

## Varying the Preparation Guide and Group Discussion in a Classroom Analysis of Interteaching

**Catherine M. Gayman**

Troy University  
[cgayman@troy.edu](mailto:cgayman@troy.edu)

**Stephanie T. Jimenez**

University of Pittsburgh at Johnstown  
[Sts97@pitt.edu](mailto:Sts97@pitt.edu)

*Abstract: Interteaching is a strategy that shifts the emphasis from passive student learning to active engagement through the use of preparation guides, small group discussions, clarifying lectures, and frequent testing. Several classroom studies have demonstrated that interteaching leads to better student comprehension and higher test scores. However, the specific strategies used in these studies vary slightly. The goal of the present study was to compare two different ways of implementing the preparation guide and group discussions to determine which method led to higher academic success. A group design was used in two sections of a psychology course over two semesters. One section experienced the standard interteaching method, where students completed the entire preparation guide prior to class and engaged in small group discussions during class. The second section was divided into two groups and each group was given half of the preparation guide to complete. Students, then, went through two rounds of group discussions: first, in a dyad with a member that completed the same portion of the preparation guide and then in a larger group with another dyad who completed the other portion of the preparation guide. Students in the second section scored more points on exam questions that came from their half of the preparation guide and they demonstrated less of a preference for interteaching than those who experienced the standard interteaching method. Results from this study indicate that instructors should have students read and complete the entire preparation guide to allow for more effective implementation of interteaching.*

*Keywords: interteaching, preparation guide, higher education*

Interteaching is an evidence-based behavioral teaching method that continues to accrue empirical evidence supporting its effectiveness (e.g. Arntzen & Hoium, 2010; Saville, Zinn, & Elliot, 2005; Saville, Zinn, Neef, Van Norman, & Ferreri, 2006). Studies on interteaching in college courses have shown that its use leads to higher student academic achievement, more positive student course evaluations, greater student participation, and higher student and instructor enjoyment compared to standard lecture-centered teaching (see Saville, Lambert, & Robertson, 2011b; Sturmey, Dalfen, & Fienup, 2015 for reviews).

Interteaching was first described by Boyce and Hineline (2002), who developed and outlined the components of this behavioral teaching method. Here, students first read the chapter and complete questions covering the material prior to class. These questions, called preparation guides (prep guides), ensure that students have read the chapter and pulled out the most important definitions and concepts prior to attending class. Prep guides also encourage students to synthesize information from the reading and apply the concepts to relevant situations (Saville, Cox, O'Brien, & Vanderveldt, 2011a). The majority of in-class time is then spent with students engaging in the interteaching session. This involves students discussing the prep guide questions in groups of two to four, while the

instructor migrates between groups to facilitate the discussion and answer any questions students have over the material.

After students complete the interteaching session, they fill out a record sheet where they rate the quality of the discussion and indicate what topics they would like to be reviewed in a lecture (Boyce & Hineline, 2002). The instructor then creates a clarifying lecture for the next class period based on the feedback received from students in the record sheet. This clarifying lecture is brief and takes up only about one third of the next class period. Another important component of interteaching is using frequent probes, or exams, which contain questions directly related to the material covered in the prep guide. Interteaching may also include quality points, which are an explicit cooperative contingency where students earn points if everyone in their discussion group correctly answers certain exam questions that were covered during the discussion. However, interteaching has been shown to still be effective without the use of quality points (e.g., Gayman, Hammonds & Rost, 2018; Saville & Zinn, 2009).

There seem to be several aspects of interteaching that are associated with higher exam scores as well as better long-term retention. For example, prep guides, interteaching sessions, and clarifying lectures provide a contingency for distributed practice. That is, students are exposed to the material on three separate occasions: once before class when they are completing the reading and prep guide, a second time in class when they discuss the material during the interteaching session, and a third time when the instructor presents a clarifying lecture. Distributed practice has repeatedly been shown to increase information retention (Roediger, 2013). Interteaching also makes use of frequent testing, which has been demonstrated to aid in recall and long-term retention (Pan, Pashker, Potter, Rickard, 2015; Roediger & Karpicke, 2006). Further, Felderman (2014) showed, using interteaching, that students who had smaller, more frequent exams earned higher grades than those who experienced fewer exams.

Since the seminal article by Boyce and Hineline (2002), over 30 studies have been published investigating interteaching and its components (see Querol et al., 2015; Saville et al., 2011b; Sturmeiy et al., 2015 for reviews). Several modifications of the basic method have since been examined. For example, interteaching seems to be equally effective regardless of whether the clarifying lecture occurs immediately after the group discussion or at the start of the next class period (Saville et al., 2011a). It is also equally effective regardless of whether the interteaching sessions are broken up into smaller or larger groups (e.g. 2 vs. 4) (Truelove, Saville, & Van Patten, 2013); although students seem to prefer working in larger groups (Goto & Schneider, 2010; Scoboria & Pascual-Leone, 2009). Variations on the implementation of the prep guide have also been examined. Filipiak, Rehfeldt, Heal, and Baker (2010) manipulated whether points were contingent on the completion of prep guides and found that although contingent points led to more students completing the prep guide, they did not increase exam scores. Another study demonstrated that writing one's own prep guide questions led to slightly higher exam scores, although not statistically higher, than simply answering the questions provided on a prep guide by the instructor (Cannella-Malone, Axe, & Parker, 2009).

