2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
Paper ID #39710

Improving Student Perceptions of Learning through Collaborative Testing

Dr. Brian P. Helmke, University of Virginia

Brian Helmke (he/him/his) is currently Associate Professor of Biomedical Engineering at the University of Virginia. He received the B.S.E. in bioengineering from the University of Pennsylvania, the B.S.Econ. from The Wharton School of the University of Pennsylvania, and the Ph.D in bioengineering from the University of California, San Diego. Brian's research interests include cardiovascular physiology, cellular mechanobiology, and nanotechnology-based biomaterials. He is also interested in equity-based and collaborative pedagogies and alternative grading strategies.

Improving Student Perceptions of Learning through Collaborative Testing

Research on college teaching and learning has proposed active learning is a good practice [1]. Active learning could be defined as "an interactive and engaging process for students that may be implemented through the employment of strategies that involve metacognition, discussion, group work, formative assessment, practicing core competencies, live-action visuals, conceptual class design, worksheets, and/or games" [2]. Active learning approaches in science, technology, engineering, and mathematics (STEM) courses decrease failure rates and increase performance on summative assessments [3]. In Fink's taxonomy of significant learning experiences [4], course design should be learner-centered, so students are actively engaging not just with technical content in an engineering course but also reflecting on their own learning processes as they strive to become lifelong learners.

One strategy for promoting reflection and engagement is through opportunities for collaborative learning. Collaborative learning represents a joint intellectual effort by students (and often instructors) to mutually create or find meaning, solutions, or products [5]. Intentionally designed collaborative learning strategies can be mapped to Fink's taxonomy to support meaningful learning for students. This idea means that activities should be structured so that group members are interdependent and work together to achieve the learning objectives.

High-stakes assessments such as midterm and final tests are commonly used to evaluate retention and recall in biomedical engineering courses. In courses such as physiology, such assessments are professionally authentic for pre-health or pre-professional students planning to take the MCAT and/or licensure exams. In large format courses, grading time can be reduced by using electronic platforms to automate test administration and scoring. For students, these assessments often induce test anxiety that negatively impacts their learning [6].

Collaborative testing, or two-stage testing, has been used to introduce collaborative learning to high-stakes assessments in a number of STEM disciplines (e.g., [7]–[10]). In collaborative testing, students first complete a test individually, and then they complete all or part of the test again working in groups [5]. Test scores are typically computed as a weighted average of the individual and group scores. Collaborative testing has been demonstrated to improve performance and motivation of students [8]–[10]. However, it is not clear whether collaborative testing reduces affective components of student motivation such as test anxiety or whether situating collaborative learning in a test environment impacts students' learning strategies. In this study, a collaborative testing strategy was implemented in an introductory biomedical engineering physiology course to test the hypothesis that collaborative testing improves learning outcomes and student perceptions of their learning.

Methods

Course and study design

Interventions were made in a sophomore-level physiology course in biomedical engineering. The course is a required core course in the biomedical engineering major and is taught fall and spring semesters by two different instructors. Course objectives are the same in each section, and the instructors collaborate extensively on course and assessment design to minimize differences in learning outcomes between the sections. Both sections are roughly structured around reading assignments in the *Guyton and Hall Textbook of Medical Physiology* [11]. Main topics in the course include fundamentals of cell membranes, ion transport, and excitability; muscle contraction; cardiovascular physiology and control mechanisms; and respiratory physiology. The course is divided into four units of 3-4 weeks duration, and a summative test is administered at the end of each unit. A previously described pre-/post-course concept inventory [12] is administered to assess learning gain during the semester. The concept inventory is a ten-question multiple choice quiz administered through a web-based audience response system. The questions are designed to assess core concepts in homeostasis, cell and tissue excitability, muscle contraction, the cardiovascular system, and the respiratory system.

The Control group consisted of students in a fall section of the physiology course, which was taught using a blended format of interactive lectures, frequent formative assessments, and the four summative unit tests. Formative in-class and out-of-class activities have been described previously [12], and students completed the unit tests individually with a time limit of 50 min. The class mean test scores for each of the four unit tests during the semester were used to represent the Control group.

