Journal of the Scholarship of Teaching and Learning, Vol. 18, No. 2, June 2018, pp. 105-117.
doi:10.14434/josotl.v18i2.22539

An Assessment of Group Size in Interteaching

Rocío Rosales
University of Massachusetts Lowell

James L. Soldner
University of Massachusetts Boston

Abstract: A key component of interteaching, as described by Boyce and Hineline (2002), is the opportunity for students to participate in "dyadic" or pair discussions. Although the rationale for pair discussions is evident, only one study involving interteaching, to date, has evaluated the relative effectiveness of student performance when group size is manipulated. The present investigation was designed to further evaluate the effect of group size during student discussions on quiz scores in an introductory psychology course with a diverse group of learners. An alternating treatments design was implemented whereby students were assigned to work in a dyad or in groups of 4-5 students to discuss a preparation guide. All of the major components of interteaching were in effect during both conditions (i.e., availability of prep guides and quality points, clarifying lectures, and frequent test probes). Results showed a small advantage for performance following discussion in dyads. A social validity measure indicated students favored discussion in larger groups. Implications of these findings and suggestions for future work will be discussed.

Keywords: Interteaching, group size, pair discussion, component analysis, college students

Interteaching is an empirically supported, behaviorally-based teaching strategy that offers an alternative to lecture for classroom instruction. Interteaching includes several key components: 1) preparation guides that generally serve as a study guide for students and cover a small amount of assigned reading material; 2) an in-class discussion with one classmate to review responses to the preparation guides completed before class; 3) completion of a record sheet to provide feedback to the instructor on the quality of the in-class discussion and topics that were most challenging; 4) clarifying lectures to cover class topics based on feedback provided on record sheets; 5) frequent test probes on the material from the assigned reading material and prep guides; and 6) quality points, a cooperative contingency whereby additional points are delivered if all members of a respective discussion group perform at or above a certain criteria on a question(s) or an entire test probe. For a more complete description of these components, please see Saville, Lambert, and Robertson (2011).

The authors who developed interteaching describe it as a “mutually probing, mutually informing conversation between two people (Boyce & Hineline, 2002, p. 220); thereby highlighting the importance of one of the key components of interteaching - the pair discussion. The pair discussion component requires students to come to class prepared to interact with a classmate. During the pair discussion, students share information on preparation guides assigned for each class period, and provide feedback on their discussion to the instructor in the form of a

record sheet. In this respect, interteaching, and pair discussion, in particular, is similar to other teaching strategies, such as reciprocal peer tutoring and cooperative learning, which promote a collaborative learning and achievement model (Johnson & Johnson, 1975; Leung, 2015; Slavin, 1996, 2014).

There is emerging empirical evidence to support the use of interteaching when compared to traditional lecture (Arntzen & Hoium, 2010; Saville, Zinn, & Elliot, 2005; Saville, Zinn, Neef, Van Norman, & Ferreri, 2006; Soldner, Rosales, & Crimando, 2015). A growing number of studies have also shown consistent academic achievement when interteaching is used with no direct comparisons to lecture. Notably, many studies have also evaluated students' perspectives and preference for interteaching via social validity measures. For example, Saville and colleagues (2006) asked students to complete a two-item questionnaire at the end of the semester that asked which of the two teaching methods they preferred, and to rate their extent of learning from each method. Zayak and Paulk (2014) had a similar approach to evaluate social validity, but included additional open-ended questions in the form completed by participants. The reported preference for interteaching when compared to lecture has been mixed, with at least some studies reporting student preference for interteaching while others reported preference for lecture even when interteaching was more effective (see Querol, Rosales, & Soldner, 2015 for a review of empirical research on Interteaching including the use of social validity measures).

To date, only a handful of studies have conducted component analyses to help identify the impact of each element of interteaching. For example, Saville, Cox, O'Brien, and Vanderbelt (2011) evaluated the impact of the clarifying lecture on the efficacy of interteaching. Results of this study showed better performance on two exam scores when lectures were included. Saville and Zinn (2009) examined the impact of quality points on student performance, but reported no statistically significant differences in exam scores. In contrast, Rosales, Soldner, and Crimando (2014) reported better results on quiz performance when quality points contingency and immediate feedback on performance was included in the implementation of Interteaching in the classroom.

