
Student Attitudes, Satisfaction, and Learning in a Collaborative Testing Environment*

Christopher A. Meseke, PhD, Palmer College of Chiropractic Florida,
Rita Nafziger, MBA, Palmer College of Chiropractic Davenport, and **Jamie K. Meseke**, MSM, Walden University

Introduction: This study examined the effect of collaborative testing on student learning, attitude toward testing, and course satisfaction at a chiropractic college. **Methods:** The study compared testing performance between two cohorts of students taking an advanced neuroanatomy course: a control group ($n = 78$) and an experimental group ($n = 80$). Scores examined for each cohort included sums of quizzes, examination scores, and a comprehensive final examination. The control cohort completed weekly quizzes as individuals, while the experimental cohort completed the quizzes collaboratively in small groups. Both cohorts completed three unit examinations and the comprehensive final examination as individuals. Additionally, pretest-posttest and delayed posttest scores were examined. Multivariate analysis of variance (MANOVA) and multivariate analysis of covariance (MANCOVA) (including repeated measures MANCOVA) were used for statistical analysis. **Results:** The experimental cohort scored significantly higher compared to the control cohort on all quizzes ($F = 217.761$; $df = 1,156$; $p < .05$) and overall course grades ($F = 16.099$; $df = 1,156$; $p < .05$). There were no significant differences in either the comprehensive final (posttest) ($F = 3.138$; $df = 1,122$; $p > .05$) or the delayed posttest (taken 5 weeks after the end of the course) ($F = 0.431$; $df = 1,122$; $p > .05$) between the two cohorts. The overall scores for both cohorts on the delayed posttest were significantly lower than the posttest scores ($F = 4.660$; $df = 1,122$; $p < .05$). **Conclusions:** This project extends previous findings that students using collaborative testing have significantly increased short-term course performance compared with those students using traditional testing. No differences in learning or retention were noted. (*J Chiropr Educ* 2010;24(1):19-29)

Key Indexing Terms: Decision Making; Educational Assessment; Group Process; Teaching

INTRODUCTION

The notion that social interaction and collaboration increases cognitive development and learning in students has been well documented in research literature.¹⁻³ It is also contended that a student's potential for cognitive growth is limited to what he or she may be able to accomplish independently; the synergy of collaboration allows students

to surpass individual limitations.^{1,2} Through peer collaboration, students build on each other's knowledge to develop new attitudes, cognitive skills, and psychomotor skills beyond that which they previously possessed.^{4,5} Collaborative testing—an environment where students work together in small groups during summative assessments—has been shown to have multiple positive impacts on student achievement, including increased positive interdependence, improved personal accountability, and increased opportunities to learn the course material with formative feedback.⁵⁻¹¹

Peer collaboration may also increase problem-solving, critical thinking, and higher reasoning skills and abilities.^{1,5-10,12,13} Conversely, the traditional individualistic paradigm may contribute to a competitive environment that interferes with the learning

* This paper was selected as a 2009 Association of Chiropractic Colleges-Research Agenda Conference Prize Winning Paper. This award is funded by the National Board of Chiropractic Examiners.

process as a whole, especially the development of cooperative skills.¹³⁻¹⁷ Collaborative testing may also help students develop assertiveness and the ability to discriminate among multiple options. Reflective thinking may be strengthened through collaboration since most students appear to learn something new or reinforce their knowledge as a result of this testing style.¹⁸

Collaborative settings may be superior for peer-to-peer transfer of knowledge and previous studies in chiropractic and nurse education programs have reported greater overall student performance in the collaborative testing groups, with the majority of students indicating a preference for the collaborative testing environment.¹⁸ Results from student attitude surveys confirm that collaborative testers have more positive attitudes toward the testing process in general compared to students who take assessments individually.^{11,19-21} It should be noted, however, that when students are exposed to traditional teaching methods but tested in a collaborative manner, they may not perform as well on higher-level theory questions even if performance on lower-level concept questions is improved.^{11,19}

While researching the effects of collaborative testing on test performance without prior collaborative learning, Breedlove and colleagues reported that the effects of collaborative testing were directly related to the level of cognitive processing required by the test question.⁸ In the absence of prior collaborative learning, students reportedly performed better on collaborative tests which incorporated lower-level cognitive processing (ie, knowledge-recall questions such as fill-in-the-blank or matching), while no performance improvement was noted for higher-order, theory-based questions (applications, inferences, conclusions).⁸ Theory-based questions require a higher level of thinking than information-recall questions, and researchers have opined that collaborative testing without prior collaborative learning may not facilitate higher levels of cognitive processing.⁸ However, collaborative learning effectiveness appears to improve over time as students become more familiar with the process.²² Therefore, collaborative testing may be more effective in longer courses or as a standard assessment strategy across a curriculum.

