American Journal of Pharmaceutical Education **2015; 79 (5) Article 66.**

RESEARCH

Effect of Collaborative vs Noncollaborative Quizzes on Examination Scores in a Pharmaceutical Calculations Course

Stephanie Enz, PharmD, Donald R. Frosch, MS

Butler University College of Pharmacy and Health Sciences, Indianapolis, Indiana

Submitted July 23, 2014; accepted December 8, 2014; published June 25, 2015.

Objective. To determine if midterm and final examination scores in a pharmaceutical calculations course differ among students who take weekly quizzes collaboratively, noncollaboratively, or half-collaboratively/half-noncollaboratively ("mixed").
Methods. One hundred twenty-three students enrolled in one of 4 laboratory sections: collaborative, noncollaborative, or mixed sections. Students working noncollaboratively completed a 15-minute quiz at the end of weekly laboratories. Students working collaboratively were randomly allocated into groups of 4 and worked independently for 5 minutes before working collaboratively for 10 minutes. Students in mixed sections worked collaboratively during one half of the semester and noncollaboratively during the other half of the semester in a crossover design to control for order effect. All students took midterm and final examinations independently.
Results. Mean scores for a pretest did not significantly differ between the 4 laboratory sections at the study's onset. Multivariate analysis of variance (MANOVA) revealed a significant multivariate effect for the 9 laboratory quizzes in relation to group assignment. Mean scores on the midterm examination did not significantly differ between collaborative and noncollaborative groups. On the final examination, the two mixed groups significantly outperformed the collaborative group, but did not significantly differ from one another or from the noncollaborative group.
Conclusion. Peer collaboration improves quiz scores, is favorably perceived by students and enhances their course satisfaction, but does not improve subsequent performance on midterm and final examinations taken noncollaboratively.

Keywords: collaborative, testing, cooperative

INTRODUCTION

Collaboration between pharmacists and other health care professionals is essential to improved patient outcomes,[1-4] reduction in medication errors,[5,6] and a better understanding of the pharmacist's role within the health care team.[7,8] The Center for the Advancement of Pharmacy Education (CAPE) Educational Outcome 3.4 states that all entry-level graduates from doctor of pharmacy (PharmD) programs should be prepared to collaborate with other health care professionals and in doing so demonstrate "mutual respect, understanding, and values to meet patient care needs."[9] In order for pharmacy students to develop effective collaborative skills, they must be given ample opportunities to collaborate within the classroom.

One methodology for integrating collaboration into a pharmacy curriculum is allowing students to take quizzes or examinations with one or more fellow students. Models for collaborative testing are summarized in Table 1. Instructors who decide to implement a collaborative testing model must consider how many students will be in each group, whether students will be able to choose their own groups or be assigned selectively or randomly, how many opportunities during the semester will groups have to work together, whether students will work independently before working collaboratively or only collaboratively during the examination, how much time students will have to work independently and/or collaboratively, whether a consensus answer will be required, and finally, how students will be graded for their independent and/or collaborative work.

There are two approaches to evaluating the benefit of collaborative testing. The first is to assess score improvement among students who work collaboratively compared to those who work independently. Most researchers found that students who worked collaboratively on a quiz or examination outscored those who did not.[10-20] Only Haberyan and Barnett saw no significant difference

Corresponding Author: Stephanie Enz, PharmD, 4600 Sunset Avenue, Indianapolis, IN 46208. Tel: 317-940-9551. E-mail: senz@butler.edu

Table 1. Comparison of Collaborative Testing Methods

Author	Group Size	Group Formation	Collaboration Required	Group Consensus Required	No. Examinations or Quizzes	Method	Time for Collaboration	Scoring
Bloom 2009	3-4	Self-selected; students could change groups each examination	Yes	No	4 examinations	Independent then collaborative	<20 minutes	75% of independent score + 25% of collaborative score
Cortright et al. 2003	2-3	?	No	?	2 examinations	Independent then collaborative on subset of questions only	No information	No information
Giuliodori et al. 2008	2	Self-selected	No	Yes	3 examinations	Independent then collaborative	30 minutes	Only used independent score
Haberyan and Barnett 2010	2	Self-selected	No	Yes	1 or 2 examinations	Collaborative only	Entire testing period	Both students received same score for examination
Leight et al. 2012	2-4	Self-selected	No	Yes	1 examination	Independent then collaborative	Equal to independent testing period	Increase in individual score of 5, 3, or 1% based on collaborative score
Lusk and Conklin 2003	2-3	Randomly assigned	Yes	No	3 examinations	Independent then collaborative	20 minutes	-
Meseke et al. 2008	2-3	Randomly assigned	Yes	No	6 quizzes	Collaborative only	-	Students scored based on individual answers
Mitchell and Melton 2003	2-3	Randomly assigned	No	No	1 examination	Independent then collaborative	10 minutes	Both answers used for a combined score
Rao et al.2002	2-3	Assigned by instructor	Yes	Yes	4 quizzes	Independent then collaborative	30 minutes	80% of independent score + 20% of collaborative score
Slusser and Erickson 2006	6-7	Assigned based on GPA	Yes	No	8 quizzes	Independent then collaborative	2 minutes	Score from one quiz randomly selected (all scores recorded)
Woody et al. 2008	3-4	Self-selected	No	No	2 examinations	Collaborative only	Entire testing period	Students scored based on individual answers
Zimbardo et al. 2003	2	Self-selected	No	Yes	2 examinations	Collaborative only	Entire testing period	Both students in pair received same score

between students who worked in pairs on an examination and those who worked independently.[21] The other approach to evaluating this type of testing is to ascertain if students experience a long-term benefit from the collaborative testing model—in other words, to determine if students who work collaboratively gain a better understanding than students who work independently of material during a quiz or an examination as evidenced by performance on a subsequent examination.