The Intervention group was comprised of three consecutive fall semesters with the same instructor as the Control group. All aspects of the course design were the same as the Control group except for the design of the unit tests. Each unit test consisted of two parts delivered through an online audience response system: an individual and a team part. All students in each class were formed into teams using a CATME survey [13], guided by principles of effective team formation [14]. Teams were re-mixed for each test in a manner that each student worked with different teammates for each of the four tests during the semester. The individual part had a 50-min time limit and consisted of 30 questions with formats of multiple choice, check all that apply, put steps in order, or fill in the blank. During the team part, students had 15 min to discuss the same 30 questions in teams and to change any or all of their answers from the individual part. Each team member submitted individual answers, so they had agency to disagree with their team's consensus answer. The individual part comprised two-thirds and the team part comprised one-third of the total test score [5].

Several class mean scores for each unit test were computed to represent the Intervention group: the individual score (just the individual part of the test), the team score (just the team part of the test), and the total score (2/3 × the individual score + 1/3 × the team score). The mean percent increase of team score over individual score was also computed. Finally, the average learning gain, a measure of the points gained relative to the points available to gain by making corrections on the team part, was computed as

$$\text{learning gain} = \frac{\text{team score} - \text{individual score}}{\text{total points available} - \text{individual score}}$$

and reported as a percent gain.

Student perceptions of the testing environment

Students in the Intervention group were asked to answer three Likert-type questions about their testing preferences after the first unit test each semester. Questions were administered during the next class period after the test using a web-based audience response system. Answers were encoded on a five-point scale as “strongly disagree” (1 point), “disagree” (2 points), “not sure” (3 points), “agree” (4 points), and “strongly agree” (5 points).

Motivation and learning strategies

The Motivated Strategies for Learning Questionnaire (MSLQ) [15] was administered using the course learning management system as part of an anonymous end-of-semester survey. The MSLQ is designed to assess motivation and use of learning strategies in a university-level course. Scores for each scale in the MSLQ were compiled according to authors’ instructions. Since MSLQ data for the Control group were not available, the MSLQ was administered in a spring section of the physiology course that had comparable individually completed unit tests [12]. A third group, the Teamwork group, was comprised of responses from another course taught by the same instructor that included formative teamwork training and team-based coursework but individually completed tests of course content.

Statistics

Data were collected and analyzed according to a protocol approved by the Institutional Review Board for Social and Behavioral Sciences (IRB-SBS).

Pre- and post-course concept inventory scores within each group (Control and Intervention) are reported as mean ± standard deviation and were compared by paired *t*-test. One-way ANOVA was used to compare between groups. The null hypothesis that means were not different was rejected when $p<0.05$.

Test scores were aggregated as mean class score for each test, and the mean and standard deviation of the class scores across multiple tests were reported for each group. The class scores for the Intervention group were compared to those for the Control group using the unpaired *t*-test. For the Intervention group, the mean class team score was compared to the mean class individual score using the paired *t*-test. In each comparison, the hypothesis that means were not different was rejected when $p<0.05$.

For scales selected from the Motivated Strategies for Learning Questionnaire (MSLQ), responses were compiled according to the authors’ instructions [15], and a mean score for each group was computed. The mean score for the Intervention group was compared to that for the Control group or the Teamwork group using the unpaired *t*-test, and the hypothesis that means were not different was rejected when $p<0.05$.

TABLE 1. Means of class unit test scores from Intervention (n=12 tests) and Control (n=4 tests) groups. All unit test scores are out of 100 points. *Within the Intervention group, Team and Total scores were significantly higher than Individual scores (one-way ANOVA, p<0.05).