Other studies that examined group dynamics during collaborative testing have reported that group testing appeared to foster deductive reasoning and critical thinking.^{12,23} However, Castor found that students reportedly did not like the challenge of

completing short-answer (ie, higher cognitive) questions collaboratively, because they found it difficult to reach consensus.²² Furthermore, when students engaged in discussion over the short-answer questions, the benefits of the group discussions did not always correlate with the written test answers (students may have benefited from the discussion in the long run but still may have answered the question incorrectly). Students also reported the belief that concept (ie, lower cognitive) questions did not enhance their critical thinking skills, because they were often able to reach consensus quickly, or simply resorted to majority rule barring significant disagreement.²³

Predictably, collaborative testing has been reported to have value in the reduction of test anxiety.^{5,6,8,11,14,15,18,19,24-27} While some degree of angst may be valuable in preparing for a test, the test should not become an anxiety-driven event. Higher levels of anxiety may interfere with optimal learning (and consequently lower grades), which may lead to negative emotional and physical consequences.^{14,28} A decreased ability to recall learned material has also been associated with increased levels of anxiety.⁸ Collaborative testing may be used to change the testing environment, thereby significantly reducing test anxiety. A study examining introductory psychology students involved in collaborative testing reported less anxiety among students both while studying and during testing, increased practice in negotiating differences of opinion, knowledge sharing, and overall enhanced learning.¹⁵ Ultimately, this decreased anxiety may lead to more accurate assessments of students.²³

As previously noted, collaborative testing has also been noted for its value as an educational method of enhancing critical thinking skills and developing a higher-level knowledge of material.^{1,5-10,12-14,18} The collaborative environment encourages students to become active learners and may positively impact students' attitudes regarding the clarity and importance of course material as well as enhance critical thinking skills and depth of understanding.^{5,6,14,15,18,29} Peer collaboration may also play a role in reducing the competitive nature of testing (and the educational experience as a whole) as well as serve as a conduit to improve interpersonal skills and critical thinking.^{5,29,30}

Testing collaboratively apparently benefits both high- and low-performing students. Giuliadori and colleagues studied team testing versus individual

testing among students taking a veterinary physiology course.²¹ The study was designed so that students completed exams individually first, then immediately paired with another classmate to answer the same questions. Using individual test scores, students were sorted into “high performing” or “low performing” categories. As anticipated, group test scores were significantly higher than individual test scores. The collaborative testing effect was particularly large for the overall population and the low performers, but was conversely small for the high performers.²¹ The study also evaluated student satisfaction with testing format and found that a majority of students favored the collaborative format. Overall, the researchers concluded that the collaborative testing paradigm benefits all students and is particularly beneficial for lower-performing students.

In addition to its value as a teaching and assessment method, collaborative testing appears to improve the learner’s understanding of the course material due in part to the immediate formative feedback and peer-to-peer instruction inherent in the collaborative process.^{3,5,6,14,18,29–31} Also, although scant research has been done in the area to date, studies have shown a positive correlation between collaborative testing and improved student retention of course material. Researchers at East Carolina University and Wayne State University studied the effects of collaborative testing on material retention among students enrolled in an exercise physiology class.³¹ The research was prompted by previous studies that indicated student retention of course material is short-lived.³² Results of their study confirmed previous reports that student retention of course material is short-lived in general, but retention appears to be significantly increased when collaborative testing is applied. Similarly, other researchers have evaluated collaborative testing among nursing students. In these studies, both short- and long-term material retention was improved among students who completed assessments collaboratively.^{14,20}

Conversely, researchers at the University of Northern Colorado studied the effects of collaborative testing on material retention among students enrolled in undergraduate psychology courses.³³ Although students reported satisfaction with the collaborative model, and course performance was improved as a result of participating in group testing, retention was neither significantly improved nor significantly decreased among collaborative participants. In a similar study among nursing

students, Lusk and Conklin also found no significant increase in material retention among students who completed assessments collaboratively.⁵ These studies suggest that the benefits of collaborative testing may outweigh the concerns that collaborative testing hampers learning, is “cheating,” or is not a good strategy for student learning.

Although most studies reviewed overwhelmingly applaud the benefits of collaborative learning in general, and collaborative testing in specific, there are notable issues to be considered. Some studies have reported student concern that unprepared peers may be able to earn undeservedly higher exam scores.²⁶ Other disadvantages cited by students include second-guessing oneself, dysfunctional or weak groups, and the possibility of arguments escalating when group consensus cannot be achieved.³⁴

The relative lack of conclusive research in the area suggests opportunity for further study. The current study examines student learning on course-based assessments and postcourse follow-up examinations comparing two cohorts taught with the same materials and methods but assessed differently. The study also investigates student attitudes and satisfaction of the two cohorts.

METHODS

Approved by the institutional review board of Palmer College of Chiropractic, this project used a nonequivalent control group design with two cohorts (cohort one, “experimental,” $n = 80$, enrolled January 2008 to March 2008; and cohort two, “control,” $n = 78$, enrolled April 2008 to June 2008).³⁵ The Doctor of Chiropractic Program (DCP) in-program precourse grade point averages (GPAs) were compared to examine homogeneity between the cohorts. Before the course, a 70-question pretest was administered to all students individually (points were not part of the course point total). Five weeks after the end of the course, a delayed posttest was administered individually to students in both cohorts. The delayed posttest content was identical to both the pretest and comprehensive final examination, but points were not part of the course point total. Students met for 6 hours of lecture per week (60 hours total) during a 10-week academic term. The instructor, lecture format, and material were identical for each cohort. Quiz and examination questions focused on the same content for the

two cohorts, with minor modification in either the stem or the distracters. Course grades were derived from a combination of assessments: six weekly quizzes (15 points each, 34% of total points), three unit examinations (45 points each, 40% of total points), and a comprehensive final examination (70 points, 26% of total points). The scores from the assessments (both quizzes and unit examinations), the mean of the sums of the quiz scores, the mean of the sums of the examination scores, and the comprehensive final examination were compared for the two groups to examine overall differences in cohort performance. The prefinal point totals, comprehensive final exam point totals, and final course grades of the two cohorts were also compared. Multivariate analysis of variance (MANOVA) was used for statistical comparison of the cohorts (except the final examination scores) using SPSS version 17.0 (SPSS Inc, Chicago, IL). The comprehensive final examination and delayed posttest scores of the two cohorts were compared using repeated measures multivariate analysis of covariance (MANCOVA) with pretest scores as the covariate.^{35,36} Additionally, the mean of the sums of quizzes, the mean of the sum of exams, and the prefinal point totals of the two cohorts were compared.