Another methodological alteration to the standard interteaching method was conducted by Goto and Schneider (2009; 2010). Here, students were first divided into two groups. Prior to class, one group was assigned to complete a prep guide covering the first half of the reading material and the other group was assigned to complete a prep guide covering the second half of the material. In class, there were two rounds of group discussions. First, students formed dyads and discussed the questions from their assigned prep guide with students assigned to complete the same half of the prep guide. Subsequently, students formed larger groups of four (i.e. two students assigned to read the first half of the material joined with two students assigned to read the second half of the material). Students then taught those in the other dyad the answers to their prep guide questions, using collaborative, discussion-based peer-tutoring. Goto and Schneider (2009; 2010) reported in two studies that students

rated this version of interteaching as being more enjoyable than lecture-centered instruction. However, they did not compare this variation of interteaching to the original method where all students complete a prep guide covering all material prior to class and then spend the duration of the interteaching session discussing all prep guide questions. The current study aimed to systematically investigate the difference in test scores between the prep guide and discussion modification described by Goto and Schneider and the original methodology described by Boyce and Hine (2002).

## Method

### *Participants and Setting*

A total of 33 undergraduate students were enrolled in one of two psychology of learning courses at a small northeastern university. Participants ranged in age from 19-27 yr ( $M = 21.48$ ). The majority were female (78.78%) and Caucasian (84.84%). All participants held either junior (30.30%) or senior (69.69%) standing. Each of the two sections took place in sequential semesters. The first section (SEC 1) had 16 students and met in the fall semester of 2017. The second section (SEC 2) had 17 students and met in the spring semester of 2018. Both sections met on Mondays, Wednesdays, and Fridays for 50 min.

Since participants were not randomly assigned to groups, several demographic characteristics were collected through self-report on the first day of the semester to see if the two sections were significantly different from one another. Students were asked to report their cumulative grade point average (GPA), the number of psychology courses previously taken, total number of credits taken that semester, if they were employed and how many hours they worked, whether they had a significant other, how many children they had, if they participated in a sport that semester, and if they were part of a sorority or fraternity. None of these demographics differed significantly across the two sections.

### *Procedure*

*SEC 1 Interteaching.* Students were given an 8-12 question prep guide to complete before coming to each class. These prep guides were made available on the online learning platform, Courseweb, at least 7 days before the class meeting. Each prep guide corresponded with 5-10 pages of reading from the course textbook. Once in class, students were divided into groups of four, which were formed using the phone application Team Maker Pro™ that randomly assigned students into groups. This was done so students worked in different groups every class. Students were then given approximately 30 min to discuss the questions on the prep guide. During this time, the instructor walked from group to group and answered any questions students had. The last 5 min were used to fill out a record sheet where students indicated the quality of the group discussion and any topics they wanted clarification on. These record sheets were used by the instructor to create a 15-20 min clarifying lecture to be given the next class period. For example, prep guide questions discussed in groups on Monday were reviewed in a clarifying lecture on Wednesday.

*SEC 2 Interteaching.* This procedure was similar to the interteaching used in SEC 1, with two main exceptions. First, students were randomly divided into two groups: Group A ( $n = 8$ ) was given a prep guide that contained the first half of the questions used in SEC 1 to complete prior to class, and Group B ( $n = 9$ ) was given a prep guide that contained the second half of the questions. These prep guides were exactly the same as those used in SEC 1, but split in half. The second main difference involved the group discussions. First, students from each group formed dyads or triads and discussed the questions on their assigned prep guide. That is, Group A students formed dyads and discussed the

half of the prep guide assigned to them, and Group B did the same. Then, two students from Group A would form a larger group with two students from Group B and they would go over all prep guide questions assigned to both groups. Students from Group A taught students from Group B the answers to their prep guide questions and vice versa. During both discussions, the instructor migrated between groups, answering any questions students had.

*Exams.* Each section received the same six unit exams and a comprehensive final. Exams were each worth 9% (54% cumulatively) of students’ total grade and consisted of about 30 multiple choice and 8 short answer questions. Points were evenly distributed across question type. All short answer questions were taken from the prep guides. The comprehensive final had a similar format to the unit exams and was worth 12% of students’ total grade. The final exam consisted of eight multiple choice and one to two short answer questions which were recycled from each of the six unit exams.

Interobserver agreement (IOA) data were collected on 64 of the short answer questions and was calculated by taking the total number of questions where both observers agreed divided by the total number of questions scored, then multiplied by 100. An observation was considered to be in agreement when the same number of points was assigned to the question by both observers. Bailey and Birch (2017) suggest that if IOA is above 80%, then the data can be considered reliable. Out of 64 questions, 58 were agreed on by both observers, yielding an IOA of 90.62%. Thus, the short answer questions were reliably graded.