	Control (n=4)	Intervention (n=12)				
		Indiv	Team	Total	Team vs. Indiv	Learning Gain
Mean	70.8	72.6	83.4*	76.2*	+18%	+36%
SD	8.6	3.8	4.1	3.9	3%	8%

Results

Course learning assessments

Overall learning in each group was assessed using a 10-question concept inventory administered on the first and last days of the course [12]. In the control group (n=38 students), concept inventory scores increased from 42±14 (mean±SD) points (out of 100) to 66±17 points, corresponding to a learning gain of 41±24%. In the intervention group (n=194 students over three semesters), concept inventory scores increased from 41±12 points to 69±16 points, corresponding to a learning gain of 46±31%. In both groups, the post-course scores were significantly higher than the pre-course scores (paired t-test, p<0.05), indicating learning occurred during the semester. The average learning gains were not different between groups (one-way ANOVA, p=0.4).

Mean class scores for each unit test were analyzed to compare assessment outcomes for Control and Intervention groups. In the Control group (4 tests), the mean of unit test scores was 70.8±8.6 (mean±SD) points (Table 1). In the Intervention group (12 tests over three semesters), the mean of class scores for individual parts of the unit tests was 72.6±3.8 points, and the mean of scores for team parts was 83.4±4.1. The mean of total scores, computed as 2/3 × the individual score + 1/3 × the team score, was 76.2±3.9 points. Neither the mean of individual scores nor the mean of total scores in the Intervention group were statistically different from the mean of Control group scores (one-way ANOVA, p=0.6 and p=0.1, respectively). Within the Intervention group, both the mean of team scores and the mean of total scores were significantly higher than the mean of individual scores (one-way ANOVA, p<0.05). On average the mean of team scores was 18±3% higher than the mean of individual scores, resulting in a learning gain of 36±8%.

Student perceptions of the testing environment

The Intervention group spanned three different semester offerings of the course in 2017 (n=64 students), 2018 (n=65), and 2019 (n=70). Figure 1 shows responses to three Likert-style questions about student perceptions of the collaborative testing activity broken down by semester and summed for the total Intervention group. 78% of students agreed or strongly agreed that discussion of the test with a team helped improve their scores, whereas 5% disagreed or strongly disagreed, and 18% were not sure. 74% of students agreed or strongly agreed that they felt more confident in their learning as a result of the team activity on the test, whereas 9% disagreed or strongly disagreed, and 17% were not sure. 82% of students agreed or strongly agreed that we should do a team activity on every test, whereas 4% disagreed or strongly disagreed, and 15% were not sure.

Motivation and learning strategies

Students in the Intervention group in fall 2018 (n=75 students) and in a Control group in spring 2019 (n=43 students) completed the Motivated Strategies for Learning Questionnaire (MSLQ) [15]. In order to determine whether structured teamwork activities in class impacted student motivation and learning independently of the collaborative test, a Teamwork group (n= 59 students) was comprised of responses from a different course taught by the same instructor that included formative teamwork training activities and summative teamwork assignments, but students in the Teamwork group completed unit tests individually. Scales measuring test anxiety, elaboration, peer learning, and help seeking were chosen for analysis (Table 2). The mean score for the Elaboration scale was significantly higher in the Intervention group than in either the Control or Teamwork groups (unpaired t-test, $p<0.05$). The Intervention group score for the Peer Learning scale was lower than the Teamwork group score (unpaired t-test, $p<0.05$) but not different from the Control group score (unpaired t-test, $p=0.9$). For the Text Anxiety and Help Seeking scales, the Intervention group score was not different from either the Control Group or Teamwork group scores.

