
ORIGINAL ARTICLE

Requiring students to justify answer changes during collaborative testing may be necessary for improved academic performance*

Niu Zhang, MD and Charles N.R. Henderson, DC, PhD

Objective: Three hypotheses were tested in a chiropractic education program: (1) Collaborative topic-specific exams during a course would enhance student performance on a noncollaborative final exam administered at the end-of-term, compared to students given traditional (noncollaborative) topic-specific exams during the course. (2) Requiring reasons for answer changes during collaborative topical exams would further enhance final-exam performance. (3) There would be a differential question-type effect on the cumulative final exam, with greater improvement in comprehension question scores compared to simple recall question scores.

Methods: A total of 223 students participated in the study. Students were assigned to 1 of 2 study cohorts: (1) control – a traditional, noncollaborative, exam format; (2) collaborative exam only (CEO) – a collaborative format, not requiring answer change justification; and (3) collaborative exam with justification (CEJ) – a collaborative exam format, but requiring justification for answer changes.

Results: Contrary to expectation (hypothesis 1), there was no significant difference between control and CEO final exam scores ($p = .566$). However, CEJ final exam scores were statistically greater (hypothesis 2) than the control ($p = .010$) and CEO ($p = .011$) scores. There was greater collaboration benefit when answering comprehension than recall questions during topic-specific exams ($p = .000$), but this did not differentially influence study cohort final exam scores ($p = .571$, hypothesis 3).

Conclusion: We conclude that test collaboration with the requirement that students explain the reason for making answer changes is a more effective learning tool than simple collaboration that does not require answer change justification.

Key Indexing Terms: Collaboration; Educational Measurement; Chiropractic; Education

J Chiropr Educ 2017;31(2):96–101 DOI 10.7899/JCE-16-5

INTRODUCTION

Collaborative learning can refer to any instructional method in which students work together in small cohorts toward a common goal.¹ Learning strategies in collaborative learning are active and student-centered, with a variety of collaborative learning strategies being shown to promote knowledge and skill acquisition at all grade levels and in multiple professions.^{2–4}

Quizzes and examinations are the most common methods of assessing academic achievement and assigning course grades. They facilitate learning assessment in a large number of students over a short period of time. The

instructor is provided feedback regarding what students have learned, and students discover the scope and depth of their knowledge. Collaborative testing, in which students work together to discuss examination questions and arrive at consensus to test answers, is an extension of collaborative learning that has been reported to decrease test anxiety, and increase critical thinking, communication, and team building skills.^{5,6} It also has been studied in a wide range of disciplines, including mathematics,⁷ nursing,⁸ and language training.⁹ While examining the effects of collaborative testing on academic performance, 2 chiropractic education studies reported that students involved in collaborative testing achieved higher topical exam scores and course grades compared to noncollaborative testing. However, noncollaborative final exam scores at the end of term were not significantly improved.^{10,11} Woody et al.¹² reported that, although students reported satisfaction with the collaborative

*This paper was selected as a 2016 Association of Chiropractic Colleges – Research Agenda Conference Prize Winning Paper – Award funded by the National Board of Chiropractic Examiners.

