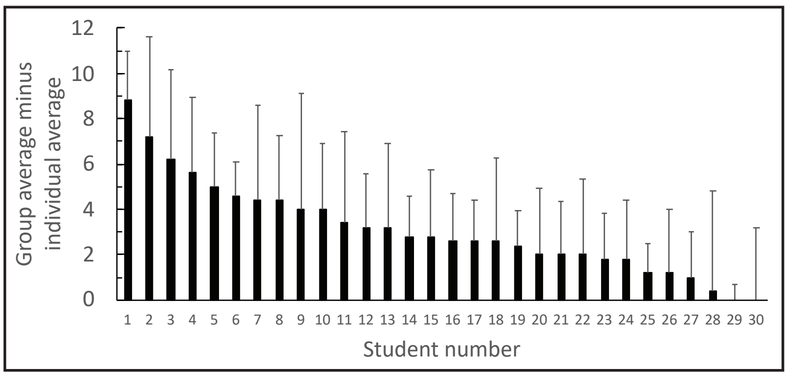

Figure 2: *Student benefit from group component of two-stage exams. For each student, their individual average exam score calculated based on the five exams was subtracted from their calculated group average exam score based on the five exams. The positive values indicate that student group performance was higher than student individual performance. Data are presented as average ± standard deviation.*

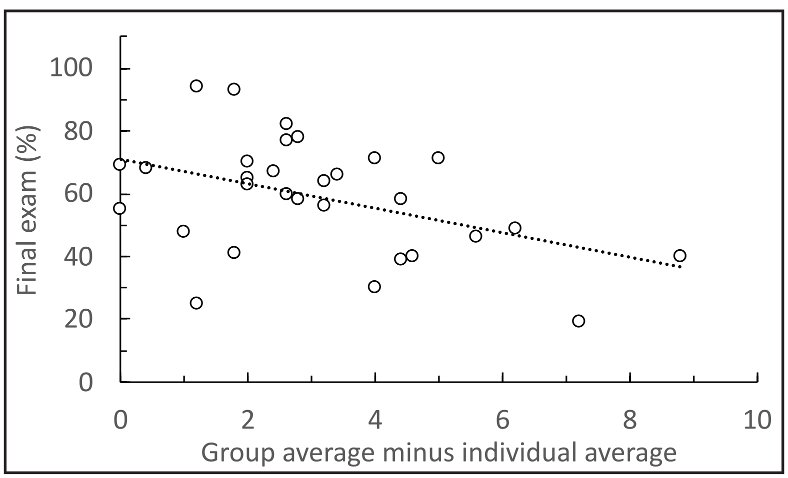

Figure 3: *Impacts of group component of two-stage exam (G-I difference) on final exam performance. Each open symbol represents an individual student, and the dashed line is a linear regression between the G-I difference (x-axis) and final exam score (y-axis). The negative slope suggests that the students who benefited the most from the group component of the two-stage exams had lower performance on the final exam.*

To account for possible confounding effects of student aptitude on final exam performance, a multiple linear regression model was constructed using final exam score as the outcome variable and student college GPA and G-I difference as explanatory variables. The result from this model (model intercept = 40.1 ± 16.8 (p = 0.02), GPA estimate = 11.4 ± 5.9 (p = 0.06), G-I estimate = -3.4 ± 1.5 (p = 0.02), adjusted R2 = 0.24) was consistent with the linear regression shown in Figure 3, suggesting that student G-I difference indeed was related to final exam score.

Student Perception of Two-stage Exam

At the end of the course, students completed an online survey to provide feedback about their perceptions of group work and the two-stage exams. As shown in Table 1, students were overwhelmingly positive with regards to working in groups and the two-stage exams. Notably, 100% of students agreed or strongly agreed that the group component of the two-stage exam helped them learn the material more than if only an individual component was used. Additionally, only 3.8% of students agreed that students in their groups unfairly benefited from the group component of the exam.

In addition to these items in Table 1, students were also asked about other features of two-stage exams. All 26 students who completed the survey agreed that the two-stage exams should be used again in this MEB course, and 76.9% reported that the instructor should form groups (as opposed to students forming their own groups). Lastly, when asked what percentage of the grade should be apportioned to the group part of the exam, 46.2% said 25%, and 53.8% said 20%. No students selected the 15%, 10%, or 5% grading options.