Soldner, Rosales, Crimando, and Schultz (2017) conducted a component analysis of the pair discussion component on interteaching in an online graduate rehabilitation course. In this study, two conditions were randomly assigned across participants and class sessions. The first condition included all key components of interteaching, including the pair discussion. The second condition included all key components with the exception of the pair discussion. Results indicated the pair discussion condition resulted in higher student quiz scores. Additionally, social validity findings indicated the majority of students preferred interteaching with the pair discussion. This component analysis provides preliminary empirical evidence for the importance of including the pair discussion component in the implementation of interteaching in the classroom. Although, further investigation is needed regarding the particular and most effective methods used for compiling students into discussion groups, including the number of respective students in each discussion group.

One rationale for grouping students in dyads is that it may help to control for the possibility of students underpreparing for in-class discussions with a classmate. That is, when students are in larger groups the possibility of "free-riders" or social loafing tends to increase (Chidambarum & Tung, 2005; Latane, 1981). The possibility of social loafing can be remedied if students are in dyads for discussion. Placing students in dyads for discussions also presents equitable opportunities for each student to share their perspective on each prep guide item and thus may improve the overall quality of the discussion. However, placing students into dyads

may present some concerns. Specifically, instructors may feel that traversing the classroom and facilitating discussion equally among all the groups when students are in dyads poses a challenge because they are limited in the time they can spend with each dyad. This challenge may be especially evident when teaching assistants are not available during class time or in classes with larger student enrollment (Scoboria & Pascual-Leone, 2009). For this reason, it is important to systematically evaluate the feasibility and validity of assigning larger groups during class discussions. In addition, a study by Goto and Schneider (2010) reported students preferred to work in larger groups during interteaching. Other researchers have suggested, for example, that cooperative learning works best when group size is smaller, although this research has not been evaluated within the framework of interteaching specifically (Johnson & Johnson, 2009).

A recent study by Truelove, Saville, and Patten (2013) systematically manipulated group size for interteaching sessions and evaluated the impact on student performance. Students from two sections of an undergraduate psychology course served as participants. In one class section, students were assigned to work in groups of two during all interteach sessions; and in the second section, students were assigned to work in groups of four during all interteach sessions. Students were assigned to either groups of two or groups of four in a random fashion throughout the semester. The researchers reported no statistically significant differences between the sections on six exams administered throughout the semester, or on a cumulative final or the total number of points earned. The authors of this study outlined some notable limitations to their study.

First, a 'no discussion' condition was not included. As previously mentioned, at least one prior study has now examined the impact of the pair discussion component, showing favorable results for the inclusion of the pair discussion in interteaching (Soldner et al., 2017). Second, the participants were comprised of a relatively homogeneous sample of psychology students enrolled in a "more selective" university, as designated by the Carnegie Foundation (Truelove at al., 2013). This institutional distinction refers to the rigorous admissions criteria, as well as strong academic backgrounds of the students enrolled at the university.

Overall, the limited amount of empirical research on the impact of pair discussion as a component of interteaching, and the effect of group size for the pair discussion, in particular, warrants further attention. Therefore, the aim of this study was to evaluate the impact of group size during pair discussions with a more heterogeneous group of undergraduate students enrolled in an introductory psychology course at a university located in an urban area of the United States. Specifically, the research question in this study was: What is the impact of group size during interteach sessions on chapter quizzes administered following in-class discussion in dyads compared to groups of 4-5 students?

Method

Participants and Materials

Thirty-five undergraduate students aged 18-40 (M_{age} = 22, N_{male} = 15; N_{female} = 20) enrolled in an introductory psychology course provided informed consent at the end of the semester for use of their class-related information for research purposes. The course fulfilled a general education requirement and met twice per week for 75 minutes. All participants were full-time students at the time of the study and the majority (N = 25) worked at least one part-time job during the course of the semester. The majority were underclassmen ($N_{freshman}$ = 18 and $N_{sophomores}$ = 10; see Table 1) and there were representatives from 18 different majors (see Table 2).