Effective peer instruction is a key component during collaboration. Crouch and Mazur noted that effective peer instruction engaged students through activities, requiring them to apply core concepts presented and then explain those concepts to fellow students. [22] Researchers who addressed this hypothesis had mixed findings as to whether collaborative testing resulted in improved learning and higher test scores on subsequent examinations. Both Bloom and Cortright and colleagues found that students who worked collaboratively during an examination scored significantly higher when retested on the same material at a later date.[10,11] Conversely, several studies showed no significant difference in subsequent examination scores between students who worked collaboratively and those who worked independently.[13-15,18,20]

Researchers who surveyed students reported a positive student reaction to collaborative testing. From the students' perspective, collaborative testing led to increased understanding of the material.[11,13] These findings however, must be tempered by the findings of Gurung and colleagues, who highlighted the disconnect between student perception of and reality about their own learning.[23] Interestingly, students in the Slusser and Erickson study had a positive view on how collaborative testing affected their course grade even when the end result was not a higher grade.[20] Students also reported less stress and anxiety when taking a collaborative examination vs a traditional examination.[13,21]

With no "tried and true" method for the collaborative testing model, we set out to carefully construct a model to benefit students by allowing them to support and teach each other on calculations-based quizzes in the laboratory setting of a pharmaceutical calculations course. Based on recommendations in the literature, groups size was kept small, groups were randomly assigned, membership in each group was kept the same throughout the semester, and each team was required to submit a consensus answer.[24-26] According to Cohen, the size of a collaborative group should be kept small to encourage each member of a group to participate.[24] Johnson and colleagues agreed that a group should be made up of 2-4 members in order to achieve meaningful face-to-face interaction.[25] Of the research on collaborative testing we reviewed, almost all reported using groups within the range of 2-4 members.[10-19,21] Oakley and colleagues recommended the instructor assign the collaborative groups instead of allowing students to self-select because "left to their own devices," stronger students tend to seek one another out, leaving weaker ones to fend for themselves.[26] Of the studies in which students were assigned to groups, the methods and timing of randomization varied. Lusk and Conklin, Meseke and colleagues and Mitchell and Melton randomly assigned students to a dyad or group just prior to each quiz or examination.[14-16] Only Slusser and Erickson selectively assigned groups during the second week of the semester based on grade point averages (GPAs) in order to attain balanced levels of academic achievement in each group.[20] In an attempt to control for extraneous variables and to develop what Slusser and Erickson referred to as "a sense of group identity, trust, support, and accurate communication,"[20] we chose to randomly assign collaborative groups. Because Slusser and Erickson mentioned the difficulty in achieving this connection when the group only had 1 or 2 opportunities to work collaboratively, we chose to assign collaborative groups at the beginning of the semester and keep them the same throughout the semester. One challenge of operationalizing a collaborative testing format is that, unless there is a reason or incentive for the members to interact, students may choose to work independently.[24] To prevent this occurrence, we required a consensus answer from the group.

Most studies we reviewed seemed to concur that allowing students time to work independently before working collaboratively gave them an opportunity to assess their understanding of the material and form their own conclusions prior to collaborating with their peers.[10-14,16,17,20] For this reason, we decided to provide students with 5 minutes of independent work before collaborating with group members.

We did not locate published studies on collaborative testing in a pharmacy program or on quizzes or examinations that are strictly calculations-based. The primary objective of this study was to determine if mean scores on the calculations portion of the midterm and final examinations differed between students who took low-stakes weekly quizzes throughout the semester, working either collaboratively in small groups, noncollaboratively (independently), or half collaboratively and half noncollaboratively ("mixed"). A secondary objective was to determine how collaborative quizzes would be perceived by students. We hypothesized that calculation-based weekly quizzes taken collaboratively would result in increased scores on the calculations portion of the midterm and final examination as compared to quizzes taken independently. We further hypothesized that collaborative quizzes would be favorably perceived by students exposed to this method.