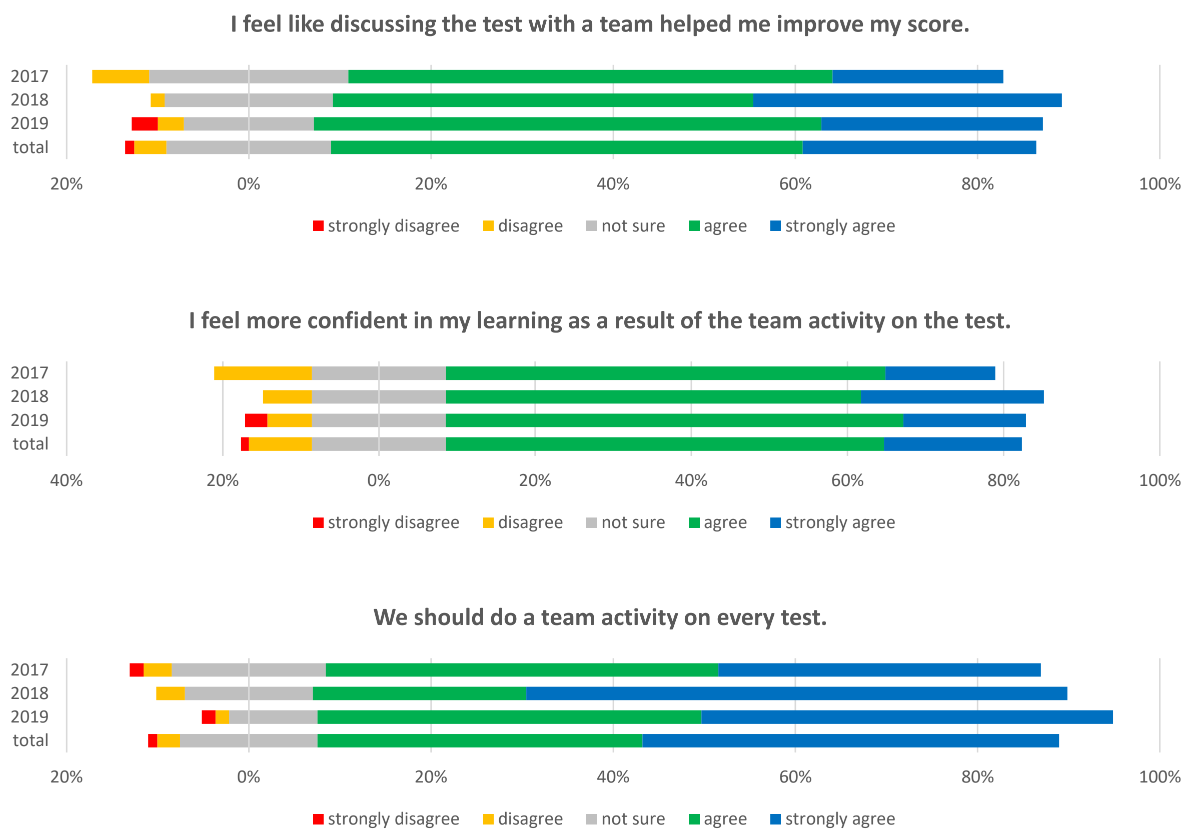

FIGURE 1. Student perceptions of the collaborative testing activity. Likert category results for the Intervention group total and broken into 3 semesters (2017, 2018, 2019).

TABLE 2. Selected scales from the Motivated Strategies for Learning Questionnaire (MSLQ). Values are mean±SD. *Intervention group score was significantly lower than Teamwork group score for Peer Learning scale (*t*-test, $p<0.05$). **Intervention group score was significantly higher than both Control and Teamwork group scores for Elaboration scale (*t*-test, $p<0.05$).

	Control	Intervention	Teamwork
Test Anxiety	4.61±1.24	4.83±1.37	4.58±1.41
Elaboration	5.23±0.83	5.57±0.83**	4.97±1.04
Peer Learning	4.40±1.74	4.35±1.59*	4.89±1.13
Help Seeking	4.68±1.28	4.50±1.53	4.71±0.94

Discussion

In collaborative learning, groups construct knowledge by co-creating products with shared meaning [5], building on a social constructivism theory. Students learn by interacting with and/or imitating others, i.e., peers and instructors. Since a learning goal in biomedical engineering is to join a professional community of practice, collaborative learning strategies have been described using a situated learning framework [16], [17]. Knowing is described as an individual's ability to participate successfully in the activities of the group or learning community. By connecting learners and supporting the ability to co-create meaning or products, collaborative learning strategies promote a shift of an individual's participation from the periphery to the center of the community through peer-to-peer interactions. The change in the learner is situated in the modes of participation in the professional community.