Table 1. Demographic information for all students

	N
Sex	
Male	15
Female	20
M Age	22
M GPA	3.1
Class Standing	
Freshman	18
Sophomore	10
Junior	2
Senior	5
Employment Status	
Yes	25
No	8
No Response	2
Ethnicity	
White	28
Hispanic/Latino	1
Black/African-American	5
Prefer not to answer	1

Table 2. Declared Major for all students

	N
Business/Pre-Business/Business Management/Marketing	7
Nursing/Pre-Nursing	6
Engineering/Civil/Mechanical/Electrical Engineering	5
Exercise Science	2
Psychology	2
Accounting	1
Art Education	1
Criminal Justice	1
Dental Hygiene	1
Dietetics	1
Early Education	1
English	1
Environmental Studies	1
Fashion	1
Geography	1
Philosophy	1
Professional Writing and Editing	1
Social Work	1
	1
Total	35

The instructor for the course (the first author) created all preparation guides and made them available to students at least one week before the scheduled in-class discussion of the assigned reading material. Each prep guide reviewed 5 to 10 pages of material and included definition, conceptual, and applied questions (see Appendix 1 for a sample prep guide used in this course).

Dependent Measures and Experimental Design

We implemented an alternating treatments design whereby we placed students in a dyad or in a group of 4-5 to discuss the assigned prep guide during each class meeting. The experimenters assigned the condition for each class session in a quasi-random fashion by a coin flip before the start of the semester. We implemented a rule such that the same condition (i.e., groups of 2 vs. groups of 4-5) occurred during no more than two consecutive sessions. Assignment to work with one or more classmates was done by the instructor. Students learned of the condition in place for each class session following the clarifying lecture, immediately before the instructor assigned them to a group. The instructor made group assignments according to the student's first name, last name, color of clothing worn that day, seating arrangement (e.g., students sitting in close proximity to one another, students sitting across the room from one another). The instructor made assignments such that students had few opportunities to work with the same classmates throughout the semester. However, given the number of students and the number of prep guides,

we did not ensure that students worked with a different classmate for every single group discussion.

The primary dependent variable for the study was average performance on 15-pt. chapter quizzes. The instructor based quiz questions on the information presented in one prep guide per quiz, and required students to provide or apply a definition, recall information from the readings, or apply knowledge to novel examples. Question format included multiple-choice, fill-in-the-blank and short-answer-type questions. Ten quizzes were included as part of the data analysis, with an even number assigned to each condition.

A secondary dependent measure consisted of a brief social validity questionnaire created for the purpose of this study. The instructor distributed a brief questionnaire to the students at the end of the semester. The questionnaire included the following questions: **1)** Overall, which type of group did you like working with more? **2)** Overall, which type of group did you feel you learned the most with? Students were required to respond to each question by selecting one of the following responses: **A**) Pair Discussion (working with just one other student), **B)** Larger group (working in a group of 4-5 students), or **C**) No preference or I learned the same amount from both types of groups. We added this measure as a way for the students to provide feedback on their experience in the course and as an assessment of any possible correlations between outcome performance and responses on the self-report. Previous studies have used this or similar social validity questionnaires as a way to evaluate student perceptions of interteaching (Saville et al., 2006; Soldner et al., 2015).

Interobserver Agreement

Two independent graders scored seven of the quizzes throughout the semester (70% of total quizzes administered). The instructor omitted student names and made copies of the paper quizzes before grading to prevent any potential instructor bias. The instructor of the course served as the primary grader for each quiz. The instructor provided the secondary grader (a graduate teaching assistant for the course) with an answer key to score the quizzes assigned for interobserver agreement (IOA). We calculated IOA by summing the number of agreements and dividing by the number of agreements plus disagreements and multiplying by 100. For research purposes, IOA is "acceptable" if it is above 80% and "good" when it is above 90% (House, House, & Campbell, 1981). Mean overall IOA was 89.84% (range 83.81-95.04%).

Procedure

The procedure for this study was as follows: the students had access to the preparation guides at least one week before each due date on the Blackboard site used for the class. Students were responsible for completing each assigned prep guide before arriving to class. Upon arrival to class, the instructor delivered a clarifying lecture on the previous class session's topic based on student feedback on record sheets completed during the previous session. This lecture lasted approximately 25 minutes. The remainder of the class time was devoted to discussion on the prep guide due for that class session and completion of a record sheet and assigned quiz.

The instructor assigned students to work with either one other student (dyad) or in a group of 4-5 students. The larger group was comprised of four or five students depending on the number of students present for class on the days designated to the larger group condition. Following each interteach session, students were given approximately 5-6 minutes to complete a

record sheet to provide feedback on the quality of the discussion and ask for clarification or further help on topics they found most challenging. Students submitted the record sheets to the instructor at the end of each class session. Students received participation points for each completed record sheet. Participation points comprised 33% of the students' final grade.