The students in the control cohort completed the weekly quizzes as individuals, while the experimental cohort was randomly assigned into groups of three each week to complete the quizzes collaboratively. Research Randomizer was used for the student group randomizations.³⁷ Randomizations were made prior to distributing the weekly quizzes with students not aware of their group assignments until the time of the quiz.³⁸ Both cohorts were allotted 40 minutes for each quiz. Though allowed to discuss the quiz questions and answers in their groups, each student returned an individual answer form. Both cohorts completed the three unit examinations and the comprehensive final examination as individuals. After the administration of the third unit exam, a survey of student attitudes regarding the specific testing method was administered to all students in each cohort. The survey was scored using a 4-point Likert scale (Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree = 1). MANOVA was used for statistical comparison of the cohort's survey scores. Due to the nature of anonymous surveys, in-program precourse GPA could not be used as a covariate in the analysis of the survey data.

RESULTS

The DCP in-program precourse GPA of the two cohorts did not differ significantly (control mean = 3.29, SD = 0.454; experimental mean = 3.25, SD = 0.489) ($F = 0.261$; $df = 1,155$; $p > .05$). Overall, the experimental cohort differed from the control cohort (Wilks' lambda = 0.196; $F = 56.573$; $df = 10,138$; $p < .01$) based on the in-class assessments (Table 1). All quizzes were significantly higher for the experimental group at $p < .05$. Although unit examination one did not differ between the two cohorts ($F = 2.220$; $df = 1,145$; $p > .05$), the control cohort scored significantly higher on both unit examinations two and three ($F = 8.8441$, $df = 1,147$; $p < .01$ and $F = 19.164$; $df = 1,147$; $p < .01$, respectively) (Table 1).

The reliability indices (Cronbach's alpha) ranged between 0.81 and 0.85 for the pretest, the comprehensive final exam, and the delayed posttest for both cohorts. No significant difference was noted between the experimental group and the control group for the pretest scores ($F = 2.459$; $df = 1,156$; $p > .05$), the final examination scores ($F = 3.138$; $df = 1,122$; $p > .05$), or the delayed posttest scores ($F = 0.431$; $df = 1,122$; $p > .05$) (Table 2). For both cohorts, overall scores for the comprehensive final were higher than the delayed posttest ($F = 4.660$; $df = 1,122$; $p < .05$).

In general, the two cohorts differed in their self-reported attitudes (Wilk's lambda 0.579; $F = 6.862$; $df = 9,85$; $p < .01$) with the experimental group reporting a more positive attitude ($F = 18.884$; $df = 1,93$; $p < .01$) (Table 3). However, on survey item 8, the control cohort reported greater perceived preassessment preparation compared to the experimental cohort (control mean = 3.52; SD = 0.77; experimental mean = 2.84; SD = 0.81; $F = 15.800$; $df = 1,93$; $p < .01$). Likewise, survey items 5 and 9 did not differ between the cohorts ($F = 3.552$; $df = 1,93$; $p > .05$; and $F = 3.366$; $df = 1,93$; $p > .05$, respectively) (Table 3).

DISCUSSION

Student Performance

This project examined the effect of collaborative testing on student performance, retention of course material, and attitudes and satisfaction related to the course testing paradigm (collaborative versus solo).

Table 1. MANCOVA results for grades in an advanced neuroanatomy course

Dependent Variable	Control	Experimental	<i>F</i> Statistic	Degrees of Freedom	Significance
	Mean (SD)	Mean (SD)			
Quiz 1	9.92 (2.48)	12.84 (1.808)	59.241	1,141	.000
Quiz 2	10.68 (1.805)	13.70 (1.588)	114.837	1,141	.000
Quiz 3	10.50 (2.813)	13.54 (1.211)	64.987	1,141	.000
Quiz 4	10.36 (2.006)	14.00 (1.201)	161.945	1,141	.000
Quiz 5	10.99 (1.933)	13.61 (1.392)	92.054	1,141	.000
Quiz 6	13.01 (1.651)	14.98 (0.157)	91.557	1,141	.000
Exam 1	28.78 (5.903)	27.76 (4.863)	2.334	1,141	.129
Exam 2	32.49 (5.111)	30.14 (4.597)	8.451	1,141	.004
Exam 3	40.51 (2.167)	38.90 (2.840)	22.155	1,141	.000
Sum of quizzes	64.12 10.130	82.33 4.328	217.761	1,156	.000
Sum of exams	102.03 10.367	96.84 10.155	10.098	1,156	.000
Prefinal point total	166.14 (18.797)	179.16 (11.428)	27.839	1,156	.000
Final point total	223.91 (23.904)	237.37 (17.903)	16.099	1,156	.000
Final grade (percent)	0.845 (0.090)	0.900 (0.068)	16.099	1,156	.000

Precourse in-program GPA used as covariate (Wilks' lambda = 0.196; $F = 56.573$; $df = 10,138$; $p < .01$). Final summative (posttest) scores are found in Table 2.