METHODS

The study was a single-blinded, nonrandomized, controlled trial comparing mean scores on the pharmaceutical calculations portions of the midterm and final examinations of students enrolled in a semester-long pharmaceutical care laboratory course. The course had a common lecture section with 4 different laboratory sections. Depending on their assigned laboratory section, students took weekly laboratory quizzes all collaboratively (C), all noncollaboratively (NC), collaboratively for the first half of the semester and noncollaboratively for the second half of the semester (C/NC), or noncollaboratively for the first half of the semester and collaboratively for the second half of the semester (NC/C). Each laboratory section met once weekly for 3 hours, and was held 11 times during the semester. The midterm was administered during the sixth laboratory, the final examination was administered during the eleventh laboratory, and weekly quizzes were administered during the remaining 9 laboratory meetings. The main focus of the course was teaching pharmaceutical calculations. The laboratory component utilized real-world scenarios as a method of applying calculations learned in lecture. Regular laboratory sessions consisted of 15 minutes of prelaboratory instruction, approximately 2.5 hours of activities during which students worked collaboratively with others in their assigned group, and 15 minutes to take a quiz.

Students eligible for the study were in their first professional year (P1) at the Butler University College of Pharmacy (N=123), who enrolled in Introduction to Pharmaceutical Care-2, a required course offered during the spring 2013 semester. Each student self-enrolled in 1 of 4 laboratory sections, which differed according to how laboratory quizzes were taken. Enrollment was on a first-come, first-served basis according to each student's registration window, as designated by the registrar's office. When enrolling, students did not know which laboratory sections were collaborative, noncollaborative, or mixed. Students remained in the same assigned section throughout the semester. Students who enrolled in the 3 laboratory sections with a collaborative component were randomly assigned to collaborative groups, each consisting of 4 students, using the random number generator, meRandom v1.0 (oPhone app, copyright 2009 by John Fullington).

During the first week of class, students were informed they were going to be participants in an ongoing study during the semester, and the collaborative testing format was explained in all laboratory sections. To allay potential concerns, students were told that a grade adjustment for quiz scores would be made at the end of the semester if quiz scores were found to appreciably differ among the 4 laboratory sections. Although given the option to opt out of the study, all 123 students elected to participate. Because laboratory sections were scheduled on different days of the week, students were asked to and signed an academic integrity pledge stating, "At no point will I discuss the content of the quizzes with any other student outside of my laboratory section." Expedited approval for this study was granted by the Butler University Institutional Review Board (IRB).

To determine if the nonrandom process of self-enrollment in the sections resulted in comparison groups similar in ability to perform pharmaceutical calculations at the onset of the study, a 10-question pretest (1 point per item) was administered to all students during the first week of classes, prior to the first laboratory. The pretest consisted of entry-level pharmaceutical calculations from the Introduction to Pharmaceutical Care-1 course taught the previous semester. An example pretest question was, "A 25 kilogram patient is to receive a drug dosed as 25 mg/kg/day, divided TID (three times a day). The concentration of the medication in an oral solution is 5% w/v (weight/volume). How many milliliters should this patient receive per dose?"

Each week at the conclusion of the laboratory session, students were given a calculations-based quiz focused on material presented in prelaboratory assignments and reinforced during that week's session. The quiz consisted of 1 or 2 open-ended, case-based questions requiring application of learned material. For example, one quiz question asked students to determine the flow rate of an intravenous infusion for a 75 kg patient receiving 9 grams of calcium gluconate in 500 milliliters of $D_{5\%}W$ at a rate of 0.1 mEq/kg/h (milliequivalents/kilogram/hour). Quizzes were closed-book and closed-notes. Each quiz was worth 10 points. Nine weekly quizzes were administered, 5 during the first half of the semester and 4 during the second half of the semester. Students assigned to work noncollaboratively (ie, independently) were given 15 minutes to complete the quiz and submit it to the instructor. Students assigned to work collaboratively were given 5 minutes to work independently on the quiz before collaboration within groups began. Each collaborative group then worked for 10 minutes to achieve a consensus response to each question. Each collaborative group was given a single answer sheet to submit its answers to the quiz question(s). Because all students within their designated collaborative group received the same grade, each member of the group was asked to sign the answer sheet indicating agreement with the group's answers. A student disagreeing with the group's final consensus answers could choose not to sign the group answer sheet and instead submit independent answers.

Students took 2 laboratory-based examinations during the semester—a midterm and final (cumulative) examination. The midterm and final examinations each contained 10 calculations-based questions that varied in point value according to difficulty level. Although each examination consisted of additional questions, this study was only concerned with scores earned on the pharmaceutical calculations portions of the midterm and final examinations, which were worth up to 30 points and 25 points, respectively. Examination scores were reported as the total number of points earned. Every student took the midterm and final examination noncollaboratively.

A newly developed 10-item online anonymous survey was administered at the end of the semester to any student who worked either all collaboratively or half-collaboratively to assess attitudes toward collaborative quiz administration. Subject responses were based on a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree).

The 4 laboratory sections were taught by 2 instructors. Instructor 1 taught the collaborative and noncollaborative sections; instructor 2 taught the mixed sections. Course content was identical for each section. A third instructor taught the common lecture section. To control for grading bias, all examinations were graded by instructor 1 after instructor 2 covered student names with an opaque sticker and replaced the names with code numbers. After the examinations were graded by instructor 1, they were returned to instructor 2, who replaced the number codes with student names and entered the grades into a spreadsheet accessible only to her. At no time was instructor 1 permitted access to the master data sheet that linked student names with their respective code numbers.