The instructor used the information gathered from the record sheets to create a clarifying lecture for the next class session, as described above. Following completion of each record sheet, students were administered a 15-pt. in-class quiz. Quizzes covered only the information from the prep guide discussed during each class session. In addition, quality points were awarded to students if all members of the group received a score of 80% or higher on a quiz. This cooperative contingency resulted in two 'bonus' points added to each student's respective quiz score. If any student in the dyad or group received a score below this criterion, no student from that group received quality points. Students received immediate feedback on their performance following every quiz by delivering an answer key upon submission. This immediate feedback may enhance the value and effectiveness of quality points, as has been suggested in previous research on this topic (Rosales et al., 2014).

Results

We used an Excel workbook to enter all data for future analysis, as well as creation of figures for visual inspection, and statistical analyses. The results presented in the Figures below include scores only for quizzes administered in class. We omitted quizzes completed by students outside of class for make-up purposes from the analysis and any '0's' assigned for missing class. The instructor assigned students to work in dyads for discussion prior to delivery of quizzes 1, 3, 5, 6, and 8; and assigned students to work in groups of 4-5 for discussion prior to delivery of quizzes 2, 4, 7, 9, and 10.

Figure 1 depicts results for the average quiz performance during all class sessions. These results indicate overall performance was slightly better following discussions in dyads (M = 12.01, SD = 1.31) when compared to performance following discussions in groups of 4-5 (M = 11.17, SD = 0.47). There appeared to be an increasing trend in quiz scores following discussion in dyads (see Figure 1). The difference in average scores represents a difference between a "B" letter grade for quizzes following discussion in dyads, and "C" for quizzes following discussion in groups of 4-5.

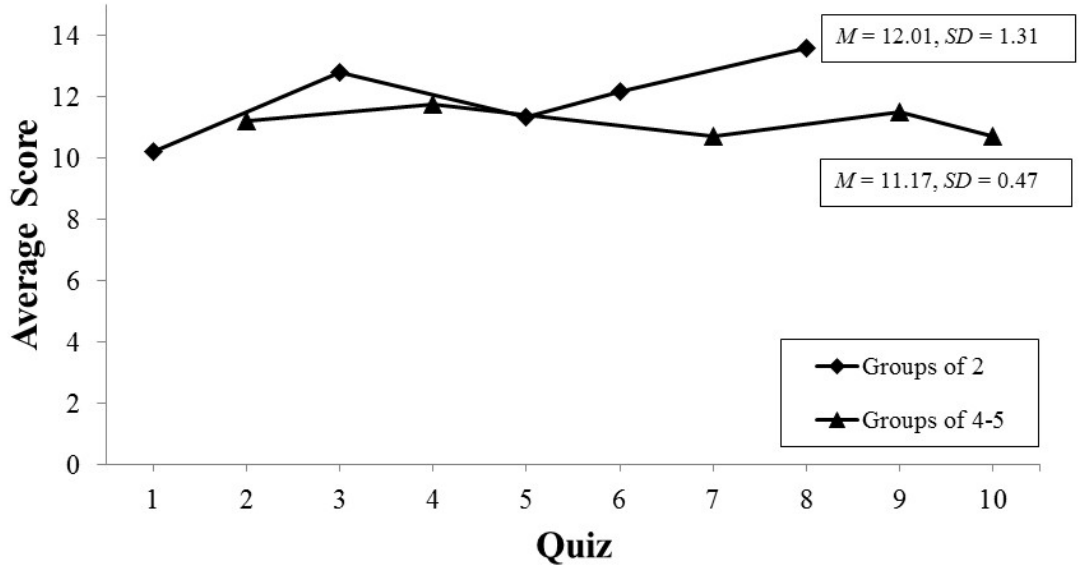

Figure 1. Average Class Quiz Performance

Figure 2 shows individual average quiz scores for all class sessions. The results indicate that 28 of the 35 students performed better on quizzes preceded by discussion in dyads. A paired-samples t-test was conducted to compare the individual score averages on quizzes preceded by a student placed in dyads and quizzes preceded by students placed in groups of 4-5. There was a significant difference in the scores for dyad discussion (M = 11.95, SD = 1.63), and discussion in groups of 4-5 (M = 11.15, SD = 1.92), t_{35} = 3.70, $p < .001$. These results confirm that students performed better on quizzes following discussion in dyads compared to discussion in groups of 4-5.