Table 2. MANCOVA results for posttest and delayed posttest scores using pretest scores as the covariate

Dependent Variable	Control Mean (SD)	Experimental	<i>F</i> Statistic	Degrees of Freedom	Significance
		Mean (SD)			
Pretest	20.55 (5.58)	19.06 (6.33)	2.459	1,156	.119
Posttest	57.77 (10.74)	58.94 (10.79)	3.138	1,122	.079
Delayed posttest (5 weeks)	51.72 (8.31)	50.49 (7.83)	0.431	1,122	.513

Posttest scores were significantly higher than the delayed posttest scores ($F = 4.660$; $df = 1,122$; $p < .05$). No significant differences were noted between the control and experimental groups on both the posttest and the delayed posttest.

Table 3. Survey items examining student attitudes related to collaborative testing (experimental group) as compared to individualistic testing (control group)

Survey Item	Experimental Group Mean (SD)	Control Group Mean (SD)	F statistic	Degrees of Freedom	p Value
1. The written exams were based on class presentations and/or reading assignments.	3.42 (0.587)	2.72 (0.696)	27.795	1,93	.000
2. The written tests were fair.	3.21 (0.600)	2.54 (0.719)	23.432	1,93	.000
3. The written administration helped to reduce test stress/anxiety.	3.33 (0.837)	2.31 (0.836)	34.095	1,93	.000
4. The results of the written tests reflect my knowledge of the subject matter.	3.00 (0.816)	2.47 (0.959)	9.500	1,93	.003
5. There was sufficient time for completion of the written exams.	3.77 (0.427)	3.60 (0.528)	3.552	1,93	.063
6. The written tests increased my critical thinking skills.	3.37 (0.655)	2.98 (0.656)	6.984	1,93	.010
7. The written tests increased my confidence in my judgments.	2.98 (0.859)	2.28 (0.840)	16.304	1,93	.000
8. I studied harder than I normally would for the written tests.	2.84 (0.814)	3.53 (0.774)	15.800	1,93	.000
9. I was exposed to new ideas concerning topic matter as a result of the test.	3.28 (0.630)	3.02 (0.805)	3.366	1,93	.070
Overall Survey	29.19 (4.311)	24.95 (4.372)	18.602	1,93	.000

Wilk's lambda = 0.579; $F = 6.862$; $df = 9,85$; $p < .001$. The survey was scored using a Likert-type scale (Strongly Agree = 4, Agree = 3, Disagree = 2, Strongly Disagree = 1).

As reported in the authors' previous studies, quiz scores for the collaborative cohort were again substantially higher than those for the solo cohort.^{9,10} Although the authors previously posited that collaborative testing may better prepare students to complete unit examinations as individuals, results of this study do not concur.^{9,10} Paradoxically, the control cohort of the current project scored significantly higher on the second and third unit examinations. The significant increases in the control cohort examination performance may indicate a reverse Hawthorne effect.³⁹ The Hawthorne effect describes increased task performance based on perceived environmental change: in this case, a situation in which an experimental group is aware of their status and, as a result, strives for greater performance. A reverse Hawthorne effect is a situation in which the control group changes their behavior based on their status as a control group. This confounding factor is rooted in the assumption that because the control group may "know" they are

the control group in a study, they may try harder, perhaps out of mere rivalry, to achieve similar or greater results than the experimental group.^{40,41} In the current project, the control group may have been aware of the primary author's previous research in collaborative testing. Not only was there a significant difference in student's opinion of test preparation (item 8: $F = 15.800$; $df = 1,93$; $p < .01$), but the control group reported a greater amount of study (Table 2). It may be inferred the control group worked harder to raise (or maintain) their grades.^{40,41}

The experimental cohort had significantly more points prior to the final examination (74% of the course total points) ($F = 21.040$; $df = 1,141$; $p < .01$), resulting in overall course grades of the experimental group that were significantly higher (Table 1). However, there was no difference between the two cohorts on the final exam ($F = 0.706$; $df = 1,145$; $p > .05$). Previous studies have shown similar results relative to quiz scores without significant differences in final examination

scores.^{9,10,41,42} Both Giraud and Meseke and associates reported the final examination did not significantly differ between the experimental and control cohorts.^{9,10,42} These inconsistent significant differences between the two cohorts regarding the quizzes and the unit examinations relative to the subsequent final examination may be related to the length of the final examination, higher stress level for the perceived difficulty of a “final,” total course points prior to final examination, or length of the academic term.^{10,11} Researchers have also opined that as students’ overall scores increased during the term based on the collaborative testing, there may have been a tendency for students in the collaborative testing groups to prepare less for the final and study for other courses instead.^{10,42}

In this project, the cohorts were members of two separate classrooms, as similar as availability permitted. Due to this nonequivalent design, the groups could not be assumed to be demographically equivalent and MANCOVA was used for analysis of the pretest/final exam/delayed posttest data.^{35,36} Precourse GPA was used as a covariate for analysis of all course assessments, with the exception of the final examination scores. The final examination scores were compared between the two cohorts using pretest scores as the covariate.^{35,36} Lack of significant differences between the two cohorts in both precourse GPAs and pretest scores controlled for the occurrence of unwanted regression effects.^{35,43}