A priori sample size estimations were performed for the *t* test and 1-factor analysis of variance (ANOVA) analyses. To detect a small-to-medium effect size (f) of 0.30 by ANOVA on the pretest, laboratory quizzes, and final examination, an a priori sample size determination revealed 31 students were needed per comparison group (.05 alpha, df_b=3) to achieve a minimum power goal of 80%. To detect a small-to-medium effect size of 0.30 by *t* test analysis of the midterm examination, an a priori sample size determination revealed 175 students were needed per comparison group (.05 alpha, 2-tailed) to achieve a minimum power goal of 80%. Because we were limited to the approximately 60 students in the 2 comparison groups of the midterm examination, we would need to measure a post hoc effect size index (d) greater than 0.5 to achieve our minimum power goal. Post hoc power determinations were performed for all tests in which a meaningful effect size was measured.

The statistical program SPSS v21 (IBM Corp., Armonk, NY) was employed for data analysis. All tests were nondirectional, using a .05 alpha level of significance. Bivariate Pearson *r* correlations were used to evaluate the relationship between the pretest and all quizzes (combined), the midterm examination, and the final examination. Independent *t* tests were used to evaluate mean differences between 2 unrelated groups, and independent group ANOVA *F* tests were used to evaluate mean differences among 3 or more unrelated groups. The Bonferroni correction was employed to limit the amount of alpha error that occurs when performing multiple analyses. Whenever a significant *F* value was found, an eta-square (η^2) value was reported to gauge the size of the detected overall effect size (where, according to Cohen, 0.01=small, 0.09=medium, and 0.25=large).[26] Brown-Forsythe ANOVA tests were performed in place of standard ANOVA tests whenever the Levene statistic indicated the assumption for the homogeneity of variances was violated. In the latter case, the Tamhane multiple comparison test replaced the Tukey's Honestly Significant Difference (HSD) test when performing pairwise comparisons to determine which group mean differences contributed to a significant *F* value. Whenever a significant pairwise comparisons was found, a standardized effect size (Cohen's d) was calculated (where, according to Cohen, 0.20=small, 0.50=medium, and 0.80=large).[26]

To evaluate performance on the weekly laboratory quizzes, an initial MANOVA was performed, using laboratory group assignment as the independent variable and the 9 laboratory quizzes as the (combined) dependent variable. Thereafter, univariate ANOVA was performed on each of the 9 quizzes. In instances where the assumption for the homogeneity of variances was violated, both Brown-Forsythe ANOVAs and Kruskal-Wallis H tests were performed to evaluate differences between laboratory groups.

Two-variable chi-square tests of independence (2x2 cross-tabulations) were performed to explore relationships between the mixed-collaborative and solely-collaborative groups, and Likert scale responses to each question (dichotomously categorized as: disagree, strongly disagree, or neutral; and agree or strongly agree). Chi-square corrections using Fisher exact tests were performed whenever any cell in a cross tabulation had an f_e value less than 5.

RESULTS

To help establish the validity of the pretest, Pearson *r* bivariate correlations were performed on the pretest and all quizzes (combined), the midterm examination, and the final examination. These correlations were 0.19, 0.18, and 0.28, respectively. All correlations were significant at $p < 0.05$.

Mean scores for the pharmaceutical calculations pretest did not significantly differ among the noncollaborative ($\overline{X} = 8.9$, $s = 1.7$), noncollaborative/collaborative ($\overline{X} = 8.4$, $s = 1.8$), collaborative/noncollaborative ($\overline{X} = 8.6$, $s = 1.9$), and collaborative ($\overline{X} = 8.1$, $s = 1.7$) groups at the study's onset, according to a 1-factor independent groups ANOVA ($F(3,119) = 1.032$, MSE=3.200, $p = 0.38$; Table 2). The calculated effect size of $f = 0.16$ on the pretest was not a meaningful difference to detect.

Only 2 students, on one occasion each, submitted independent answers because they disagreed with their group's answers. To evaluate performance on the weekly lab quizzes, an initial MANOVA was performed, using group assignment as the independent variable and the 9 quizzes as a combined dependent variable. The MANOVA revealed a significant multivariate effect for the quizzes in relation to group assignment (Wilks' Lambda=0.38, $F_{(3,324.82)} = 4.79$, $p = 0.00$). Univariate analysis for the effect of group assignment showed significant differences between groups for all quizzes except 5 and 8. In each of the 7 significant quiz analyses, however, the Levene statistic was significant at the $p < 0.05$ level, indicating a violation of the assumption of the homogeneity of variances. Consequently, Brown-Forsythe ANOVAs were reported in place of conventional ANOVA results. Brown-Forsythe ANOVAs confirmed the findings of the MANOVA for quizzes 2, 3, 4, and 9, but were unable to be performed for quizzes 1, 6, and 7 because at least one group had no variance (Table 2). For this reason, nonparametric statistics (Kruskal-Wallis *H* tests) were used to either confirm or deny the findings obtained from the univariate ANOVAs (Table 2). Kruskal-Wallis tests revealed significant findings for 6 of the 7 quiz analyses: quiz 1 ($H(n=123) = 29.75$, $p = 0.00$), quiz 2 ($H(n=123) = 17.60$, $p = 0.00$), quiz 4 ($H(n=123) = 20.76$, $p = 0.00$), quiz 6 ($H(n=123) = 11.06$, $p = 0.01$), quiz 7 ($H(n=123) = 13.19$, $p = 0.00$), and quiz 9 ($H(n=123) = 20.75$, $p = 0.00$). On quiz 3, significant results were found on univariate ANOVAs but not on the Kruskal-Wallis H test ($H(n=123) = 5.64$, $p = 0.13$).