*Social validity.* At the end of each semester, students were asked to anonymously complete an 11-question survey rating their experience with interteaching (see Table 1). Each question was rated on a slider scale from 0 (“strongly disagree/dislike”) to 100 (“strongly agree/like”). Anchors between these two extremes included “disagree/dislike” (25), “neutral” (50), and “agree/like” (75). Students were given bonus points for completing the survey, which took students about 15 min. These statements asked students to rate various aspects of interteaching, such as their preference for interteaching or standard lecture (Q1, Q2, Q3, Q4), their perceptions of whether interteaching increased their comprehension and critical thinking (Q5, Q6, and Q7), how well they felt interteaching prepared them for class (Q8, Q9, and Q10), and their commitment to interteaching (Q11). The majority (7/11) of the questions used were similar to those used by Goto and Schneider (2010).

**Table 1. Mean and standard deviation ratings from social validity questionnaire**

Statements/Questions	SEC 1	SEC 2	<i>p</i> -value
	Mean ± SD	Mean ± SD	
<b>Preference for Interteaching or Lecture</b>			
1. What is your preference for standard lecture? (lecture, assignment, exam)	73.63 ± 18.31	70.76 ± 12.29	.60
2. What is your preference for interteaching? (Prep guides, group discussions, clarifying lectures.	67.63 ± 30.06	45.76 ± 22.59	.02*
3. The combination of interteaching and lecture is more effective than lecture alone.	72.93 ± 23.24	52.82 ± 23.13	.02*
4. Which teaching method do you feel you learn more from?	56.54 ± 37.06	28.59 ± 16.57	.01*
<b>Perceptions of Interteaching on Critical Thinking</b>			
5. Interteaching prep guide questions fostered critical thinking.	79.06 ± 11.30	67.29 ± 23.84	.08
6. Group discussions in class fostered critical thinking.	77.34 ± 14.13	62.18 ± 20.41	.02*

7. Participating in interteaching sessions helped me understand complex scientific concepts rather than memorizing factual information.	70.87 ± 22.84	52.53 ± 22.79	.03*
<b>Preparation for Class</b>			
8. Participating in group discussions contributed to enhancing my motivation toward coursework.	73.87 ± 17.43	65.82 ± 27.04	.33
9. I was better focused during the clarifying lecture since I was already exposed to the topic through interteaching.	80.82 ± 16.32	64.31 ± 29.43	.06
10. Interteaching prep guides helped me read materials and prepare well for class.	79.20 ± 19.53	71.12 ± 20.91	.26
<b>Commitment to Interteaching</b>			
11. Please rate your overall commitment to interteaching this semester.	81.94 ± 17.44	62.23 ± 27.10	.02*

**Note:** A 100-point slider scale ranged from “strongly disagree” or “dislike” (0) to “strongly agree” or “like” (100), except for Q4 and Q11. Q4 had anchors of “standard teaching (e.g. lecture, assignments, exams)” (0), “no preference” (50) and “interteaching (e.g. prep guides, group discussions)” (100), whereas Q11 ranged from “very low” (0) to “very high” (100). Asterisks denote statistical significance of  $p < .05$ .