The second threat to internal validity is referred to as selection maturation. The higher grades on the quizzes for the experimental cohort may be explained as a type of growth process occurring primarily in the experimental group.^{35,36} While both cohorts participated in the same assessments, cognitive development may have occurred at a greater rate in the experimental cohort due in part to the collaborative experience. Nowak and colleagues reported that team testing effectiveness increased over time as group members matured in the collaborative process.⁴⁴ It has also been reported that a learning curve for collaborative testing may exist such that an instructor may be required to instruct students in the function of collaborative testing.²² In the present study, this question of the learning curve was addressed by exposing the experimental cohort to the process of collaborative testing in the prior academic term. While a maturation process (learning curve) may have occurred in students within the experimental cohort leading to increased quiz scores, this in no way should be considered an actual threat

to internal validity.^{9,10,22,44} It is logical to assume that the performance of the experimental group on the collaborative quizzes and overall grades may be due in part to a change in the student’s learning process, even in light of the nonsignificant differences between the experimental and control cohort’s overall course performance (grade). Because both the final examination and delayed posttest scores were significantly higher compared to the pretest scores overall, it may be assumed that student learning has occurred. The fact that the experimental cohort scores on both the comprehensive final and the delayed posttest were not significantly lower than those of the control cohort indicates the value of this interactive assessment technique.^{10,44} Although learning may be an elusive concept to quantify, several aspects of learning may be examined based on the pretest, the comprehensive final exam, and the delayed post-test. When compared to the pretest, the control and experimental groups showed an increase in the posttest scores of 281% and 309%, respectively (Table 2). There was a 251% increase in knowledge in the control group and 264% in the experimental group when the delayed posttest scores were compared to the pretest.

Knowledge Retention

Although it may be argued that giving the same pretest, final examination, and delayed posttest may lead to “memorizing the test,” a legitimate counterpoint is that memorization is part of Bloom’s taxonomy (knowledge) and memorization is a legitimate mode of learning.^{13,45} Most assessment items in this course were written in terms of the “knowledge” or “comprehension” levels of Bloom’s taxonomy. Higher-level questions often can be used as learning tools in themselves. For instance, if students are challenged to analyze information rather than merely memorizing it, there may be greater retention of information.⁴⁶ It is the potential of collaborative testing to improve critical thinking that is most promising.

The delayed posttest scores of both the collaborative and control groups were significantly lower than the respective comprehensive final exam scores (Table 2). As with the comprehensive final exam scores, the delayed posttest scores did not differ between the two cohorts. Even 5 weeks after the course’s conclusion, knowledge loss based on the delayed posttest and the comprehensive final

scores for the control and experimental cohorts was only 10.47% and 14.34%, respectively. The problem of knowledge loss (knowledge decrement) in health science students has previously been documented.^{1,47-52} However, there are scant data regarding the effects of collaborative testing. As in the current study, Tucker reported no difference in retention of course material among nursing students in a collaborative paradigm, whereas Cortright reports significant improvement in retention.^{30,48} It is currently unknown what effect collaborative testing has on long-term retention in the chiropractic setting or how changes in retention translate to performance on licensing exams.

Student Satisfaction

When student comments on the survey were examined several themes became apparent. Students who participated in collaborative testing were largely positive about the experience. Several of the comments from the survey reflect the use of collaborative testing as an active learning tool. Comments such as “*Excellent tool. Effective for me as a ‘multiple exposure’ to the material,*” “*I learned many key points from others during the tests,*” “*I enjoyed the group quizzes because you could talk through questions and it was a good learning experience,*” and “*The group quizzes were great because you would teach each other, and I would always take notes while going through the questions.*” During the quizzes the authors noted a great deal of peer-to-peer teaching within the collaborative group structure. This may represent an additional aspect of involvement of the student toward learning even during the test.⁵³ It is apparent from these comments that the students of the experimental cohort felt the collaborative experience to be a positive one that helped increase their learning and understanding of the course material. These student comments directly relate to survey item six (“*the written tests increased my critical thinking skills*”) (experimental mean = 3.37, SD = 0.655; control mean = 2.98, SD = 0.656; $F = 6.984$; $df = 1,93$; $p = .01$). Several comments from the experimental group noted the value of discussion/debate in the learning process. These comments echo Vygotsky’s concept of the zone of proximal development.^{1,2} Full cognitive development requires social interaction which, in students, leads to increased learning and accomplishment.^{1,2}