Because of the increased likelihood of making a type-I error by running 9 separate univariate quiz analyses, only Kruskal-Wallis results significant at the $p < .0056$ level (using the Bonferroni correction) were reported. Only quizzes 1, 2, 4, 7, and 9 reached this criterion of significance. In these quizzes, the highest scores were noted in those groups exposed to collaborative quiz taking during part or all of the semester. For each of these quizzes, Tamhane multiple comparison tests were used to determine which pairwise group comparisons contributed to the overall significant univariate ANOVAs and are summarized in Table 2. The calculated $f = 0.51$ on the 9 quizzes, when summed, was 0.51, a meaningful difference that yielded approximately 99% power.

Table 2. Noncollaborative (NC), Mixed (NC/C and C/NC), and Collaborative (C) Group Comparisons

	NC (n=31) Mean (SD)	NC/C (n=32) Mean (SD)	C/NC (n=32) Mean (SD)	C (n=28) Mean (SD)	ANOVA *p* value	Kruskal-Wallis *p* value
Pretest Quizzes	8.9 (1.7)	8.4 (1.8)	8.6 (1.9)	8.1 (1.7)	0.38†	
Quiz 1	9.2^{a} (1.3)	8.7^{b,c} (1.5)	9.9^{c} (0.7)	10.0^{a,b} (0.0)	—‡	0.00*
Quiz 2	9.1 (1.3)	8.5^{a} (1.8)	9.7 (1.8)	9.6^{a} (1.1)	0.01‡	0.00*
Quiz 3	9.3 (1.3)	8.3^{a} (2.1)	9.3 (1.4)	9.6^{a} (0.7)	0.01‡	0.13
Quiz 4	9.5 (1.1)	8.4^{a,b} (2.0)	9.9^{a} (0.4)	9.8^{b} (0.6)	0.00‡	0.00*
Quiz 5	9.9 (0.4)	9.8 (0.6)	10.0 (0.0)	10.0 (0.0)	—‡	0.27
Quiz 6	9.6 (1.1)	10.0 (0.0)	9.3 (1.6)	10.0 (0.0)	—‡	0.01
Quiz 7	9.2^{a,b} (1.6)	10.0^{b} (0.0)	9.5 (1.3)	10.0^{a} (0.0)	—‡	0.00*
Quiz 8	8.8 (1.6)	8.9 (1.5)	8.9 (1.5)	9.5 (1.2)	0.27‡	0.27
Quiz 9	8.4^{a} (1.6)	9.8^{a,b} (0.6)	8.7^{b} (1.3)	9.1 (1.4)	0.00‡	0.00*
Examinations						
Midterm	26.4 (2.8)	— (—)	— (—)	26.4 (3.0)	0.98	
Final	20.2 (2.9)	21.4^{a} (2.3)	21.0^{b} (3.1)	18.2^{a,b} (4.1)	0.00‡	0.00

Pretest and quiz scores reported as points earned out of 10 possible points; midterm scores reported as points earned out of 30 possible points; final examination scores reported as points earned out of 25 possible points.

†1-factor independent groups ANOVA with $\alpha = .05$, 2-tailed

‡Brown-Forsythe ANOVA with $\alpha = .05$, 2-tailed

*significant, after applying the Bonferroni correction

a,b,c significantly different by Tamhane's multiple comparison test (within a given row, mean values containing the same superscript letter significantly differed from one another)

Before comparing mean group performances on the midterm examination, the noncollaborative and noncollaborative/collaborative groups were combined to reflect students solely exposed to noncollaborative quiz administration during the first half of the semester. Likewise, the collaborative and collaborative/noncollaborative groups were combined to reflect those solely exposed to collaborative quiz administration during the first half of the semester. Mean scores on the midterm did not significantly differ between noncollaborative group ($\overline{X}=26.4$, $s=2.8$) and the collaborative group ($\overline{X}=26.4$, $s=3.0$) according to an independent *t* test ($t(121)=0.03$, $p=0.98$, Table 2).