## Results

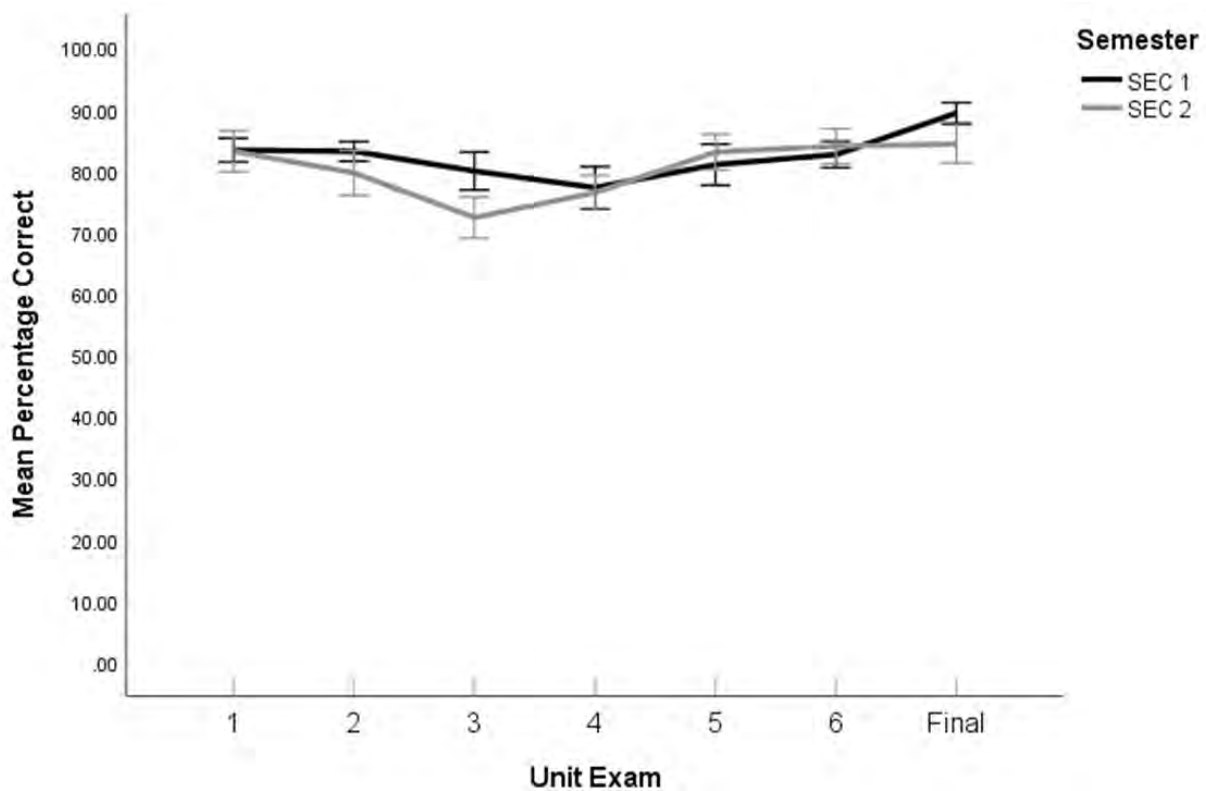
### Exams

An independent samples t-test was used to compare the mean percentage correct on exams for each of the six unit exams and the final across SEC 1 and SEC 2. Table 2 lists the means, standard deviations, and p-values. Although none of the t-tests found statistically significant differences between the two sections, a trend was found demonstrating that students in SEC 1 scored higher on all of the exams, except exams 5 and 6 (see Figure 1). This difference was most pronounced on exams 2, 3, and the final. An independent samples t-test was also used to evaluate the difference in the mean total number of points earned through all exams, including the final for SEC 1 ( $M = 454.95$ ,  $SD = 42.05$ ) versus SEC 2 ( $M = 443.38$ ,  $SD = 62.08$ ); no significant difference was demonstrated,  $t(229)=1.35$ ,  $p = .18$ . However, across all exams and the final, participants in SEC 1 scored an average of 11.58 points out of 550 (2.10%) higher than participants in SEC 2.

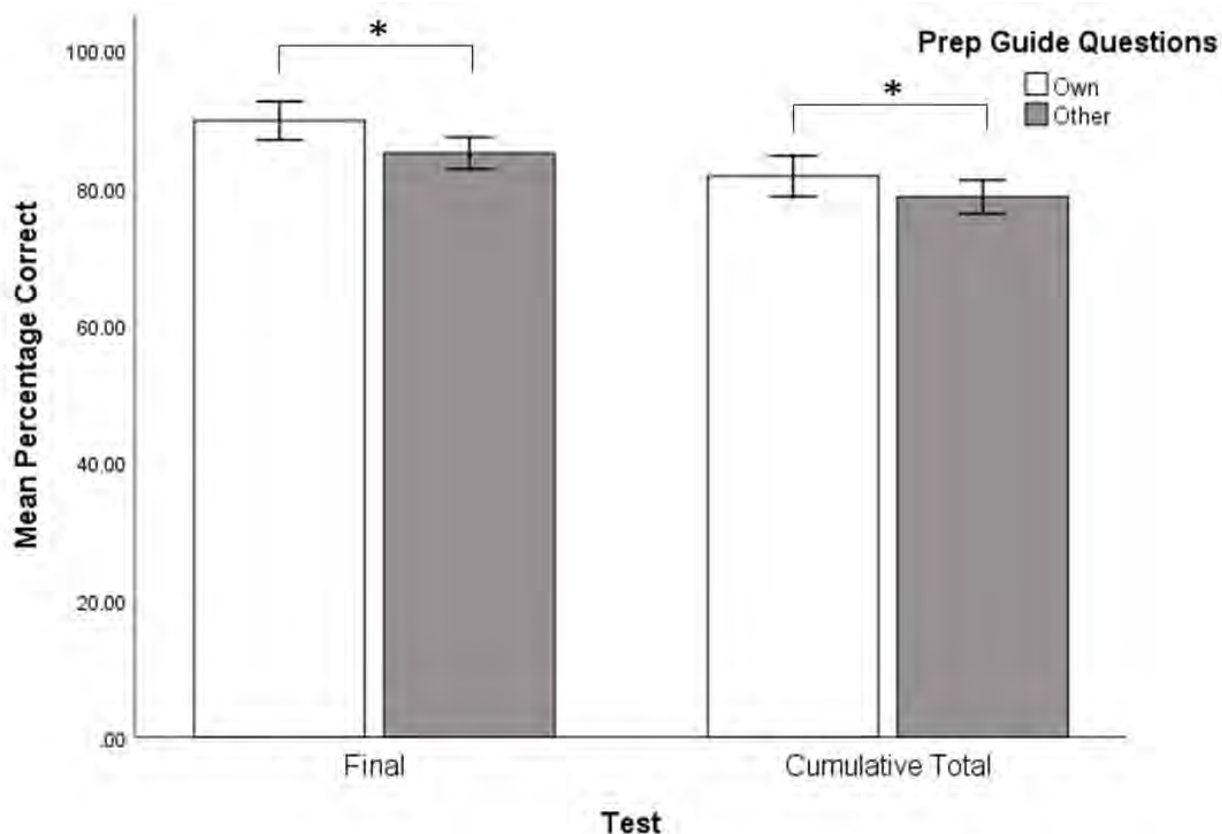
**Table 2. Mean percentage correct, standard deviations, and p-values for each for the six unit exams and the final in SEC 1 and SEC 2**