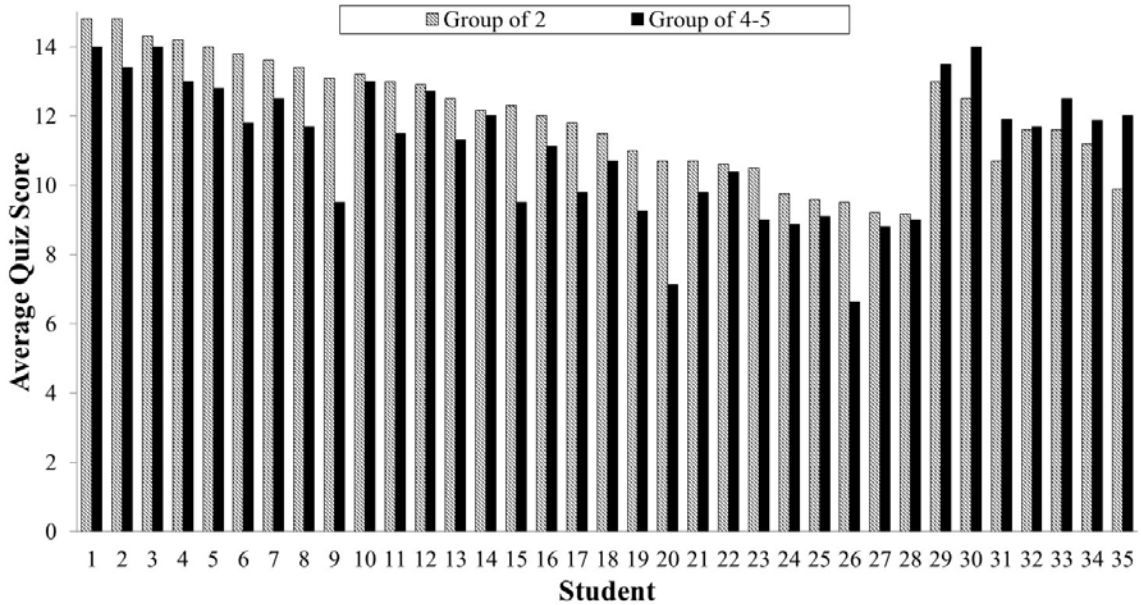

Figure 2. Average Individual Quiz Performance

Social validity measures administered at the end of the semester suggest students' overall preference was for working in larger groups, and their perceived learning was greater following discussion in larger groups (see Table 3). A common rationale provided by the students for preference and more perceived learning when working in groups of 4-5 was that when multiple students facilitated discussion with multiple opinions, it led to a better chance of achieving correct answers on prep guides. These were the subjective opinions of the students, and it did not reliably reflect the scores assigned for each quiz.

Table 3. Social Validity Data

Question	Percentage (#) of students
1. Overall, which type of group did you like working with more?	
a. Pair discussion (working with just one other student)	34.29% (12)
b. Larger group (working in a group of 4-5 students)	51.43% (18)
c. I liked both equally	14.29% (5)
2. Overall, which type of group did you feel you learned the most with?	
a. Pair discussion (working with just one other student)	25.71% (9)
b. Larger group (working in a group of 4-5 students)	51.43% (18)
c. No preference or I learned the same amount from both types of groups	22.86% (8)

Discussion

The results of this study provide support for interteaching sessions that includes discussions in dyads, as described in the seminal article on interteaching by Boyce and Hineline (2002); and are in contrast to the findings of Truelove and colleagues (2013). There are differences in the procedure of each study and in the student body demographic, that should be considered in the evaluation of these results. First, one of the primary aims of the present study was to replicate the findings of Truelove et al. with a more diverse student body. Notably, the participants in this study were from 18 different majors in the university, were mostly underclassmen, and included some non-traditional undergraduate students (age range 18-40). In contrast, students in Truelove et al. (2013) were all traditional upperclassmen Psychology majors. These differences may be of importance because students in their junior and senior college years will typically have a longer history of exposure to lecture-based classes.

Second, of the 35 student participants in this study, 43% were male. This is also a notable difference because previous research has suggested that women tend to perform better in higher education settings than men (Truelove et al., 2013). In addition, 71% were working at least part-time. Previous studies have not reported data on job status of their participants, but it is perhaps an important variable to consider for future research on interteaching. That is, if students are balancing work and school responsibilities they may benefit from specific contingencies that promote study time outside the classroom.