Mean scores on the final examination significantly differed among the noncollaborative ($\overline{X}=20.2$, $s=2.9$), noncollaborative/collaborative ($\overline{X}=21.4$, $s=2.3$), collaborative/noncollaborative ($\overline{X}=21.0$, $s=3.1$), and collaborative ($\overline{X}=18.2$, $s=4.1$) groups, according to a 1-factor independent groups Brown-Forsythe ANOVA ($F(3,93.86)=5.93, p=0.00, \eta^2=0.13$, a medium effect). This was confirmed with a Kruskal-Wallis *H* test ($H(n=123)=13.339$, $p=0.00$). The highest scores were observed in the 2 mixed groups, and the lowest scores observed in the collaborative group. Tamhane post hoc multiple comparison tests revealed this was a result of significant group mean differences between the noncollaborative/collaborative and collaborative groups ($d=1.02$, a very large effect), and between the collaborative/noncollaborative and collaborative groups ($d=0.91$, a large effect size). When the 2 mixed groups were combined to control for order effect, mean scores on the final examination significantly differed among the noncollaborative ($\overline{X}=20.2$, $s=2.9$), mixed ($\overline{X}=21.2$, $s=2.7$), and collaborative ($\overline{X}=18.2$, $s=4.1$) groups, according to a 1-factor independent groups Brown-Forsythe ANOVA ($F(2,66.24)=7.76$, $p=0.00$, $\eta^2=0.13$, a medium effect). This was confirmed with a Kruskal-Wallis *H* test ($H(n=123)=13.252$, $p=0.00$). Tamhane post hoc multiple comparison tests revealed this was a result of significant group mean differences between the mixed and collaborative groups ($d=0.97$, a large effect size, Table 2). The calculated $f=0.39$ on the final examination was a meaningful effect size to detect and yielded approximately 96% power.

All students in the 3 groups that experienced either full or partial collaborative quiz administration were asked to complete an end-course survey regarding their extent of agreement with positive statements about their collaborative experiences (Table 3). Thirteen of 28 (46.4%) students in the collaborative group, 18 of 32 (56.3%) students in the noncollaborative/collaborative groups, and 19 of 32 (59.4%) students in the collaborative/noncollaborative group completed the survey, yielding an overall survey response rate of 50 of 92 (54.3%). In

Table 3. Results of the Survey Assessing Attitudes of Mixed-Collaborative and Solely Collaborative Quiz Takers (%)

Survey Statement	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
Interaction with my collaborative group increased my confidence related to the quiz material.	0	0	6	42	52
The size of my collaborative group was appropriate for a group discussion.	0	0	0	26	74
After an initial adjustment period, every member of my group consistently contributed to the weekly discussions.	2	12	14	34	38
By helping other students understand quiz material, I achieved a better understanding of the material as well.	0	4	4	37	55
Hearing other students in my group explain quiz material gave me a better understanding of the material.	0	0	16	36	42
Being a member of a collaborative quiz group made me feel more accountable for knowing and learning the material so I could contribute to group discussions.	0	6	16	36	42
My group worked well together to reach a final consensus answer on quiz questions.	0	0	6	34	60
The collaborative quizzes reduced my stress about taking the midterm and/or the final.	2	12	36	22	28
The collaborative testing method was a positive experience for me.	0	0	4	36	60
I would like to use the collaborative testing method again in the future.	0	2	2	26	70

9 out of 10 survey items, more than 70% of Likert responses to every item were either "agree" or "strongly agree." On the one remaining survey item, "The collaborative quizzes reduced my stress about taking the midterm and/or the final," 14% either disagreed or strongly disagreed, 36% were neutral, and 50% either agreed or strongly agreed. It is also notable that the item "After an initial adjustment period, every member of my group consistently contributed to the weekly discussions," had the same combined percentage of disagree or strongly disagree as the statement regarding the stress about taking examinations.

Two-variable chi-square tests of independence (2x2 cross-tabulations) with Fisher exact test corrections were performed to explore relationships between the mixed and solely collaborative groups and Likert scale responses to each question (dichotomously categorized as: disagree, strongly disagree, or neutral; and agree or strongly agree). Analysis of item 2 was not possible because there were no responses of "disagree," "strongly disagree," or "neutral." Even before applying the Bonferroni correction, none of the remaining 9 items reached significance, suggesting that group assignment and Likert scale responses, as categorized above, were independent (unrelated). Independent *t* tests were also performed to explore mean differences in Likert ratings among group assignment (dichotomously categorized as "mixed," or collaborative) and Likert scale average responses to each survey question. After applying the Bonferroni correction, no significant group mean differences were observed, also indicating that survey responses did not significantly differ between students exposed to solely collaborative quiz taking and those exposed to mixed quiz taking.

DISCUSSION

We set out to create an environment in a pharmaceutical care laboratory in which students would support, challenge, and teach each other during low-stakes, calculations-based quizzes. The goal was for students to gain a better understanding of the quiz material from working collaboratively, which would improve their high-stakes midterm and final examination calculations scores more than if they had taken the quizzes independently.

On a week-to-week basis, students in collaborative groups outscored students who took quizzes independently. These findings are consistent with most other studies of this nature.[10-20] However, our hypothesis that collaboration during weekly quizzes would result in increased midterm and final examination calculation scores was incorrect. On the midterm examination, mean scores were no different between students who had only worked collaboratively up until that point (collaborative and collaborative/noncollaborative) and students who had only taken quizzes independently (noncollaborative and noncollaborative/collaborative). Insufficient time may have had passed for group differences to emerge. On the final examination however, both mixed groups significantly outscored the collaborative group. Students who worked solely collaboratively scored the lowest of the 4 groups.