It has previously been noted that some individuals tend to reduce their personal input in collaborative efforts while reaping the benefits of the group interaction, a phenomenon known as “social loafing,” “free-riding,” or “freeloading.”^{43,54,55} Although this behavior may be seen as detrimental to the group dynamic and counter to the collaborative paradigm, it may be advantageous to the low-achieving student (ie, receive input from others without reciprocating). It has been speculated, however, that the parasitic student, without full participation, may not learn as deeply as participatory members.^{17,56} Interestingly, this arrangement is not thought to be detrimental to high-achieving students, because they will nonetheless benefit from the discussion and group feedback.^{17,21,27,54} The issue of “freeloading” is thought not to be problematic in smaller-sized groups (three to four persons) as in the present study.⁵⁵ While students in the collaborative groups may become dependent on others’ preparation for the assessment (ie, “freeload”), most comments from the experimental group state a level of peer pressure that promoted studying and participation. Student comments such as, “*Studied so that I wouldn’t look dumb,*” “*Because we took group quizzes I wanted to be prepared to participate and therefore studied a bit more regularly than I do for other subjects,*” and “*They require you to study the material more consistently, so putting studying off is really not an option.*” Only one comment from the experimental cohort refers to freeloaders: “*Weekly testing concept—itself—very good due to being “forced” to keep-up with lecture content. Great idea—but felt at times—not all group members participated and were helped by given correct answers when they would have normally failed a quiz as such.*” Although the group may have received little input from a freeloading member, it cannot be stated that the freeloader only benefited from the quiz answers; they may have also learned from the other members. It is interesting to note that on all quizzes the experimental group not only scored higher than the control group, but they also had small standard deviations as compared to the control group. This was also shown in previous reports by Meseke and associates.^{9,10} This may be related to the group process in which the students tend to be more homogenous in thought, whereas the control group showed more variation owing to the heterogeneity of the cohort. This also may be related to the issue of weak students working with stronger students on the assessments. Because

the mean quiz grades were higher for the experimental group, it may indicate that the stronger students bolstered the weaker students without harm to themselves. This aspect of variance of the means between the cohorts suggests further consideration. The concepts of positive interdependence and individual accountability that are reflected in these statements may also indicate that the students may not have wanted to let their group down or be perceived as unprepared for the assessments.

Test Anxiety

As previously mentioned, researchers have noted collaborative testing may be a valid tool addressing the problem of test anxiety.^{5,6,8,11,14,15,18,19,24-27} As in the aforementioned studies, the experimental cohort of the current project differed from the control cohort. In fact, the collaborative testing group reported a decrease in test anxiety.^{11,19} But test anxiety may be also related to the previously described reverse Hawthorne effect. It is possible that members of the control cohort felt additional anxiety related to their performance during the assessments and the proceeding study periods. Yet others have reported that there may be no difference in levels of test anxiety between students involved in collaborative testing and solo testing.^{16,57} Based on these data, it may only be stated that test anxiety is not increased within the collaborative paradigm.

Based on the positive comments of the experimental group, it is interesting to note that the control cohort differed significantly from the experimental group on survey item 8 (*I studied harder than I normally would for written exams*) (experimental mean = 2.84, SD = 0.814; control mean = 3.53, SD = 0.774; $F = 15.80$; $df = 1,93$; $p < .01$). As previously suggested, this difference may be related to a reverse Hawthorne effect. Alternately, students who knew they would be in a collaborative group, where they would have the opportunity to discuss and teach each other, may have chosen to focus their studies on other courses. However, the student survey comments from the experimental cohort do not seem to support this hypothesis. Further, the current study indicates that students from the experimental group studied no less than their control group peers.¹¹ In addition, previous studies by Bovee and associates noted that there were no significant differences in the amount of assessment preparation between solo-testing students compared with those participating in collaborative

testing.^{11,19} Future studies may include an analysis of student assessment preparation, perhaps by total hours studied and by type of studying (solo, groups, etc.).

In contrast, students who took the quizzes as individuals had relatively negative remarks, many with a common theme of exam content. Student comments such as, “*I think you should test us more on what you tell us in class,*” “*The test material was sometimes not covered in the notes,*” “*Test material should be presented in class and should be presented earlier (ie, allowing more time to study the material,*” and “*Tests did not reflect material taught in class.*” The common theme to these comments is based on survey items 1 and 2, in both of which the experimental cohort expressed more positive attitudes as compared to the control cohort (survey item 1: $F = 27.795$; $df = 1,93$; $p < .01$; and survey item 2: $F = 23.432$; $df = 1,93$; $p < .01$) (Table 3). Further, the experimental cohort found the assessment items to be more representative of the studied course material (survey item 1: $F = 27.795$; $df = 1,93$, $p < .001$; and survey item 4: $F = 9.500$; $df = 1,93$; $p < .01$) Of note, although both cohorts took identical exams (minor alterations in either the stem or distracters), only the control cohort viewed the examinations negatively. Meseke and associates previously reported an attitude of “unfairness” from students involved with solo testing compared to the students involved in collaborative testing.¹¹ The students in that project did not refer to perceived differences in testing difficulty. The attitude survey data combined with the assessment data may illustrate a potential relationship between educational process, student performance, and satisfaction.

Although there were no significant differences in precourse GPA, when GPA was used as a covariate in MANCOVA, the experimental group nonetheless scored significantly higher on the quizzes and final course grades. Additionally, by examining the satisfaction survey results, it may be argued that the collaborative process might lead to higher performance, and in turn, to greater satisfaction. Alternately, it may be argued that in light of no significant differences in the final examination scores, the collaborative process did not impact student performance, but rather the higher attitude survey scores of the experimental group may be related to simply having a different course assessment style. In a previous project, Meseke and associates reported that although precourse GPA did not differ between control and experimental groups, the experimental

group scored significantly higher on quizzes and unit examinations.^{9,10} In either possibility, there is an increase in student satisfaction that may lead to an improved classroom dynamic, supporting an improved educational environment.