Third, there was a difference in the primary dependent measure. Namely, the present study evaluated the impact of group size on performance in quizzes worth a total of 15 points and covered a small amount of material (1 chapter from 1 prep guide); compared to exam probes covering 3-4 prep guides worth of class materials (Truelove et al., 2013). Fourth, we included a

quality points contingency for the present study, which was missing from the procedure outline by Truelove and colleagues. There is some evidence for the positive impact of quality points on student learning outcomes when they are included in interteaching (Rosales et al., 2014). It is possible that inclusion of quality points across the two conditions had an impact on student performance in this study. This was not part of the data analysis for this study, but future research should investigate the impact of individual and group allocation of quality points between conditions. In particular, it will be important to investigate if the inclusion of quality points between different discussion group sizes impacts student learning. For example, students may perceive that quality points are more attainable and subsequently more motivating following discussion in dyads compared to larger discussion groups (i.e., 4-5 students or more) because only one other student must meet the cooperative contingency requirement. This may affect individual and group performance in larger discussion groups.

Future studies on this topic should continue to evaluate relevant demographic and other relevant student characteristics and its impact on student learning and satisfaction. For example, Saville, Pope, Truelove, and Williams (2012) investigated the relations between student GPA performance and demographic variables on exam performance during interteaching and lecture and found that interteaching improved performance most notably for students with low and moderate GPAs. Although we did not systematically consider GPA as a variable in this study, the average GPA for students in this study was 3.1 (range of 2.0-4.0). This variable may also have an impact on performance when in-class discussions are in dyads vs. larger groups.

Future research should also address the limitations of the present study. First, the content of each chapter was not counterbalanced. The possibility that content of respective quizzes over the course of the class was not equally difficult cannot be ruled out. We attempted to control for this potential confound by randomly assigning the group size throughout the semester. However, in order to control for the possibility that the content of chapters assigned to dyad interteaching sessions were less difficult than those assigned to group discussions, counterbalancing of the two conditions needs to be conducted. Second, future studies should also attempt to replicate these procedures in online and blended courses, especially given the increased prevalence of this course format in higher education (Soldner et al., 2017). Third, we did not include a control condition. Instead, all class sessions included some interaction amongst the students present in class. Future studies may wish to replicate this work by including a third condition where students are required to independently review the content of the completed prep guide during class time before administration of a quiz.

Despite the noted limitations of this study, this study is the first to empirically support the effectiveness of interteaching using dyads during class discussion when compared to discussion in groups of 4-5 students. These results also provide initial evidence for the importance of group size as part of the pair discussion component of interteaching and have important implications for student learning and satisfaction.

Acknowledgements

We thank Mathew Swerdan, Joseph Veneziano, and Beatriz Querol Isabelle Del Rosario for assistance with data analysis.

Appendix 1. Sample Prep Guide

Prep Guide: Chapter 2

1. What are behavioral genetics and evolutionary psychology? Discuss how evolutionary psychologists use Darwin's principles of natural selection and genetic mutations to explain human behavior.
2. Describe the nature-nurture debate. What approach do most psychologists today take on this debate today?
3. Identical twin boys raised in different households demonstrated great similarity in their personalities when reunited 20 years later. Is this evidence for the predominance of *nature* or *nurture*? Be sure to explain why your answer is correct.
4. Differentiate between monozygotic and dizygotic twins. Which type of twins share the greatest genetic overlap?
5. Diagram a neuron, including all the major structures. Then describe the major function(s) of each. (NOTE: If you are not artistically talented, don't worry about it! It will help for you to create your own picture to use as you study for the review exam and quiz).
6. Briefly describe the process of neural communication between and within neurons. What role does an action potential and synaptic gap play in communication between neurons? What happens when there are leftover neurotransmitters in the synaptic gap?
7. Distinguish between agonists and antagonists by answering the following question: _____ such as Prozac mimic the effects of the neurotransmitter serotonin, whereas _____ include anti-psychotic drugs that block the action of the neurotransmitter dopamine.
8. Describe the main function of the endocrine system. Then, explain the difference between hormones and neurotransmitters. You may use the analogy of sending email messages described in your text to note this difference.
9. Differentiate between the central nervous system (brain and spinal cord) and the peripheral nervous system (autonomic and peripheral nervous systems), including the major functions of each system.
10. Describe recent research regarding neuroplasticity and neurogenesis, and the role of stem cells in the treatment of various brain dysfunctions. As you describe this research, make sure to provide definitions of each term.
11. What are the rooting and Babinski reflex, and why are these reflexes important to assess in infants?
12. What is the purpose of the hindbrain? Identify its <u>three key structures</u> and main functions of each.
13. What is the purpose of the midbrain and the forebrain? Identify the <u>key structures and main functions</u> of each structure.
14. What is the purpose of the corpus callosum and what happens when it is severed?
15. What part of a person's brain would be affected if after brain damage they could speak quickly and easily, but could not understand what they read or heard? (Wernicke's area)