Exam	SEC 1		SEC 2		p-value
	Mean (%)	SD (%)	Mean (%)	SD (%)	
1	83.37	7.70	83.17	13.80	.96
2	83.16	6.36	79.64	15.09	.39
3	79.95	12.41	72.39	13.73	.11
4	77.25	13.70	76.43	11.52	.85
5	81.00	13.33	83.05	11.99	.64
6	82.65	8.58	84.00	11.88	.71
Final	89.41	6.77	84.35	12.80	.17
Mean Total	82.71	7.65	80.61	11.28	.53

Further investigation into the exams from SEC 2 aimed to determine whether students were more likely to correctly answer exam questions that came from the half of the prep guide they had answered prior to class and taught to the group versus the half that had been taught to them. In terms of both the comprehensive final ( $t[16] = 2.57, p = .02, d = 1.57$ ) and the combined total from all exams, including the final ( $t[16] = 2.72, p = .01, d = .66$ ), a paired samples t-test found a statistically significant difference in the mean percentage of points earned. Participants in each group earned more points on the final and cumulatively through all exams on questions pulled from prep guides they taught to other students ( $M = 89.80, SD = 11.52; M = 81.72, SD = 12.29$ , respectively) than questions pulled from prep guides that they learned from other students ( $M = 85.08, SD = 9.57; M = 78.71, SD = 10.11$ , respectively) (see Figure 2). Further, the magnitude of the effect for the final was large according to Cohen’s (1992) criteria and the cumulative mean total number of points earned on all exams had a medium effect size.



**Figure 1. Mean percentage correct score for each of the six unit exams and the comprehensive final in SEC 1 (black line) and SEC 2 (grey line). Error bars show  $\pm 1.0$  standard error.**



**Figure 2.** Mean percentage correct on the final and cumulative total number of points earned through each unit exam for questions pulled from either a group's own prep guide (white bars) or the other group's prep guide (grey bars) in SEC 2. Error bars show  $\pm 1.0$  standard error and asterisks denote statistical significance of  $p < .05$ .

In order to better determine whether the modification suggested by Goto and Schneider (2009; 2010) resulted in lower academic performance on the half of the prep guide material assigned to other students, we compared the mean percentage of points earned on exam questions pulled from this prep guide material on the final and cumulatively through all exams in SEC 2 with the mean percentage of points from SEC 1. An independent samples t-test found no difference in percentage correct on the final ( $t[31] = 1.49, p = .14$ ) or cumulative total points earned ( $t[31] = 1.28, p = .21$ ) between SEC 1 ( $M = 89.41, SD = 6.77; M = 82.72, SD = 7.64$ , respectively) and the questions pulled from prep guides not assigned to students in SEC 2 ( $M = 85.08, SD = 9.57; M = 78.71, SD = 10.11$ , respectively). These results are somewhat at odds with the paired samples t-test further investigating exam scores in SEC 2 reported above. Here, an independent samples t-test demonstrated that a 4.33% difference on the final ( $M = 89.41$  vs  $M = 85.08$ ) and 4.01% difference in cumulative total points ( $M = 82.72; M = 78.71$ ) were nonsignificant; whereas the paired samples t-test showed that a 4.72% difference on the final ( $M = 89.80$  vs  $M = 85.08$ ) and a 3.01% difference in cumulative total points ( $M = 81.72$  vs  $M = 78.71$ ) were statistically significant. This discrepancy in statistical significance is most likely due to differences in variability associated with between- and within-subject methodologies. Between-subject designs involve higher variability from individual differences across groups, which leads to a higher standard error in an independent samples t-test. Within-subject designs, where participants are tested against themselves, tend to have less variability, and therefore a smaller mean standard error, since there are no individual differences.

*Social Validity*

As seen in Table 1, students were asked to rate 11 statements on a slider scale. Students in both sections rated interteaching highly in terms of the method fostering critical thinking, enhancing their motivation toward coursework, helping them focus, preparing them for class, and their commitment to the pedagogy (Q5, Q6, & Q8-Q11). Independent samples t-tests compared the ratings of each question between SEC 1 and SEC 2 (see Table 1 for means, standard deviations, and  $p$ -values). While participants showed no difference in their preference for standard lecture (Q1); they did demonstrate a significant difference in their preference for interteaching (Q2), how effective they thought interteaching was (Q3, and Q4), whether they thought it encouraged critical thinking (Q6 and Q7), and their commitment to interteaching (Q11).

Participants in SEC 1 rated a higher preference for interteaching ( $p = .02$ ,  $d = .82$ ) or interteaching plus lecture ( $p = .02$ ,  $d = .87$ ), than participants in SEC 2 (Q2 and Q3, respectively). When participants' preference ratings for standard lecture (Q1) and interteaching (Q2) were compared within each section, a paired samples t-test demonstrated no difference in the preference of participants in SEC 1,  $t(15) = .55$ ,  $p = .59$ ). Students in SEC 2, however, preferred standard lecture significantly more than interteaching,  $t(16) = 4.18$ ,  $p = .001$ ,  $d = 1.01$ . Moreover, they felt they learned more from standard lecture in comparison to students in SEC 1 (Q4;  $p = .01$ ,  $d = .97$ ).