Although the comprehensive final exam and the delayed posttest scores did not significantly differ, the experimental cohort nonetheless had significantly higher grades and reported an increased positive attitude as compared to the control group. While it is difficult to quantify learning, collaborative testing does seem to improve the classroom environment and encourage collegiality among students in addition to improving student course performance. Statistically significant differences between the pretest and the comprehensive final exam scores, as well as between the pretest and the delayed posttest, do appear to indicate that learning has occurred and that students in the experimental group learned just as much as their control group counterparts.

CONCLUSION

This study in the chiropractic environment extends previous reports related to collaborative testing and learning. While students involved in collaborative testing achieved significantly higher grades as compared to peers involved in solo testing, no significant differences were noted between the cohorts on either the comprehensive final examination or the delayed posttest. Results from survey items indicate that students involved with collaborative testing studied no more than they would have normally and convey higher satisfaction with the content of the assessment items. Students involved in collaborative testing show increased overall course performance, better testing attitudes, and equal learning (ie, less grade variance on the quizzes and significant differences between the pretests and the posttests) as compared to students involved in traditional solo testing. Future studies may consider an analysis of assessment item depth (ie, Bloom's taxonomy) and assessment item reliability as well as investigate issues of knowledge loss, long-term learning, and student satisfaction.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

About the Authors

Christopher Meseke is the Assessment Coordinator, Department of Academic Affairs, Palmer College of Chiropractic Florida. Rita Nafziger is the Director for the Center for Teaching and Learning, Palmer College of Chiropractic. Jamie K. Meseke is with Walden University. Address correspondence to Christopher A. Meseke, Academic Affairs, Palmer College of Chiropractic, Florida Campus, 4777 City Center Parkway, Port Orange, FL 32129 (e-mail: christopher.meseke@palmer.edu). This article was received June 1, 2009, revised September 14, 2009, and accepted September 19, 2009.

REFERENCES

1. Doolittle PE. Vygotsky's zone of proximal development as a theoretical foundation for cooperation learning. *J Excell Coll Teach* 1997;8(1):83-103.
2. Vygotsky LS. Thinking and speech. In: Vygotsky LS (ed). *Collected works*, vol. 1. New York: Plenum; 1987, p. 211.
3. Adams P. Exploring social constructivism: theories and practicalities. *Education* 3-13 2006;34(3):243-57.
4. Damon W, Phelps E. Critical distinctions among three methods of peer education. *Int J Educ Res* 1989;13:9-19.
5. Lusk M, Conklin L. Collaborative testing to promote learning. *J Nursing Educ* 2003;42(3):121-24.
6. Slusser SR, Erickson RJ. Group quizzes: an extension of the collaborative learning process. *Teach Sociol* 2006;34(3):249-62.
7. Russo A, Warren S. Collaborative test taking. *Coll Teach* 1999;47(1):18.
8. Breedlove W, Burkett T, Winfield I. Collaborative testing and test performance. *Acad Exch Q* 2004;8(3):36-40.
9. Meseke JK, Nafziger R, Meseke CA. Facilitating the learning process: a pilot study of collaborative testing versus individualistic testing in the chiropractic college setting. *J Manipulative Physiol Ther* 2008;31(4):308-12.
10. Meseke CA, Nafziger R, Meseke JK. Student course performance and collaborative testing: a prospective follow-on study. *J Manipulative Physiol Ther* 2008; 31(8):611-15.
11. Meseke CA, Bovee ML, Gran DF. The impact of collaborative testing on student performance and satisfaction in a chiropractic science course. *J Manipulative Physiol Ther* 2009;32:309-14.
12. Johnson DW, Johnson RT, Smith KA. Cooperative learning returns to college: what evidence is there that it works? *Change* 1998;20(4):27-35.
13. Allen MJ. *Assessing academic programs in higher education*. Boston: Anker Publishing; 2004, pp. 3, 34-37.
14. Phillips AP. Reducing nursing students' anxiety level and increasing retention of materials. *J Nursing Educ* 1988;27(1):35-41.
15. Ligeikis-Clayton C. Shared test taking. *J NY Nurses Assoc* 1996;27(4):4-6.
16. Wiggs CM, Bohmfalk DW. Collaborative testing: promoting critical thinking in BSN students. *Commun Nurs Res* 2006;39:349.
17. Webb NM. Assessing students in collaborative groups. *Theory Pract* 1997;36:205-13.
18. Pray Muir S, Tracy DM. Collaborative essay testing. *Coll Teach* 1999;47(1):33-35.