References

Arntzen, E., & Hoium, K. (2010). On the effectiveness of interteaching. *The Behavior Analyst Today, 11*, 155–160. http://dx.doi.org/10.1037/h0100698

Boyce, T. E., & Hineline, P. N. (2002). Interteaching: A strategy for enhancing the user friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst, 25*, 215–226.

Chidambarum, L., & Tung, L. L. (2005). Is out of sight, out of mind? An empirical study of social loafing in technology-supported groups. *Information Systems Research, 16(2),* 149−168.

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. doi:10.1111/j.15414329.2009.00087.x

House, A.E., House, B.J., & Campbell, M.B. (1981). Measures of Interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment, 3(1),* 37-57.

Johnson, D.W., & Johnson, R.T., (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38(5),* 365-379.

Latane, B. (1981). The psychology of social impact. *American Psychology, 36*, 343−356.

Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology, 107*(2), 558- 579. http://dx.doi.org/10.1037/a0037698

Querol, B. I. D., Rosales, R., & Soldner, J. L. (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

Rosales, R., Soldner, J. L., &, Crimando, W. (2014). Enhancing the impact of quality points in interteaching. *Journal of the Scholarship of Teaching and Learning, 14*(5), 1–11. doi:10.14434/josotlv14i5.12746

Saville, B. K., Cox, T., O'Brien, S., & Vanderbelt, A. (2011). Interteaching: The impact of lectures on student performance. *Journal of Applied Behavior Analysis, 44*, 937–941. doi:10.1901/jaba.2011.44-937

Saville, B. K., Lambert, T., & Robertson, S. (2011). Interteaching: Bringing behavioral education into the 21st century. *The Psychological Record, 61*, 153–166.

Saville, B. K., Pope, D., Truelove, J., & Williams, J. (2012). The relation between GPA and exam performance during interteaching and lecture. *The Behavior Analyst Today, 13,* 27-31.

Saville, B. K., & Zinn, T. E. (2009). Interteaching: The effects of quality points on exam scores. *Journal of Applied Behavior Analysis, 42*, 369–374. doi:10.1901/jaba.2009.42-369

Saville, B. K., Zinn, T. E., & Elliot, M. P. (2005). Interteaching versus traditional methods of instruction: A preliminary analysis. *Teaching of Psychology, 32*, 161–163. doi:10.1207/s15328023top3203_6

Saville, B. K., Zinn, T. E., Neef, N. A., Van Norman, R., & Ferreri, S. J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39(1)*, 49–61. doi:10.1901/jaba.2006.42-05

Scoboria, A., & Pascual-Leone, A. (2009). An 'interteaching' informed approach to instructing large undergraduate classes. *Journal of the Scholarship of Teaching and Learning, 9(3),* 29-37.

Slavin, R. E. (2014). Cooperative learning and academic achievement: Why does group work work? *Annals of Psychology, 30*(4), 785-791. http://dx.doi.org/10.6018/analesps.30.3.201201

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43-69.

Soldner, J. L., Rosales, R., & Crimando, W. (2015). A comparison of interteaching and classroom lecture in rehabilitation education. *Rehabilitation Counselors and Educators Journal, 8*(1), 91-100.

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. doi: 10.1891/2168-6653.31.4.372

Truelove, J. C., Saville, B. K., & Van Patten, R. (2013). Interteaching: Discussion group size and course performance. *Journal of the Scholarship of Teaching and Learning, 13 (2)*, 23–30.

Zayak, R.M., & Paulk, A.L. (2014). Interteaching: Its effects on exam scores in a compressed schedule format. Journal of the Scholarship of Teaching and Learning, 14(1), 1-12. doi:10.14434/josotl.v14i1.3649