TABLE 1 Summary of student survey responses about group work and two-stage exams.				
Statement	**% Strongly Agree**	**% Agree**	**% Disagree**	**% Strongly Disagree**
Working with other students helps me learn.	61.5	34.6	0.0	3.8
I enjoyed working with my group in this course.	50.0	46.2	0.0	3.8
I feel that I have made new friends because of my assigned group from this class.	34.6	42.3	19.2	3.8
Assigned groups should be formed in every chemical engineering course.	23.1	57.7	15.4	3.8
The group part of the exam helped me undestand the material more clearly than if we did not have the group part of the exam.	73.1	26.9	0.0	0.0
Students in my group unfairly benefited from the group part of the exam.	3.8	3.8	69.2	23.1
Everyone in my group contributed equally to the group part of the exam.	26.9	73.1	0.0	0.0
Group exams should be used in academic year chemical engineering courses.	61.5	34.6	3.8	0.0

Instructor Observations of Two-stage Exam

While the individual component of the two-stage exam proceeded as a normal exam would (very quiet, students focused, etc.), the classroom environment was extremely dynamic and engaging during the group component of the two-stage exams. Students were very enthusiastic and animated, with groups seemingly working very well together and fully cooperating with each other. Every student was participating to some extent in their groups, and never was a student simply sitting idly while the rest of their group did all of the work. While most groups worked together as a team to answer each of the two exam questions sequentially, on a few occasions some groups ripped the two pages of the exam apart and split into smaller groups to work on each problem at the same time. Overall, the two-stage exam environment felt very positive and logistically proceeded very smoothly.

DISCUSSIONS

The results from this study demonstrate that chemical engineering student performance on the group component of two-stage exams in an MEB course were significantly higher than individual performance. However, the increased group performance did not seem to translate into later success as measured by final exam scores. Additionally, students felt extremely positive towards two-stage exams and recommended their use in future offerings of this course and others. These findings are in line with the literature on two-stage exams used in other STEM disciplines and provide a foundation for future assessment of two-stage exams in chemical engineering courses.

The findings from this study largely agree with those of other studies in that performance on the group component of a two-stage exam is higher than individual performance.[16-28] While the impacts of two-stage exams have not yet been reported for chemical engineering courses, the results of this study did match the results from Fengler and Ostafichuk,[17] who assessed a mechanical engineering course and found that group performance exceeded individual performance. While the reasons for the immediate benefits on the group component may vary, it is likely more than groups simply rewriting the answers from the "best" student in the group, as many studies have reported that groups usually outperform any individual in the group.[16-17] In addition, chemical engineering students from this study and mechanical engineering students from the Fenger and Ostafichuk study[17] had strong positive attitudes towards the use of two-stage exams. Both MEB and the mechanical engineering course from the Fengler and Ostafichuk study[17] were sophomore-level courses, so it would be interesting to determine if the positive attitudes are due to the two-stage exams being used in lower-division courses (i.e. would upper-division students have similar attitudes). There could also potentially be a beneficial longitudinal impact on upper-division courses that evolves from the use of two-stage exams in lower-division courses, in that students learn to work cooperatively and make friends early in a curriculum, and therefore over time improve as students. As there are clear benefits for using two-stage exams in MEB, further use and assessment of two-stage exams in other chemical engineering courses would be valuable to determine if students benefit and have favorable views towards them.

While there was an immediate benefit of the group component on student exam scores, there was not a longer-term positive impact with regards to the final exam score. This result is evident from the inverse relationship between final exam scores and the difference between group component averages and individual component averages: students who received the largest benefit from the group component of the two-stage exam performed the lowest on the final exam (which was entirely an individual effort). This result is not novel, as other studies have also reported mixed results or negative results in terms of retention or the ability of students to recall knowledge from an earlier point of time to apply it during a current assessment.[11,22,31,32] These studies, however, assessed courses that were taught during standard semesters and measured retention anywhere from nine days to seven weeks after the two-stage exam, and thus the mixed retention results or lack of retention could be due to a variety of reasons. In this case, a possible reason for lack of a longer-term benefit of the group component of the two-stage exam is the accelerated nature of the MEB course in this study, which was taught in six weeks as compared to a typical 16-week semester. Students worked together in groups on weekly two-stage exams, but even with this intense exam schedule, there may not have been enough time for students to transfer the skills acquired during the group component of the two-stage exam to the individual final exam. Further research is warranted to determine if there are longer term benefits to using two-stage exams in a typical 16-week semester MEB course.