Participants in SEC 1 also agreed more strongly than participants in SEC 2 with statements related to interteaching increasing critical thinking ( $p = .02$ ,  $d = .86$ ) and their ability to understand complex scientific concepts ( $p = .03$ ,  $d = .80$ ) (Q6 and Q7, respectively). Additionally, participants in SEC 1 had a higher commitment to interteaching than participants in SEC 2 (Q11;  $p = .02$ ,  $d = .86$ ). All of these ratings differences between SEC 1 and SEC 2 demonstrated a large effect size (Cohen, 1992). Overall, participants in SEC 1 perceived the standard interteaching method as being more preferable, leading to better critical thinking, and increasing commitment to the pedagogy in comparison to the variation on interteaching experienced by participants in SEC 2.

**Discussion**

The goal of the present study was to evaluate the effect of modifying the in-class interteaching session using the format first described by Goto and Schneider (2009; 2010). Using this modification, students in SEC 2 were assigned to read and complete prep guides covering *either* the first *or* the second half of the reading prior to class. During the first half of the discussion, these students reviewed answers to their prep guide with another student who also read the same half of the material. During the second half of the discussion, students combined with another dyad who read the other half of the material. Each dyad then taught the other the information covered by the prep guide questions that they were assigned to answer prior to class. This differed from SEC 1, in which students were assigned to a standard interteaching session, where all students were assigned to read and answer all prep guide questions for a given unit prior to class and there was only one round of discussions in class (Boyce & Hineline, 2002). The current study found no significant difference in exam scores between these two prep guide and discussion methods. This finding is, to some extent, congruent with Filipiak et al. (2010), who manipulated whether points were contingent on the completion of prep guides. They demonstrated that, although points increased the completion of prep guides, it did not translate to an increase in exam scores. Further, Cannella-Malone et al. (2009) did not find differences in exam scores when students had to either simply answer prep guide questions or generate their own questions. Taken together, these findings suggest that there may be little to no relationship between variations of the prep guide and exam scores.



Further examination of the data from SEC 2 found that students answered more exam questions correctly that were based on material from the half of the prep guide that they had been assigned to complete in comparison to the other half. This effect was seen when points from all exams were aggregated together, as well as on the cumulative final. Moreover, students in SEC 2 were less likely to rate interteaching as preferable to lecture or helpful in increasing critical thinking or comprehension.

Despite the fact that the differences in exam scores between sections was not statistically significant, visual inspection of the data (see Figure 1) showed a trend in which students scored higher across all exams, except exams 5 and 6, in SEC 1 compared to SEC 2. It is possible that there was not a statistically significant difference in exam scores between the sections due to a reduction in statistical power from testing between- versus within-subject effects. For example, when within-subject effects were evaluated for the mean percentage correct on the final pulled from students' own prep guide versus the other group's prep guide ( $M = 89.80$ ;  $M = 85.08$ , respectively), that 4.72%-point difference was statistically significant. However, when between-subject effects were evaluated between the mean percentage correct on the final in SEC 1 and the mean percentage correct from the prep guide questions taught to students in SEC 2 ( $M = 89.41$ ;  $M = 85.08$ , respectively), that 4.33%-point difference was not statistically significant.

One explanation for the slightly higher exam scores in SEC 1 may be because the original interteaching method used in that section is more effective than the modified interteaching method used in SEC 2. This hypothesis is supported by the fact that students in SEC 2 answered more exam questions correctly that were based on the half of the prep guide that they had been assigned to complete prior to class. Students in SEC 2 scored an average of 3.01% higher cumulatively across all unit exam questions ( $M = 81.72$ ;  $M = 78.71$ , respectively), and 4.72% higher on the final exam questions ( $M = 89.80$  vs  $M = 85.08$ , respectively) that came from the half of the prep guide that they were assigned. That is, it is possible that the slightly lower average exam scores in SEC 2 were caused by students being more likely to miss exam questions from the half of the prep guide that was not assigned to them. It is likely that average exam scores would have been higher in SEC 2 if all students had read all material, completed prep guides in their entirety, and engaged in a discussion covering all material like the standard interteaching implemented in SEC 1. Ultimately, additional research is needed to determine if this potentially underlies the current results.

There are a couple of plausible explanations for why students from SEC 2 scored higher on exam questions that came from material they taught to other students. One may be that students learning material from peers is not as beneficial as reading the material themselves prior to class. In other words, students may not have taught the material as well as the text. Support for this idea comes from students' course evaluation comments that they did not believe classmates had sufficient knowledge of the material to teach it to others. Goto and Schneider (2010) also reported student comments along those lines, as well as students noting that they just read their prep guide answers to students in the other group without having a full discussion of the topic. Another interpretation of why students in SEC 2 scored more points on exam questions that came from their portion of the prep guide is that they had two additional exposures to that material prior to the exam compared to material that came from the other half of the prep guide. Students first contacted the material from their half of the prep guide when they answered those prep guide questions prior to class. Then, they contacted that same material when they went over it dyads before the larger group discussion of the entire prep guide. Students were only required to contact the material from the other half of the prep guide once, when the larger group discussion occurred. A third possibility may be that teaching their half of the prep guide to other students increased learning of that material. Active learning, where students are required to participate in the learning process as opposed to passively listen, has been shown to increase student achievement and knowledge retention (for a meta-analysis see Freeman et

al., 2014). It is possible that students may have built up fluency with the concepts while explaining them to others. In fact, the interteaching session is an active learning component of interteaching and is likely one of the components that makes the method so effective (Soldner, Rosales, Crimando & Schultz, 2017).