In this context, collaborative learning pedagogies represent an opportunity to increase students' motivation and support students' development of effective learning strategies. In content-heavy biomedical engineering courses, high-stakes assessments such as tests or exams often elicit anxiety, low expectancy to succeed, and low self-efficacy for students, corresponding to decreased motivation. Collaborative testing has been demonstrated to improve students' performance [8]–[10], leading to the hypothesis that collaborative testing also improves students' perceptions of their learning environment based on development of effective learning strategies.

Collaborative testing has been reported to improve retention and recall (e.g., [7]–[10]), support transferable skill development [7], and increase engagement [8]. Results in the physiology course described here support these outcomes. Individual test scores were not different across multiple semesters including the Control and Intervention groups, and team scores in the Intervention groups were increased by an average of 18% relative to the individual scores. Notably, each test contained questions that were categorized as either "retention and recall" or as "comprehension" [12]. Comprehension questions assess students' abilities to apply and integrate concepts in ways that are useful for engineering problem-solving approaches, consistent with the idea of transferable skill development. The average learning gain corresponding to the increased team scores was 36%. Importantly, the learning gain appeared greater for students scoring lower on the individual part, implying that those students benefited most from collaborative discussions. This observation agrees with previous reports [10] and leads to the suggestion that collaborative tests support an equitable classroom environment. This idea remains to be investigated.

The format and weighting scheme were chosen to be consistent with previously reported evidence of the positive impacts of collaborative testing. The weight assigned to the team portion of the test has been varied from 20% [7], [10] to 33% [5]. Test formats may also include varying the timing of the team part or may include open-ended short answer prompts [9]. The test format for the unit tests in this physiology course was limited to multiple choice, check all that apply, put steps in order, and fill in the blank questions because (1) this format allowed comparison to the previous semester in which tests were completed individually and (2) this format is professionally authentic to common tests such as medical board exams, professional licensure exams, and graduate comprehensive exams. It would be interesting to explore adjusting the weight of the team portion of the total test score, since it may impact test anxiety and student motivation.

Most students expressed satisfaction with the collaborative test, as indicated by survey responses after the first test of the semester. Students believed that the team part of the test improved their score and increased their confidence in their learning. Active discussion among team members during the team part of the test demonstrated that students were engaged in their learning, which agrees with other reports [8]. In end-of-course surveys, several students mentioned that they viewed the collaborative tests as a learning activity rather than as an evaluation or judgement. Although mindset was not measured directly, it is possible that these comments are consistent with a shift in mindset from a performance focus (earning points) to a learning focus.

In order to determine whether increased performance and satisfaction were connected to changes in motivation and learning strategies, students were asked to complete the Motivated Strategies for Learning Questionnaire (MSLQ) [15]. Scales associated with motivation were based on value-expectancy theory. Of particular interest to this study was the affective component associated with test anxiety. Surprisingly, test anxiety was not different in the Intervention group compared to the Control group, in contrast with previous reports [18], [19]. Other factors may impact test anxiety in addition to the test itself, including the classroom climate, class size, and nature of the content. Low-stakes practice of group discussions similar to the collaborative test may also help to reduce anxiety of the unknown for students who have not previously participated in this testing method. Nevertheless, it is not clear why students responded positively to the survey questions in Fig. 1 but did not report reduced test anxiety.

Based on the situated learning framework, three MSLQ scales associated with learning strategies were examined. Elaboration strategies such as note taking, summarizing, and creating analogies help students build connections among course concepts and store information into long-term memory. The Peer Learning scale measured to what extent students actively collaborate with peers on course content. The Help Seeking scale is associated with students' ability to recognize limitations in their knowledge and to identify a peer or instructor to assist. These three scales were chosen because of the potential synergy among these strategies. Elaboration with peers not only supports memory through collaboration but also identifies gaps in collective knowledge. Students have opportunities to close those gaps by seeking help from other peers or through discussions with instructors. In the physiology class, students had frequent opportunities to discuss with peers and answer questions from the instructor using an online audience response system, and the instructor and teaching assistants moved throughout the classroom checking in on groups during those discussions.