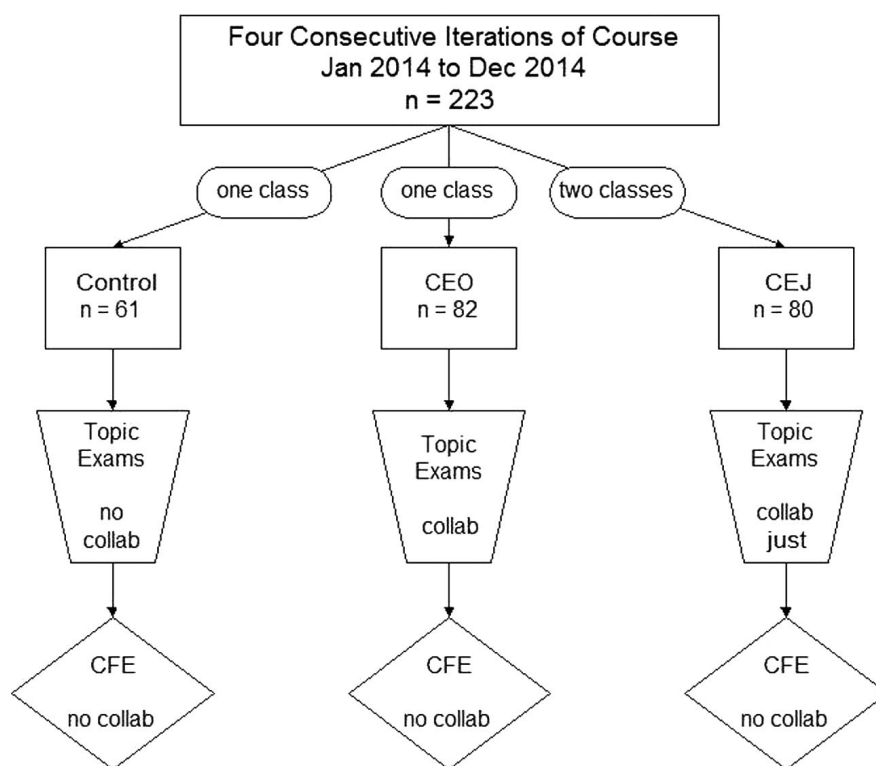


Figure 1 - Study design flowchart. A total of 223 3rd-quarter students participated in the study across 4 consecutive iterations of a 3-credit class presenting immunology and endocrinology content (January 2014–December 2014). Each class offered 20 lectures of 50 minutes each. Class members in the 4 course offerings were assigned to 1 of 3 study cohorts: (1) Control (traditional, noncollaborative, exam format), (2) CEO – a collaborative exam format, not requiring answer change justification), and 3) CEJ – a collaborative exam format requiring answer change justification). Therefore, exam collaboration occurred during topic-specific exams in only the CEO and CEJ cohorts.

model and course performance was improved as a result of participating in cohort testing, material retention was not significantly improved. In addition, students have expressed concerns that unprepared peers may obtain higher scores with collaborative exams by simply copying answers.¹³

In consideration of these studies, we were interested in evaluating the summative and formative use of collaborative exams in a chiropractic education program. We hypothesized that: (1) Collaborative topic-specific tests during the course will enhance student performance on a noncollaborative, cumulative final exam administered at the end-of-term, compared to a control cohort of students given traditional (noncollaborative) topic-specific exams during the course. (2) Requiring students to provide reasons for changing answers during collaborative topical exams will further enhance performance on the cumulative final exam administered at the end-of-term. (3) There will be a differential final exam question-type effect across the collaboration study cohorts (Collaborative Exam Only [CEO] and Collaborative Exam with Justification [CEJ]), with substantially greater score increases, compared to the control cohort, for comprehension versus recall questions.

METHODS

Student Participants

The Palmer College of Chiropractic institutional review board granted this educational method study an exemption from formal review, and permission was obtained from all students to use deidentified performance assessments for this study and subsequent publications. A sample size calculation was performed for an anticipated 5%-point change in cumulative final exam scores (our primary outcome). We considered this to be a meaningful change in exam performance, representing a 0.4 effect size (Cohen's *d*). Consequently, the required cohort sample size (*n*, at 2-tailed $\alpha = .05$ and $\beta = .80$) was estimated to be *n* = 60 students/cohort.

A total of 223 3rd-quarter students participated in the study across 4 consecutive iterations of a 3-credit class presenting immunology and endocrinology content (January 2014–December 2014; Fig. 1). Each class offered 20 lectures of 50 minutes each. Class members in the 4 course offerings were assigned randomly, en bloc, to 1 of 2 study cohorts: (1) Control – a traditional, noncollaborative, topical exam format (1 class, *n* = 61); (2) CEO – a collaborative topical exam format, not requiring answer change justification (1 class, *n* = 82); 3) CEJ – a collaborative topical exam format, but requiring justification

Table 1 - Fixed-choice List of Reasons for Switching Answers on Exams

1.	Group members backed answers with facts
2.	Overthought the first time
3.	Misread the question the first time
4.	Group members had the same answer
5.	Too many of same letter answer in a row the first time
6.	Underprepared the first time
7.	Did not understand the material the first time
8.	Unsure of the answer, chose different options to increase chance of getting points
9.	Misread the question – dyslexic
10.	Stress on individual test, prefer group test
11.	Peer pressure
12.	Wrong answer the first time
13.	Guessed first time
14.	Ran out of time
15.	Mismatched Scantron the first time
16.	Deductive reasoning as a unit

tion for any answer changes (2 classes combined, $n = 43$ and $n = 37$; totaling 80 students).

This course was offered 4 times during the academic year. The instructor, lecture format, and course content were identical in each presentation of the course. Only topical exam formats varied across the 4 classes, as described for the study cohorts above. Demographic data (sex, age, academic degrees, and ethnicity) also were collected.