After using two-stage exams in this MEB course, the following recommendations can be made for using them in other chemical engineering courses. First, be sure to explain to students early in the course about why two-stage exams are being used (citing the literature as appropriate). As student buy-in (or acceptance) to active learning is an important feature for successful use of these techniques,[39,40] explaining the rationale for using two-stage exams will let students know why they will be assessed in a unique fashion. This may also improve student attitudes towards the two-stage exams, even though the literature shows overwhelmingly positive results towards students' acceptance of two-stage exams.[12,13,17] The timing and logistics of the two-stage exam are critical as well. Since the course in this study was taught in the summer, it met for three two-hour and 50-minute sessions a week, allowing ample time for conducting two-stage exams (each exam took about one hour). However, doing this in a typical 50-minute class may be more difficult, as each component of the exam would have to be shortened. If possible, evening exams could be held which would allow for more time to implement the two-stage exams.[12,13] Additionally, the group component of the two-stage exam could be modified such that students only complete a portion of the exam, which would save time. Lastly, it is important to observe group dynamics and perhaps require group peer evaluations at some point during the course in order to promote positive collaboration during two-stage exams. As the groups in this summer MEB course were permanent for all six weeks of the course, students completed a mid-course group evaluation where they used a rubric to assess each other's contributions to group efforts in-class, on homework, and on two-stage exams. As accountability is key for successful group work,[41,42] this type of peer evaluation may help all members of a group contribute equally to their group. If the instructor observes, however, through visual observation or group feedback that a group is not working effectively, it is important to intervene and discuss the dynamics of that group with the students and offer assistance if needed to help the group work more effectively going forward.[43]

Limitations

This study sought to describe and characterize a novel use of two-stage exams in an MEB course, but there are some limitations. This MEB course was taught in an accelerated six-week summer session while covering the same amount of material as in a typical 16-week semester. Due to the increased pace, the results from this study may be different when compared to using two-stage exams in a normal semester. In addition, this summer course had a lower enrollment than a semester course, yielding a small number of students in the study population. However, the population demographics were similar to those in the academic year MEB course. While students' responses to working in groups and two-stage exams were extremely positive, students worked in permanent groups not only during the two-stage exams, but also to turn in weekly homework sets and during in-class problem solving activities. If groups are only used during two-stage exams, or if groups change throughout the course, students may not be as familiar with each other, which could result in two-stage exams being less effective. Therefore, repeating this study with a larger number of students during the academic year and with varying group structures (rotating groups, groups only used during exams, etc.) is warranted to confirm and extend these results.

ACKNOWLEDGEMENTS

I would like to thank James Cooke for helpful discussions about implementing two-stage exams in the course, Gabe Walton for the use of his custom group-formation software, and the students enrolled in the material and energy balances course for being a part of the study.