Interestingly, parallel trends in these learning strategies scales did not emerge. Only Elaboration was higher in the Intervention group than in the Control group. One possible explanation is that the classroom environment supported peer learning and help seeking equally in the two groups, but students in the Intervention group felt like they had extra opportunities to teach each other and check their understanding in high-stakes moments. To test this idea, the MSLQ data were compared to data from another course taught by the same instructor in which the pedagogy included activities explicitly centering teamwork skills (Table 2). This Teamwork group participated in structured activities specifically focused both on communication and community building in groups and on working together to achieve a learning objective. A student in the Teamwork group commented, “I think that anyway [*sic*] that a professor/class can facilitate or teach effective communication early in a group causes more effective group work and is a beneficial learning experience…Most students don't realize the importance of communication early, and then when an issue arises communication can break down pretty easily.” Students in the Teamwork group appear to recognize the value of establishing community norms within their teams to support synergistic learning strategies among team members. Despite the extensive teaching about teamwork strategies, students in the Teamwork group took traditional tests individually. The MSLQ Peer Learning score was significantly lower for the Intervention group than for the Teamwork group, as expected. However, the Elaboration score was higher for the Intervention group than for the Teamwork group, supporting the idea that the chance to explain, summarize, and check understanding in high-stakes situations may support students' development of elaboration strategies.

It was somewhat surprising that only the Elaboration score was different in the Intervention group compared to the Control group. Factors such as persistence of team composition, type of learning activities, and course content might have impacted this result. This physiology course is a first course in the biomedical engineering major; students at this institution declare their major after completing a common first-year core curriculum. Teams in the Intervention group were different for each test in order to help students get to know more of their peers in the class. Keeping teams together throughout the semester may be required to increase in peer learning strategies, as observed in the Teamwork group taken from another course with activities designed to support team cohesion and interdependence. Learning activities and course content focused around engineering design or problem-solving process rather than retention, recall, and comprehension may also produce different effects on Peer Learning and Help Seeking. In fact, the Teamwork group was taken from a course involving mathematical problem-solving processes for transport phenomena [20], and the Peer Learning score was higher than in the Intervention group. It would be interesting in the future to determine whether persistent teamwork or course content was the primary determinant of the increased Peer Learning score.

A few limitations to interpreting the MSLQ data exist. Students in each group completed the MSLQ at the end of the course, and prompts were directed at the course rather than at the collaborative test activity specifically. As a result, it is difficult to separate effects of this specific activity from effects of other components of the course design. Since the MSLQ was administered to a control group in a course section taught by another instructor, student-instructor interactions or other aspects of the individual instructors may have influenced the results. Finally, the Teamwork group was comprised of junior-level students in a course with substantial content differences from the physiology course, creating the possibility that content differences may play a role in the motivation and/or learning strategies developed by the students in these two different courses.

In summary, a collaborative testing strategy involving an individual portion and a team portion of each unit test was implemented in a sophomore-level physiology course. Compared to traditional individual tests, the individual parts of the collaborative tests were not different, but team scores were significantly increased. A majority of students expressed satisfaction and increased confidence after completing the collaborative tests; however, test anxiety was not decreased. Students increased their use of elaboration strategies in the course but did not increase use of peer learning or help seeking strategies. Questions remain about the interaction between length of time in teams and the use of elaboration and peer learning strategies. Future work could investigate these questions.

References

[1] A. W. Chickering and Z. F. Gamson, "Seven principles for good practice in undergraduate education," *AAHE Bull.*, vol. 39, no. 7, pp. 3–7, Mar. 1987.

[2] E. P. Driessen, J. K. Knight, M. K. Smith, and C. J. Ballen, "Demystifying the meaning of active learning in postsecondary biology education," *CBE—Life Sci. Educ.*, vol. 19, no. 4, p. ar52, Oct. 2020, doi: 10.1187/cbe.20-04-0068.

[3] S. Freeman *et al.*, "Active learning increases student performance in science, engineering, and mathematics," *Proc. Natl. Acad. Sci.*, vol. 111, no. 23, pp. 8410–8415, Jun. 2014, doi: 10.1073/pnas.1319030111.

[4] L. D. Fink, *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses (Revised and updated ed.)*. San Francisco: Jossey-Bass, 2013.