Exam Administration Procedures

Two topic-specific exams and 1 cumulative final exam were administered to each of the 4 classes taking a combined Immunology and Endocrinology course. Students in the 2 collaborative exam cohorts (CEO and CEJ) completed a topic-specific exam individually and handed in their answer sheets. Immediately thereafter, the students were allowed to form into groups of 2 to 3 students and work as a team to complete the same exam. Each student completed a second individual answer form for this second administration of the exam. In addition, students in the CEJ cohort submitted an explanation sheet, selecting from a fixed-choice list of reasons for making answer changes (Table 1). This list was compiled from a survey of students in a previous collaborative learning study by Zhang and Henderson.¹⁴

The 2 topic-specific exams and the cumulative final examination administered to all classes were of the single-best-response, multiple-choice format containing recall and comprehension questions. Exam questions were identical for all study cohorts. The cumulative final examination was administered individually (without collaboration) in all study cohorts.

Data Analysis

Data were initially examined graphically to reveal underlying distribution patterns and identify outliers. We summarized and analyzed our data using SPSS version 22 (IBM, Chicago, IL). Statistical test assumptions were verified, standardized effect sizes were recorded, and

Table 2 - Demographics of Study Subjects ($n = 223$)

Cohort ^a	Control	CEO	CEJ
Student count ^b	61 (100)	82 (100)	80 (100)
Sex			
Male	37 (61)	55 (67)	53 (66)
Female	24 (39)	27 (33)	27 (34)
Academic degree			
Bachelor	55 (90)	80 (98)	73 (91)
Graduate	6 (10)	2 (2)	7 (9)
Age			
<30 y	39 (64)	76 (93)	64 (80)
30 – 40 y	18 (29)	3 (4)	14 (18)
>40 y	4 (7)	3 (4)	2 (2)
Ethnicity/race			
Caucasian	44 (72)	63 (77)	56 (70)
Hispanic	4 (7)	11 (13)	15 (19)
Black	3 (5)	3 (4)	4 (5)
Other	10 (16)	5 (6)	5 (6)

^a Control, traditional, noncollaborative, topic exams; CEO, collaborative topic exams without requiring answer change explanation; CEJ, collaborative topic exams requiring answer change explanation.

^b Values in parentheses are percent of total cohort.

means with 95% confidence levels (CI) were calculated. Study hypotheses were evaluated at the .05 family-wise α level. We applied a Mixed Design ANOVA to evaluate factorial effects on cumulative final exam scores within and across the study cohorts. After ANOVA, planned contrasts identified significant cohort and interaction effects. Significant main effects for the dichotomous “question-type” factor were identified by direct comparison of means.

Collaboration net benefit also was examined. Net benefit data were evaluated for the 2 collaboration study cohorts (CEO, CEJ) during the topic-specific exams. Collaboration net benefit was defined as the total net change in exam answers for the topic-specific exams. If a student changed a question response that was incorrect to one that was correct, that constituted a single positive benefit point (+1). If the change was from correct to incorrect, that constituted a single negative benefit point (–1). Answers that were not changed were assigned 0 benefit points.

RESULTS

Demographic Information

Demographic data are summarized in Table 2. Values within parentheses are percentages. In our sample, sexes were similar in distribution across study cohorts, with a slightly greater percentage of males. Academic degree, age, and ethnicity were markedly skewed within all cohorts in favor of bachelor degrees, age < 30 years, and Caucasians.

Study Cohort Effects (Hypotheses 1 and 2, Table 3, Fig. 2)

There was a statistically significant main effect for study cohorts, $F_{(2, 220)} = 5.29$, $p = .006$, $\eta^2 = .046$. However, contrary to expectation (Hypothesis 1), there was not a statistically significant difference between Control and CEO final exam scores ($p = .566$). By contrast, CEJ final

Table 3 - Cumulative Final Exam Scores [95% CI]

	Control Cohort		CEO Cohort		CEJ Cohort	
Score Total	72.7 [69.56, 75.85]		73.2 [70.42, 75.88]		78.8 [76.40, 81.34]	
	Recall	Comprehension	Recall	Comprehension	Recall	Comprehension
Question Type	78 [75.0, 81.3]	53 [48.6, 57.6]	78 [75.5, 80.9]	55 [51.0, 59.0]	84 [81.7, 86.3]	61 [56.3, 64.6]

exam scores were statistically greater than the Control ($p = .010$) and CEO ($p = .011$) scores (Hypothesis 2).