19. Bovee M, Gran D. Effects of collaborative testing on student satisfaction survey. *J Chiropr Educ* 2005; 19(1):47.
20. Durrant LK, Pierson G, Allen EM. Group testing and its effectiveness in learning selected nursing concepts. *J R Soc Health* 1985;105(3):107–11.
21. Giuliadori MJ, Lujan HL, DiCarlo SE. Collaborative group testing benefits high- and low-performing students. *Advan Physiol Educ* 2008;32:274–8.
22. Castor T. Making student thinking visible by examining discussion during group testing. *New Dir Teach Learn* 2004;100:95–99.
23. Gokhale AA. Collaborative learning enhances critical thinking. *J Technol Ed* 1995;7(1):22–30.
24. Helmericks SG. Collaborative testing in social statistics: toward Gemeinstat. *Teach Sociol* 1993;21:287–97.
25. Hancock DR. Exploring the effects of group testing on graduate students' motivation and achievement. *Assess Eval High Educ* 2007;32(2):215–27.
26. Mitchell N, Melton S. Collaborative testing: an innovative approach to test taking. *Nurse Educ* 2003;28(2): 95–97.
27. Zimbardo P, Butler L, Wolfe V. Cooperative college examinations: more gain, less pain when students share information and grades. *J Exp Educ* 2003;71(2): 101–26.
28. Chapell MS, Blanding ZB, Silverstein ME, et al. Test anxiety and academic performance in undergraduate and graduate students. *J Educ Psychol* 2005;97(2): 268–74.
29. National Research Council. How people learn: brain, mind, experience, and school. Expanded edition. Bransford JD, Brown AL, Cocking RR (eds). Washington, DC: National Academy Press; 2000, pp. 279–80.
30. Cortright RN, Collins HL, Rodenbaugh DW, DiCarlo SE. Peer instruction enhanced meaningful learning: ability to solve novel problems. *Adv Physiol Educ* 2004;29:107–11.
31. Cortright RN, Collins HL, Rodenbaugh DW, DiCarlo SE. Student retention of course content is improved by collaborative-group testing. *Adv Physiol Educ* 2003;27: 102–8.
32. Richardson DR. Comparison of naive and experienced students of elementary physiology on performance in an advanced course. *Adv Physiol Educ* 2003;23:S91–95.
33. Woody WD, Woody LK, Bromley S. Anticipated group versus individual examinations: a classroom comparison. *Teach Psychol* 2008;35:13–17.
34. Hickey BL. Lessons learned from collaborative testing. *Nurse Educ* 2003;31(2):88–91.
35. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin; 1966, pp. 40, 46–50.
36. Dimitrov DM, Rumrill PD. Pretest-posttest designs and measurement of change. *Work* 2003;20:159–65.
37. Urbanick G, Plous S. Research Randomizer [document on the Internet]. © 2008 [cited 2009 Sept 2]. Available from <http://www.randomizer.org/form.htm>.
38. Klecker BM. Formative classroom assessment using cooperative groups: Vygotsky and random assignment. *J Instr Psychol* 2003;30(3):216–9.
39. Draper SW. The Hawthorne, Pygmalion, placebo, and other effects of expectation: some notes. [document of the Internet] © 2008 May 11 [cited 2009 Sept 2]. Available from: <http://www.psy.gla.ac.uk/~steve/hawth.html>.
40. Zdep SM, Irvine SH. A reverse Hawthorne effect in educational evaluation. *J School Psychol* 1970;8:89–95.
41. Lynch B. Cooperative learning in interdisciplinary education for the allied health professions. *J Allied Health* 1983;13(2):83–93.
42. Giraud, G. Cooperative learning and statistics instruction. *J Stat Educ* [serial on the Internet]. © 1997 [cited 2009 Sept 2]. 5(3) [about 11 pages]. Available from: <http://www.amstat.org/publications/jse/v5n3/giraud.html>.
43. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand McNally College Publishing; 1979, pp. 51–56.
44. Nowak L, Miller S, Washburn J. Team testing increases performance. *J Educ Bus* 1996;71(5):257.
45. Bloom B, Englehart M, Furst E, Hill W, Krathwohl D. Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain. New York: Longmans, Green; 1956, pp. 201–7.
46. Andre T. Does answering higher-level questions while reading facilitate productive learning? *Rev Educ Res* 1979;49(2):280–318.
47. Conway MA, Cohen G, Stanhope N. Very long-term memory for knowledge acquired at school and university. *Appl Cogn Psychol* 1992;6:467–82.
48. Tucker DA. A study of the effects of collaborative and dual testing methods on retention of course material among baccalaureate nursing students (doctoral dissertation). Memphis State University; 1993.
49. Harrison A. Using knowledge decrement to compare medical students' long-term retention of self-study reading and lecture materials. *Assess Eval High Educ* 1995;20(2):149–60.
50. D'Eon MF. Knowledge loss of medical students on first year basic science courses at the University of Saskatchewan. *BMC Med Educ* 2006;6:5.
51. Mateen FJ, D'Eon MF. Neuroanatomy: a single institution study of knowledge loss. *Med Teach* 2008;30(5): 537–9.
52. Woloschuk W, Mandin H, Harasym P, Lorscheider F, Brant R. Retention of basic science knowledge: a comparison between body system-based and clinical presentation curricula. *Teach Learn Med* 2004;16(2): 116–22.
53. Lord T, Baviskar S. Moving students from information recitation to information understanding: exploiting Bloom's taxonomy in creating science questions. *J Coll Sci Teach* 2007;36(5):40–44.
54. Simkin MG. An experimental study of the effectiveness of collaborative testing in an entry-level computer programming class. *J Inf Syst Educ* 2005;16(3): 273–80.
55. North AC, Linley PA, Hargreaves DJ. Social loafing in a cooperative classroom task. *Educ Psychol* 2000;20(4): 389–92.
56. Webb NM. Group collaboration in assessment: multiple objectives, processes, and outcomes. *Educ Eval Policy Anal* 1995;17(2):2239–61.
57. Hanshaw, LG. Test anxiety, self-concept, and the test performance of students paired for testing and the same students working alone. *Sci Educ* 1982;66(1):15–24.