Based on the current study, it is impossible to determine whether it was reading and completing prep guides over the material or requiring students to teach that information to others that caused a higher number of exam questions to be answered correctly that came from the half of the prep guide assigned to students in SEC 2. It would only be possible to parse out these two potential causes by running a third group that read and completed a prep guide covering the entire unit of material, but then only taught half of the prep guide to another group. If students then scored higher on exam questions that came from the half of the prep guide that they taught to other students, it would suggest that teaching other students the material is the factor that led to higher comprehension. Future research could investigate a situation similar to this in order to separate the effect of simply reading and answering prep guides as opposed to teaching prep guide questions to another student on academic success. Nevertheless, the current study does indicate that the traditional interteaching method described by Boyce and Hine (2002) may lead to better student learning.

Overall, students in both sections indicated that interteaching fostered critical thinking, enhanced their motivation towards coursework, helped them focus on the clarifying lecture, prepared them for class, and that they were committed to the teaching method (Q5, Q6, & Q8-Q11). This corroborates results of Goto and Schneider (2010), where students rated similar questions, on average, at around 4 on a scale ranging from “strongly disagree” (1) to “strongly agree” (5). Students in both sections seemed to indicate that interteaching was helpful to their learning.

However, when responses on the social validity questionnaire were compared across sections, students who experienced the traditional interteaching method (SEC 1) rated interteaching as more preferable and effective (Q2 and Q3), indicated that they learned more from it (Q4), that it fostered critical thinking and understanding complex concepts (Q6 and Q7), and that they were more committed to the interteaching method (Q11) compared to students who experienced the modified interteaching method (SEC 2). Further, preference for interteaching was only slightly lower than preference for standard teaching in SEC 1, but this preference was much more pronounced in SEC 2. On a slider scale from “strongly dislike” (0) to “strongly like” (100) students in SEC 1 gave an average preference rating of 73.63 to standard teaching, and an average interteaching rating of 67.63. The preference of students in SEC 2 for standard lecture was significantly stronger than their preference for interteaching ( $M = 70.76$ ;  $M = 45.76$ , respectively).

Students rating interteaching as less preferable than standard lecture is unusual in the interteaching literature (see Querol et al., 2015; Saville et al., 2011; Sturme et al., 2015 for reviews), so it seems likely that the lower preference for interteaching seen in SEC 2 was a result of the modified interteaching method that students experienced. To some extent, this could be due to the modified interteaching session where half of the time was spent in small groups of 2-3. Previous research has demonstrated that students prefer working in larger groups of 4 more than smaller groups of 2 (e.g. Goto & Schneider, 2010; Scoboria & Pascual-Leone, 2009; Truelove et al., 2013). Further, course evaluation comments from students in SEC 2 indicated that they felt they did not learn the material taught by other students as well as they learned the material on their own prep guides. Students also reported that they felt it was confusing at times to start a reading in the middle of the chapter without reading the previous section. This is corroborated by anecdotal reports of students from Goto and Schneider (2010), where several peer assessments indicated that students thought their classmates did not know the material well enough to teach it and that they were concerned about the different levels of commitment and preparedness of other students. Thus, perhaps peer-learning is only helpful when

both students are familiar with the material and one peer clarifies the information to the other, as opposed to one peer teaching the other novel material.

Future studies should investigate additional ways to modify and standardize the traditional interteaching method in order to maximize its effectiveness. For example, instructions given to students about what interteaching is and why it is more effective than traditional lecture may have an effect on student commitment to the methodology, which may influence academic performance. Also, instructions to students for how to engage in a high-quality discussion could be developed and systematically investigated. These instructions and how students are graded for discussion contributions likely vary widely across instructors and research studies, as there are no published standard instructions or grading rubrics. It is likely that instructions and feedback about the quality of a student's contributions to the discussion could influence both learning and how favorably interteaching is viewed by students.

The present study indicates that students who were exposed to the interteaching modification laid out by Goto and Schneider (2009; 2010) (SEC 2), scored higher on exam questions that came from the half of the prep guide they were assigned to complete. Thus, they learned more from the half of the prep guide that they were responsible for completing and teaching to others than from the half that they discussed in class without having contacted it beforehand. Students who were exposed to the original interteaching method described by Boyce and Hinline (2002) (SEC 1), rated interteaching as more preferable, indicated that they learned more from it, agreed more strongly that it fostered critical thinking, and indicated a greater commitment to the teaching method. This shows that students seem to enjoy and value interteaching less when exposed to the modification laid out by Goto and Schneider (2009; 2010). Although future research is required, it is suggested that it may be more beneficial to students for instructors to assign students to read and complete prep guides covering the entire unit of material.