Question-Type Effects (Hypothesis 3, Table 3, Fig. 2)

Question-type had a marked main effect on final exam scores ($F_{[1, 220]} = 609.77$, $p = .000$, $\eta^2 = .735$). On the final exam, recall question scores were substantially greater than comprehension scores within each of the 2 study cohorts. While recall question scores were essentially the same in the Control and CEO cohorts, they were higher in the CEJ cohort. Similarly, comprehension scores were essentially the same in the Control and CEO cohorts, but were higher in the CEJ cohort. The relative increase in final exam recall and comprehension scores in the CEJ cohort compared to the Control cohort were quite similar. Therefore, contrary to expectation (Hypothesis 3), question-type did not differentially influence study cohort final exam scores ($F_{[1, 220]} = .562$, $p = .571$, $\eta^2 = .005$).

Collaboration Net Benefit

Of students who changed their topic-specific exam answers after group collaboration, 66.6% changed incorrect answers to correct answers (positive benefit points).

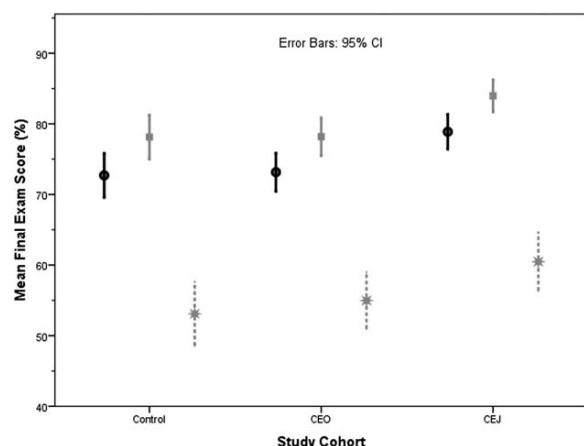


Figure 2 - Study cohort and question-type effects on final exam scores (Hypotheses 1–3). These plots demonstrate a strong main effect for question-type on the noncollaborative final exam ($p = .000$). Recall question scores were substantially greater than comprehension scores. However, the recall and comprehension question-type trends across the study cohorts are very similar, indicating no question-type \times cohort interaction ($p = .562$). Error bars: mean with 95% CIs, are shown for: Total Final Exam Score – heavy black bars with round center elements, Recall Question Final Exam Score – solid gray bars with square center elements, and Comprehension Question Final Exam Score – dashed gray bars with star center elements.

Only 7.4% changed correct answers to incorrect answers after peer discussion (negative benefit points).

In the topic-specific exams, we found statistically significant main effects on collaboration benefit for question-type ($F_{[1,160]} = 915.36$, $p = .000$, $\eta^2 = .85$), and study cohort ($F_{[1,160]} = 8.12$, $p = .005$, $\eta^2 = .048$). Student collaborations produced greater net benefit for comprehension questions compared to recall questions and students in the CEJ cohort benefited more than the CEO cohort. In the CEJ cohort, the top 3 fixed-choice reasons (Table 1) for changing an answer after group collaboration were: Group members backed answers with facts (26%), Guessed first time (14%), and Misread the question the first time (14%). The remaining 12 fixed-choice reasons each constituted less than 5% of all responses, with 4 reasons each representing less than 1% of all responses: Mismatched Scantron the first time, Stress on individual test, Ran out of time, and Too many of same letter answer in a row the first time.

DISCUSSION

We expected increased final exam scores in both collaboration cohorts (CEO and CEJ) when compared to the Control cohort (Hypothesis 1). However, in our study, Control and CEO final exam scores were statistically equivalent ($p = .566$, Table 3, Fig. 2). This finding is consistent with that of Meseke et al.¹⁰ In a study of collaboration testing among chiropractic students, they reported that final exam scores did not differ significantly between collaboration and control cohorts. Other researchers also have reported no improvement in content retention, as measured by subsequent noncollaborative exam scores. In a college large-enrollment introductory biology class, Leight et al.¹⁵ reported higher scores for collaborative testing exams, but no improvement in content retention as measured by cumulative exam questions in a subsequent noncollaborative exam. Similarly, Enz and Frosch¹⁶ reported that collaborative quizzes were favorably perceived by students in a pharmaceutical calculations course and improved course satisfaction, but made no improvement in a noncollaborative mid-term and final exam.