Even though other researchers had results similar to ours,[13-15,18,20] we hypothesized improved outcomes with collaboration as a result of choices we made when designing our collaborative testing method. We incorporated various methods from the literature. For example, we set group sizes at 4 students. This allowed the face-to-face interaction recommended by Johnson and colleagues,[25] which encouraged and facilitated participation by all group members. With larger groups of 6-7 students, like those used by Slusser and Erickson,[20] students may have found it easier to observe only. Unlike Leight et al and Woody et al, who allowed students to self-select their collaborators,[13,18] we randomly assigned groups. By doing this, we heeded the recommendation of Oakley and colleagues to not permit students to self-select their groups,[26] which minimized the possibility of students forming academically homogenous groups that could compromise collaboration, while also controlling for extraneous variables. Lusk and Conklin and Meseke and colleagues assigned students to different groups for each examination and did not require them to come to a consensus answer,[14,15] both of which were factors we felt were counterproductive to effective collaboration. Despite incorporating these methodologies, there was no improvement on subsequent examination scores in the collaborative groups.

Unlike previous researchers, our study included a priori sample size estimations and post hoc power analysis. The percent of achieved power on the combined quizzes and final examinations was 99% and 96%, respectively. We also reported η^2 values and Cohen *d* effect sizes for the significant ANOVA findings associated with the final examination, all of which ranged from medium to large effect sizes.

One limitation of our study was our inability to randomly allocate students into the 4 laboratory sections because of registration protocols. This could potentially have resulted in stronger students, with honors status or higher number of transfer credits, enrolling in laboratories with better perceived days, times, or instructors. However, our pretest demonstrated that the 4 comparison groups did not differ in their ability to perform pharmaceutical calculations.

According to Cohen, requiring students to come to a consensus answer prevents students from opting out of working with the other members of their group.[24] However, using only the collaborative answer to score all members of the group may have led to grade inflation and lack of accountability. Students who were struggling with the material were able to score well on the quiz by relying on the knowledge and competency of other group members. For these students, consistently high quiz scores may have led to an elevated sense of confidence regarding comprehension of the material, a decreased likelihood of the student seeking help from a peer or instructor, and ultimately a lower examination score. Using the collaborative method on high-stakes examinations could potentially lead to further grade inflation and overestimation of students' abilities, which in turn, could mislead professors, administrators, and even future employers. To incorporate more individual accountability within the group, scoring systems similar to ones used by Bloom and Rao and colleagues could have been used.[10,17] Both studies weighted independent answers much higher than the collaborative answer (75/25 and 80/20, respectively).

From our findings, it appears that working collaboratively for only half the semester is more beneficial than working collaboratively for the entire semester. On the final examination, our mixed groups (that worked collaboratively for only half of the semester) significantly outperformed the fully collaborative group. One possible explanation for this finding is that weaker students had not yet experience sufficient collaboration to become overly dependent on their collaborative group, yet they still benefited from the experience. Another possible explanation is that a single instructor taught the 2 mixed laboratories. However, because new material was only taught in the common lecture section, this is unlikely a contributing factor.

Both Bloom and Cortright et al had results in direct contrast to ours and the other researchers cited in this article.[10,11] Both found collaborative test takers scored significantly higher when retested as compared to students who worked independently. One explanation for this difference is that the questions on the examinations in the other 2 studies were identical to questions posed to them previously, whereas our questions were only similar in scope of material. In addition, Bloom's examinations were more knowledge-based (true/false, multiple-choice, and matching), which only required recall of information from the previous examination. The quizzes and examinations in our course required application of math concepts to solve pharmacy practice-based problems.

Student perception of collaborative testing was positive in many of the published studies.[11-18,20,21] Responses from students in this study who worked collaboratively for a portion of the semester or the entire semester were equally positive, with over 70% of students who worked collaboratively agreeing or strongly agreeing to almost every item on the survey except one. Like students in the Slusser and Erickson study,[20] they positively regarded the collaborative process despite the fact that the end result was not necessarily positive. Similar to findings in the Gurung and colleagues study,[23] there was a disconnect between their perceptions of the effectiveness of collaborative quiz taking and their ability to perform well on a high-stakes examination. The one survey item for which responses differed from the rest asked whether collaborative quizzes reduced their stress about taking the midterm and/or final examinations. Only 50% agreed or strongly agreed, 14% disagreed or strongly disagreed and the remaining responses were neutral. Potentially, students who disagreed or strongly disagreed with this statement were overly reliant on other group members during the quizzes—leaving them feeling less prepared for the high-stakes examinations, which increased their stress levels. Those overly reliant students may have caused 14% of students to disagree or strongly disagree with the statement "every member of my group consistently contributed to the weekly discussions."