REFERENCES

1. Prince M (2004) Does active learning work? A review of the research. *Journal of Engineering Education* 93(3):223-231. 10.1002/j.2168-9830.2004.tb00809.x.
2. Freeman S, et al. (2014) Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111(23):8410-8415. 10.1073/pnas.1319030111.
3. PCAST (2012) *Report to the President, Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*. Executive Office of the President, President's Council of Advisors on Science and Technology. Washington, D.C.
4. AAAS (2011) *Vision and Change in Undergraduate Biology: A Call to Action*. Washington, DC.
5. Felder RM, Felder GN, and Dietz EJ (1998) A longitudinal study of engineering student performance and retention. V. Comparisons with traditionally-taught students. *Journal of Engineering Education* 87(4):469-480. 10.1002/j.2168-9830.1998.tb00381.x.
6. Bullard L, Felder R, and Raubenheimer D (2008) Effects of active learning on student performance and retention in chemical engineering. *ASEE Annual Conference*.
7. Amos DA, Pittard CM, and Snyder KE (2018) Active learning and student performance in a material and energy balance course. *Chem. Eng. Ed*. 52(4):277-286.
8. Silberman M (1996) *Active Learning: 101 Strategies To Teach Any Subject* (ERIC).
9. Smith MK, et al. (2009) Why peer discussion improves student performance on in-class concept questions. *Science* 323(5910):122-124. 10.1126/science.1165919.
10. Council NR (2000) *How People Learn: Brain, Mind, Experience, and School:* Expanded Edition. National Academies Press.
11. Cooke JE, Weir L, and Clarkston B (2019) Retention following two-stage collaborative exams depends on timing and student performance. *CBE—Life Sciences Education* 18(2):ar12. 10.1187/cbe.17-07-0137.
12. Wieman CE, Rieger GW, and Heiner CE (2014) Physics exams that promote collaborative learning. *The Physics Teacher* 52(1):51-53. 10.1119/1.4849159.
13. Rieger GW and Heiner CE (2014) Examinations that support collaborative learning: The students' perspective. *Journal of College Science Teaching* 43(4):41-47.
14. Newstetter WC and Svinicki MD (2015) Ch. 2: Learning Theories for Engineering Education Practice. *Cambridge Handbook of Engineering Education Research*. Cambridge University Press. Cambridge, United Kingdom. 10.1017/CBO9781139013451.005.
15. LoGiudice AB, Pachai AA, and Kim JA (2015) Testing together: When do students learn more through collaborative tests? *Scholarship of Teaching and Learning in Psychology* 1(4):377.
16. Bruno BC, et al. (2017) Two-stage exams: A powerful tool for reducing the achievement gap in undergraduate oceanography and geology classes. *Oceanography* 30(2):198-208. 10.5670/oceanog.2017.241.
17. Fengler M and Ostafichuk PM (2015) Successes with two-stage exams in mechanical engineering. *Proceedings of the Canadian Engineering Education Association (CEEA)*. doi.org/10.24908/pceea.v0i0.5744.
18. Jang H, Lasry N, Miller K, and Mazur E (2017) Collaborative exams: cheating? Or learning? *Am J Phys* 85(3):223-227. 10.1119/1.4974744.
19. Cao Y and Porter L (2017) Evaluating student learning from collaborative group tests in introductory computing. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*. pp 99-104.
20. Knierim K, Turner H, and Davis RK (2015) Two-stage exams improve student learning in an introductory geology course: Logistics, attendance, and grades. *Journal of Geoscience Education* 63(2):157-164. 10.5408/14-051.1.
21. Eaton TT (2009) Engaging students and evaluating learning progress using collaborative exams in introductory courses. *Journal of Geoscience Education* 57(2):113-120. 10.5408/1.3544241.
22. Leight H, Saunders C, Calkins R, and Withers M (2012) Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE—Life Sciences Education* 11(4):392-401. 10.1187/cbe.12-04-0048.
23. Newton G, et al. (2019) Two-stage (collaborative) testing in science teaching: Does it improve grades on short-answer questions and retention of material? *Journal of College Science Teaching* 48(4):64-73.
24. Gilley BH and Clarkston B (2014) Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching* 43(3):83-91.
25. Rao SP, Collins HL, and DiCarlo SE (2002) Collaborative testing enhances student learning. *Advances in physiology education* 26(1):37-41. 10.1152/advan.00032.2001.
26. Vázquez-García M (2018) Collaborative-group testing improves learning and knowledge retention of human physiology topics in second-year medical students. *Advances in physiology education* 42(2):232-239. 10.1152/advan.00113.2017.
27. Giuliodori MJ, Lujan HL, and DiCarlo SE (2008) Collaborative group testing benefits high-and low-performing students. *Advances in physiology education* 32(4):274-278. 10.1152/advan.00101.2007.
28. Zipp JF (2007) Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology* 35(1):62-76. 10.1177/0092055X0703500105.
29. Cortright RN, Collins HL, Rodenbaugh DW, and DiCarlo SE (2003) Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education* 27(3):102-108. 10.1152/advan.00041.2002.
30. Vogler JS and Robinson DH (2016) Team-based testing improves individual learning. *The Journal of Experimental Education* 84(4):787-803. 10.1080/00220973.2015.1134420.
31. Vojdanoska M, Cranney J, and Newell BR (2010) The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Appl Cognitive Psych* 24(8):1183-1195. 10.1002/acp.1630.
32. Ives J (2014) Measuring the learning from two-stage collaborative group exams. *arXiv preprint arXiv*:1407.6442.
33. Eddy SL and Hogan KA (2014) Getting Under the Hood: How and for Whom Does Increasing Course Structure Work? *Cbe-Life Sci Educ* 13(3):453-468. 10.1187/cbe.14-03-0050.
34. Shaffer JF (2016) Student performance in and perceptions of a high structure undergraduate human anatomy course. *Anatomical Sciences Education* 9(6):516-528. 10.1002/ase.1608.
35. Lieu R, Wong A, Asefirad A, and Shaffer JF (2017) Improving exam performance in introductory biology through the use of preclass reading guides. *CBE—Life Sciences Education* 16(3):ar46. 10.1187/cbe.16-11-0320.
36. Barkley EF, Cross KP, and Major CH (2014) *Collaborative Learning Techniques: A Handbook for College Faculty*. John Wiley & Sons.
37. Team RC (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
38. Theobald R and Freeman S (2014) Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE Life Sci Educ* 13(1):41-48. 10.1187/cbe-13-07-0136.
39. Cavanagh AJ, et al. (2016) Student buy-in to active learning in a college science course. *CBE—Life Sciences Education* 15(4):ar76. 10.1187/cbe.16-07-0212.
40. Brazeal KR, Brown TL, and Couch BA (2016) Characterizing student perceptions of and buy-in toward common formative assessment techniques. *CBE—Life Sciences Education* 15(4):ar73. 10.1187/cbe.16-03-0133.
41. Gueldenzoph LE and May GL (2002) Collaborative peer evaluation: Best practices for group member assessments. *Business Communication Quarterly* 65(1):9-20.
42. Cestone CM, Levine RE, and Lane DR (2008) Peer assessment and evaluation in team-based learning. *New Directions for Teaching and Learning* 2008(116):69-78.
43. Wolfgang CN (2009) Managing inquiry-based classrooms. *Science Scope* 32(9):14. ❐