In contrast to our finding of equivalence between the CEO and Control cohorts and the studies cited above, some investigators have reported that collaboration testing significantly improved subsequent exam performance. Rao et al.¹⁷ reported that collaborative testing in a college physiology class enhanced student understanding of the course material. However, the validity of this conclusion may be challenged because their learning assessment exams all included a 20% inclusion of the collaborative testing scores. In a randomized crossover study incorporated into an undergraduate exercise physiology course, Cortright et

al.¹⁸ evaluated the academic impact of collaborative testing using 1-question multiple-choice quizzes administered at 2 to 3 lecture pauses during each 50-minute lecture. They reported that collaborative testing produced increased mastery of original material and improved ability to solve novel problems with the information learned. However, review of their study methods revealed that this improved learning was actually similar to our “collaboration net benefit,” comparing student answers on the same question before and after collaborative testing. There was no reference to noncollaborative examinations evaluating a residual improved learning effect.

It is noteworthy that, in our study, the CEJ study cohort final exam scores were greater than the Control ($p = .010$) and CEO ($p = .011$) cohorts (Table 3, Fig. 2). Our Hypothesis 2 was that justifying answer changes (CEJ) would enhance the improved final exam performance obtained by collaboration alone (CEO). This was not supported by our findings. Rather, the differing final exam outcomes for CEO and CEJ versus Control suggest that requiring justification of answer changes during topical test collaboration was necessary for subsequent improved final exam performance, not simply an enhancement of a CEO cohort effect.

Question-type demonstrated a statistically significant main effect ($p = .000$) with substantially greater scores for recall compared to comprehension questions in the final exam (Table 3, Fig. 2). This is not surprising; final exam scores would be expected to be highest for the least challenging test questions (recall questions). Also, as would be expected, beneficial answer changes during topical exam collaboration was greatest when answering comprehension questions, the more difficult exam question-type ($p = .000$, $\eta^2 = .85$). Students in the 2 collaboration study cohorts obtained a positive topical exam benefit 9 times more frequently than a negative benefit; 66.6% changed answers from incorrect to correct while 7.4% changed answers from correct to incorrect. However, the differential net collaboration benefit during topical exams did not produce a similar differential effect between question-type and study cohort effects in the cumulative final exams ($p = .566$). Recall and comprehension final exam score trends across the study cohorts were statistically equivalent.

Several mechanisms may mediate enhanced performance during collaborative testing. First, collaborative testing may reduce test anxiety by reducing the sense of competition and being part of a group may give a sense of security. Anxiety has been shown to interfere with learning and effective test taking.^{19,20} Second, it has been reported that a collaborative environment encourages students to become active learners, improves learning attitudes, and enhances critical thinking skills and depth of understanding.^{21,22} Lastly, collaborative testing provides an opportunity to discuss reasons for a particular answer, requiring deeper consideration of the course material. In addition to evaluating their own reasoning with regard to test questions, students may identify and fill in knowledge gaps. Chi et al.²³ note that when students explain their reasoning to classmates, “the learner becomes a teacher.” Michael and Modell²⁴ state that “teaching requires generation of explanations, both for oneself and for the learner.” In our study, we examined

whether the putative beneficial effects of 2 collaborative topical exams administered during the academic term would have long-term benefits on academic performance assessed by a noncollaborative, cumulative final exam administered at the end of the term. In addition, we examined whether any observed long-term benefit might be different for students in a CEO cohort, having a collaborative topical exam format, but not requiring answer change justification, versus students in a CEJ cohort, also having a collaborative topical exam format, but requiring justification for any answer changes. Students in the CEJ cohort were required to select a fixed-choice reason when changing an answer (Table 1). The most frequent answers provided by CEJ students suggested that this additional task encouraged additional consideration of the exam questions, rather than simply copying answers from cohort classmates.

CONCLUSION

We concluded that test collaboration with the requirement that students explain their reasons for making answer changes is a more effective learning tool than simple collaboration that does not require answer change justification. Moreover, this effect during collaboration tests is greater for the more challenging comprehension-type questions than simple recall questions. Finally, we found a long-term influence on learning, as reflected by higher noncollaborative, cumulative final exam scores. However, this benefit was observed in the CEJ cohort, not the CEO cohort, suggesting that requiring students during collaborative exams to explain their reasons for making answer changes may be necessary for a long-term academic benefit.

FUNDING AND CONFLICTS OF INTEREST

This work was funded internally. The authors have no conflicts of interest to declare relevant to this work.