### Acknowledgements

We would like to thank Clarissa Nepereny and Cortni Gregus for their time and effort in the data collection process.

### References

- Arntzen, E., & Hoiium, K. (2010). On the effectiveness of interteaching. *Behavior Analyst Today, 11*(3), 155-160. <http://dx.doi.org/10.1037/h0100698>.
- Bailey, J., & Birch, M. (2017). *Research methods in applied behavior analysis* (2<sup>nd</sup> Ed.). New York, NY: Routledge.
- Boyce, T., & Hinline, P. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst, 25*(2), 215-225. <https://doi.org/10.1007/BF03392059>.
- Cannella-Malone, H., Axe, J., & Parker, E. (2009). Interteaching preparation: A comparison of the effects of answering versus generating study guide questions on quiz scores. *Journal of the Scholarship of Teaching and Learning, 9*(2), 22-35.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Felderman, T. (2014). Preliminary analysis of interteaching's frequent examinations component in the community college classroom. *Journal of College Teaching and Learning, 11*, 149-156. DOI: 10.19030/tlc.v11i4.8851.

- Filipiak, S., Rehfeldt, R., Heal, N., & Baker, J. (2010). The effects of points for preparation guides in interteaching procedures. *European Journal of Behavior Analysis, 11*, 115-132. <https://doi.org/10.1080/15021149.2010.11434338>.
- Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111*(23), 8410–8415. <https://doi.org.libproxy.troy.edu/10.1073/pnas.1319030111>.
- Gayman, C., Hammonds, F., & Rost, K. (in press). Interteaching in an online asynchronous class. *Scholarship of Teaching and Learning in Psychology*.
- Goto, K., & Schneider, J. (2009). Interteaching: An innovative approach to facilitate university student learning in the field of nutrition. *Journal of Nutrition Education and Behavior, 41*, 303–304. <http://dx.doi.org/10.1016/j.jneb.2009.02.003>.
- Goto, K., & Schneider, J. (2010). Learning through teaching: Challenges and opportunities in facilitating student learning in food science and nutrition by using the interteaching approach. *Journal of Food Science Education, 9*, 31–35. <https://doi.org/10.1111/j.1541-4329.2009.00087.x>.
- Pan, S., Pashler, H., Potter, Z., & Rickard, T. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language, 83*, 53-61. <https://doi.org/10.1016/j.jml.2015.04.001>.
- Querol, B., Rosales, R., & Soldner, J. (2015). A comprehensive review of interteaching and its impact on student learning and satisfaction. *Scholarship of Teaching and Learning in Psychology, 1*(4), 390-411. <http://dx.doi.org/10.1037/stl0000048>
- Roediger, H. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest, 14*, 1-3. <https://doi.org/10.1177/1529100612454415>.
- Roediger, H., & Karpicke, J. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249 –255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Rosales, R., Soldner, J., & Crimando, W. (2014). Enhancing the impact of quality points in interteaching. *Journal of the Scholarship of Teaching and Learning, 14*, 1–11. <https://doi.org/10.14434/josotlv14i5.12746>.
- Saville, B., Cox, T., O'Brien, S., & Vanderveldt, A. (2011a). Interteaching: The impact of lectures on student performance. *Journal of Applied Behavior Analysis, 44*, 937–941. <https://doi.org/10.1901/jaba.2011.44-937>.
- Saville, B., Lambert, T., & Robertson, S. (2011b). Interteaching: Bringing behavioral education into the 21<sup>st</sup> century. *Psychological Record, 61*(1), 153-165. <https://doi.org/10.1007/BF03395752>.
- Saville, B., & Zinn, T. (2009). Interteaching: The effects of quality points on exam scores. *Journal of Applied Behavior Analysis, 42*, 369-374. <https://doi.org/10.1901/jaba.2009.42-369>
- Saville, B., Zinn, T., & Elliott, M. (2005). Interteaching versus traditional methods of instruction: A preliminary analysis. *Teaching of Psychology, 32*(3), 161-163. [https://doi.org/10.1207/s15328023top3203\\_6](https://doi.org/10.1207/s15328023top3203_6).
- Saville, B., Zinn, T., Neef, N., Van Norman, R., & Ferreri, S. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39*(1), 49-61. <https://doi.org/10.1901/jaba.2006.42-05>.
- Scoboria, A., & Pascual-Leone, A. (2009). An 'inter-teaching' informed approach to instructing large undergraduate classes. *Journal of the Scholarship of Teaching and Learning, 9*, 29–37.
- Soldner, J. L., Rosales, R., Crimando, W., & Schultz, J. C. (2017). Interteaching: Application of an empirically supported behavioral teaching method in distance rehabilitation education.

*Rehabilitation Research, Policy, and Education*, 31(4), 372-386. <http://dx.doi.org/10.1891/2168-6653.31.4.372>.

Sturmev, P., Dalfen, S., & Fienup, D. (2015). Inter-teaching: A systematic review. *European Journal of Behavior Analysis*, 16, 121-130. <https://doi.org/10.1080/15021149.2015.1069655>.

Truelove, J., Saville, B., & Van Patten, R. (2013). Interteaching: Discussion group size and course performance. *Journal of the Scholarship of Teaching and Learning*, 13, 23–30.