[5] E. F. Barkley, K. P. Cross, and C. H. Major, *Collaborative Learning Techniques: A Handbook for College Faculty*. John Wiley & Sons, 2014.

[6] B. P. Self, J. Landy, J. M. Widmann, J. Chen, and M. Kerfs, "The mechanics of SUCCESS: How non-cognitive and affective factors relate to academic performance in engineering mechanics," in *2021 ASEE Virtual Annual Conference Content Access Proceedings*, Virtual Conference: ASEE Conferences, Jul. 2021, p. 37876. Accessed: Feb. 27, 2023. [Online]. Available: https://peer.asee.org/37876

[7] A. MacKenzie, "Revelations of a collaborative midterm in an engineering technology management course," *Proc. Can. Eng. Educ. Assoc. CEEA*, Nov. 2019, doi: 10.24908/pceea.vi0.13760.

[8] J. Nease, V. Leung, S. Ebrahimi, B. Levinson, I. K. Puri, and C. D. M. Filipe, "A learner's journey towards a chemical engineering degree," *Can. J. Chem. Eng.*, vol. 99, no. 10, pp. 2149–2162, Oct. 2021, doi: 10.1002/cjce.24140.

[9] J. E. Cooke, L. Weir, and B. Clarkston, "Retention following two-stage collaborative exams depends on timing and student performance," *CBE—Life Sci. Educ.*, vol. 18, no. 2, p. ar12, Jun. 2019, doi: 10.1187/cbe.17-07-0137.

[10] J. F. Shaffer, "Student performance on and perceptions of collaborative two-stage exams in a material and energy balances course," *Chem. Eng. Educ.*, vol. 54, no. 2, pp. 52–58, 2020.

[11] J. E. Hall, *Guyton and Hall Textbook of Medical Physiology*, 14th ed. Philadelphia: Elsevier, 2020.

[12] W. H. Guilford and B. P. Helmke, "Summative versus formative assessments in teaching physiology to biomedical engineering students: a comparison of outcomes," in *2017 ASEE*

Annual Conference & Exposition Proceedings, Columbus, Ohio: ASEE Conferences, Jun. 2017, p. 28882. doi: 10.18260/1-2--28882.

[13] R. A. Layton, M. L. Loughry, M. W. Ohland, and G. D. Ricco, "Design and validation of a web-based system for assigning members to teams using instructor-specified criteria," *Adv. Eng. Educ.*, vol. 2, no. 1, pp. 1–28, 2010.

[14] B. Oakley, R. M. Felder, R. Brent, and I. Elhajj, "Turning student groups into effective teams," *J. Stud. Centered Learn.*, vol. 2, no. 1, pp. 9–34, 2004.

[15] P. R. Pintrich, D. A. F. Smith, T. Garcia, and W. J. Mckeachie, "Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ)," *Educ. Psychol. Meas.*, vol. 53, no. 3, pp. 801–813, Sep. 1993, doi: 10.1177/0013164493053003024.

[16] W. C. Newstetter and M. D. Svinicki, "Learning Theories for Engineering Education Practice," in *Cambridge Handbook of Engineering Education Research*, A. Johri and B. M. Olds, Eds., 1st ed.Cambridge University Press, 2014, pp. 29–46. doi: 10.1017/CBO9781139013451.005.

[17] J. Lave and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*. Cambridge UP, 1991.

[18] C. Pandey and S. Kapitanoff, "The influence of anxiety and quality of interaction on collaborative test performance," *Act. Learn. High. Educ.*, vol. 12, no. 3, pp. 163–174, Nov. 2011, doi: 10.1177/1469787411415077.

[19] Ioannou and A. R. Artino Jr., "Learn more, stress less: exploring the benefits of collaborative assessment," *Coll. Stud. J.*, vol. 44, no. 1, p. 189, Mar. 2010.

[20] B. P. Helmke, "Barriers to learning in a large flipped biotransport course," in *2017 ASEE Annual Conference & Exposition Proceedings*, Columbus, Ohio: ASEE Conferences, Jun. 2017, p. 27649. doi: 10.18260/1-2--27649.