About the Authors

Niu Zhang is a professor in the Life Sciences Department at Palmer College of Chiropractic Florida (4777 City Center Parkway, Port Orange, FL, 32129; niu.zhang@palmer.edu). Charles Henderson is a senior research staff member at Life Chiropractic College - West (25001 Industrial Blvd. Hayward, California, 94545. cnrhenderson@gmail.com). Address correspondence to Niu Zhang, 4777 City Center Parkway, Port Orange, FL, 32129; niu.zhang@palmer.edu. This article was received April 13, 2016, revised August 2, 2016, and accepted on November 15, 2016.

Author Contributions

Concept development: NZ, CNRH. Design: NZ, CNRH. Supervision: NZ. Data collection/processing: NZ. Analysis/interpretation: NZ, CNRH. Literature search: NZ. Writing: NZ, CNRH. Critical review: NZ, CNRH.

© 2017 Association of Chiropractic Colleges

REFERENCES

1. Bruffee KA. *Collaborative Learning: Higher Education, Interdependence, and the Authority of Knowledge*, 2nd ed. Baltimore, MD: John Hopkins University Press; 1999.
2. Baumberger-Henry M. Practicing the art of nursing through student-designed continuing case study and cooperative learning. *Nurse Educ*. 2003;28:191–195.
3. Bose MJ, Jarreau PC, Lawrence LW, Snyder P. Using cooperative learning in clinical laboratory science education. *Clin Lab Sci*. 2004;17:12–18.
4. Nolinske T. Resource list for teaching and learning. *Am J Occup Ther*. 1999;53:75–82.
5. Russo A, Warren SH. Collaborative test taking. *College Teaching*. 1999;47:18–20.
6. Hickey BL. Lessons learned from collaborative testing. *Nurse Educ*. 2006;31:88–91.
7. Berry J, Nyman MA. Small-group assessment methods in mathematics. *Int J Mathematical Educ Sci Technol*. 2002;33:641–649.
8. Hoke MM, Robbins LK. The impact of active learning on nursing students' clinical success. *J Holist Nurs*. 2005;23:348–355.
9. Ewald JD. Language-related episodes in an assessment context: a 'small-group quiz'. *Canad Mod Lang Rev*. 2005;61:565–586.
10. Meseke CA, Nafziger R, Meseke JK. Student attitudes, satisfaction, and learning in a collaborative testing environment. *J Chiropr Educ*. 2010;24:19–29.
11. Meseke CA, Bovee ML, Gran DF. Impact of collaborative testing on student performance and satisfaction in a chiropractic science course. *J Manipulative Physiol Ther*. 2009;32:309–314.
12. Woody WD, Woody LK, Bromley S. Anticipated group versus individual examinations: a classroom comparison. *Teach Psychol*. 2008;35:13–17.
13. Mitchell N, Melton S. Collaborative testing: an innovative approach to test taking. *Nurse Educ*. 2003;28:95–97.
14. Zhang N, Henderson CN. Brief, cooperative peer-instruction sessions during lectures enhance student recall and comprehension. *J Chiropr Educ*. 2016;30:87–93.
15. Leight H, Saunders C, Calkins R, Withers M. Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE Life Sci Ed*. 2012;11:392–401.
16. Enz S, Frosch DR. Effect of collaborative vs non-collaborative quizzes on examination scores in a pharmaceutical calculations course. *Am J Pharm Educ*. 2015;79:66.
17. Rao SP, Collins HL, DiCarlo SE. Collaborative testing enhances student learning. *Adv Physiol Educ*. 2002;26:37–41.
18. Cortright RN, Collins HL, DiCarlo SE. Peer instruction enhanced meaningful learning: ability to solve novel problems. *Adv Physiol Educ*. 2005;29:107–111.
19. Chapell MS, Blanding ZB, Silverstein ME, et al. Test anxiety and academic performance in undergraduate and graduate students. *J Educ Psychol*. 2005;97:268–274.
20. Zhang N, Henderson CN. Test anxiety and academic performance in chiropractic students. *J Chiropr Educ*. 2014;28:2–8.
21. Breedlove W, Burkett T, Winfield I. Collaborative Testing and Test Anxiety. *J Scholar Teach Learn*. 2004;4:33–42.
22. Lusk M, Conklin L. Collaborative testing to promote learning. *J Nurs Educ*. 2003;42:121–124.
23. Chi MTH, De Leeuw N, Chiu M-H, Lavancher C. Eliciting self-explanations improves understanding. *Cogn Sci*. 1994;18:438–777.
24. Michael JA, Modell HI. *Active Learning in Secondary and College Science Classrooms : A Working Model for Helping the Learner to Learn*. Mahwah, NJ: L. Erlbaum Associates; 2003.