CONCLUSION

The primary objective of this study was to determine if working collaboratively on low-stakes calculations-based quizzes would improve performance on the calculations portion of the high-stakes midterm and final examinations. Even after incorporating methodological suggestions from previous researchers, we found no difference between collaborative and noncollaborative groups on the midterm. On the final examination, the mixed groups each significantly outperformed the collaborative group. Therefore, our findings do not support using a collaborative quiz model throughout a semester, but do support a collaborative model for a portion of the semester. A secondary objective was to determine how students would perceive the collaborative experience. Even though students in the collaborative group scored lowest on the final examination, they still rated the collaborative experience as positively as students in the mixed collaborative groups.

ACKNOWLEDGMENT

The authors would like to thank Margaret Stratford, PharmD, for her assistance with data management.

REFERENCES

1. Carter BL, Ardery G, Dawson JD, et al. Physician and pharmacist collaboration to improve blood pressure control. *Arch Intern Med.* 2009;169(21):1996-2002.

2. Ramser KL, Sprabery LR, George CM, et al. Physician-pharmacist collaboration in the management of patients with diabetes resistant to usual care. *Diabetes Spectrum*. 2008;21(3):209-213.
3. Santschi V, Chiolero A, Burnand B, Colosimo AL, Paradis G. Impact of pharmacist care in the management of cardiovascular disease risk factors. *Arch Intern Med*. 2011;171(16):1441-1453.
4. Smith M, Bates DW, Bodenheimer T, Cleary PD. Why pharmacists belong in the medical home. *Health Aff (Millwood).* 2010;29(5):906-913.
5. Sweeney MA. Physician-pharmacist collaboration: a millennial paradigm to reduce medication errors. *J Am Osteopath Assoc.* 2002; 102(12):678-681.
6. Setter SM, Corbett CF, Neumiller JJ, Gates BJ, Sclar DA, Sonnett TE. Effectiveness of a pharmacist-nurse intervention on resolving medication discrepancies for patients transitioning from hospital to home health care. *Am J Health Syst Pharm.* 2009;66(22):2027-2031.
7. Kozminski M, Busby R, McGivney MS, Klatt PM, Hackett SR, Merenstein JH. Pharmacist integration into the medical home: qualitative analysis. *J Am Pharm Assoc.* 2011;51(2):173-183.
8. Kucukarslan S, Lai S, Dong Y, Al-Bassam N, Kim K. Physician beliefs and attitudes toward collaboration with community pharmacists. *Res Social Adm Pharm.* 2011;7(3):224-232.
9. The Center for the Advancement of Pharmacy Education. Educational Outcomes 2013. http://www.aacp.org/resources/education/cape/Open%20Access%20Documents/CAPEoutcomes2013.pdf. Accessed September 23, 2014.
10. Bloom D. Collaborative test taking: benefits for learning and retention. *Coll Teach*. 2009;57(4):216-220.
11. Cortright R, Collins H, Rodenbaugh D, DiCarlo S. Student retention of course content is improved by collaborative-group testing. *Adv Physiol Educ*. 2003;27(1-4):102-108.
12. Giuliodori M, Lujan H, DiCarlo S. Collaborative group testing benefits high- and low-performing students. *Adv Physiol Educ*. 2008;32(4):274-278.
13. 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 Educ*. 2012;11(4):392-401.
14. Lusk M, Conklin L. Collaborative testing to promote learning. *J Nurs Educ.* 2003;42(3):121-124.
15. 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):308-312.
16. Mitchell N, Melton S. Collaborative testing: an innovative approach to test taking. *Nurse Educ.* 2003;28(2):95-97.
17. Rao SP, Collins HL, DiCarlo SE. Collaborative testing enhances student learning. *Adv Physiol Educ.* 2002;26(1-4):37-41.
18. Woody WD, Woody LK, Bromley S. Anticipated group versus individual examinations: a classroom comparison. *Teach Psychol.* 2008;35(1):13-17.
19. Zimbardo P, Butler L, Wolfe V. Cooperative college examinations: more gain, less pain. *J Exp Educ*. 2003;71(2):101-125.
20. Slusser SR, Erickson RJ. Group quizzes: an extension of the collaborative learning process. *Teach Sociol*. 2006;34:249-262.
21. Haberyan A, Barnett J. Collaborative testing and achievement: are two heads really better than one? *J Instr Psychol.* 2010;37(1): 32-41.
22. Crouch CH, Mazur E. Peer instruction: ten years of experience and results. *Am J Phys.* 2001;69(9):970-977.
23. Gurung RAR, Daniel DB, Landrum RE. A multisite study of learning in introductory psychology courses. *Teach Psychol*. 2012;39(3):170-175.
24. Cohen EG. Restructuring the classroom: conditions for productive small groups. *Rev Educ Res.* 1994;64(1):1-35.
25. Johnson DW, Johnson RT, Smith KA. Cooperative learning returns to college what evidence is there that it works? *Change.* 1998;30(4):26-35. Cohen J. *Statistical Power Analysis for the Behavioral Sciences,* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
26. Oakley B, Felder RM, Brent R, Elhajj I. Turning student groups into effective teams. *J Student Centered Learn.* 2004;2(1):9-34.