Special Sections: Opportunities and Challenges of Online Instruction

JMBE Journal of Microbiology & Biology Education

Verbalized Studying and Elaborative Interrogation in the Virtual Classroom: Students with Social Anxiety Prefer Working Alone, but Working with a Peer Does Not Hurt Their Learning

Rachel Tomco Novak,[a,b] Elizabeth G. Bailey,[a] Bethany D. Blinsky,[b] Burke W. Soffe,[b] David Patterson,[a] Jordon Ockey,[a] and Jamie L. Jensen[a]

[a]*Department of Biology, Brigham Young University, Provo, Utah, USA*

[b]*College of Dental Medicine, Roseman University of Health Sciences, South Jordan, Utah, USA*

Due to public health measures enacted in response to the Covid-19 pandemic, educators and students alike have been suddenly thrust into the realm of online learning. To better understand how active and collaborative learning methods can apply to students studying in isolation, we compared the effects of two teach-and-question assignments: one that utilizes the active learning method of reciprocal peer tutoring and a solo version that requires individual verbalized studying and elaborative interrogation. We used a quasi-experimental design, with student participants enrolled in an online introductory human anatomy course. The first treatment group completed regular teach-and-question study assignments virtually with a peer, and the second treatment group completed the same assignment independently. We found no differences in exam scores between treatments, even for students with high social anxiety; however, student attitudes about the social versus individual assignment did differ for specific types of students. Students who reported experiencing high social anxiety preferred completing the active learning exercise by themselves, and students with low scientific reasoning ability preferred the partnered assignment. This research has potential implications for online classrooms. For instance, our results indicate that students who study independently, or in isolation, may have learning outcomes similar to those of students who study with a peer as long as they study actively. Because we found no negative impact on examination results, it also could be that virtually partnered or independent teach-and-question assignments could be helpful for instructors teaching large online classes to ensure all students are getting individualized feedback and attention.

KEYWORDS reciprocal peer tutoring, verbalized studying, social anxiety, elaborative interrogation, online instruction, active learning, collaborative learning, scientific reasoning ability

INTRODUCTION

In the wake of the first United States Covid-19 restrictions on gathering, higher-education programs across the nation abruptly pivoted from an in-person educational model to emergency remote teaching, with some classes being held live on virtual conferencing software or through asynchronous recordings (1, 2). Due to these precautions, many students suddenly began studying in isolation, potentially even thousands of miles from their classmates and schools (1). This led to a myriad of pedagogical challenges for teachers in higher education, including questions about how to include active and collaborative studying and learning methods within the virtual learning environment (3).

The benefits of active and collaborative learning methods are well established (4, 5). One such method of active and collaborative learning is reciprocal peer tutoring, which is when two students, who are often enrolled in the same course, take turns explaining course content to one another to further their own understanding and knowledge (6). Reciprocal peer tutoring has been found to be an effective method of learning and studying in pharmacology courses, physiology courses, and anatomy courses (7). It is also widely used in medical schools and nursing programs (6).

In a large introductory biology course, Bailey et al. found that reciprocal peer tutoring in the form of a teach-and-question assignment (TQ) increased learning gains for students and improved examination scores (8). In the TQ assignment, one student acts as a tutor and explains the course content

Address correspondence to Department of Biology, Brigham Young University, Provo, Utah, USA. E-mail: rnovak@roseman.edu.
The authors declare no conflict of interest.
Received: 17 September 2021, Accepted: 30 November 2021, Published: 4 April 2022

to a peer from memory, and the peer inquiries about the tutor's content knowledge and asks further questions to probe their understanding (8). In this way, TQ is used as a studying method between students outside of initial content attainment from the instructor or teaching assistant. In its simplest understanding, the TQ assignment is ultimately the combination between the active learning methods of verbalized studying and elaborative interrogation. Verbalized studying, or thinking aloud, is the simple effort of speaking out loud during the learning process (9). Elaborative interrogation is the process of asking "how" or "why" to increase active learning (10, 11). Both of these methods have been shown to lead to a higher level of understanding and retention and help the learner make connections that they may not otherwise create through just reading (9–13).

Collaborative learning methods, such as reciprocal peer tutoring, can be effective at helping to further learning gains for students, as well as allowing students to practice social and emotional skills and giving students space to build relationships with peers (6, 7). However, despite these positive reported effects, some students may feel uncomfortable with the social situations encountered during these learning environments (14–17). For example, some students with social discomfort in group learning situations showed lower cognitive awareness in group activities and during these activities may feel apprehension, stress, and fear (18). Due to these factors, collaborative learning methods may not offer the same learning benefits for all populations of students (14–17).

The Covid-19 pandemic has been challenging, destructive, and detrimental in countless ways, and the field of education has certainly felt the impact of this weight (3). To assist educators in implementing active and collaborative learning and studying strategies virtually, this research seeks to understand if the positive effects of reciprocal peer tutoring can be experienced by students who are studying online, potentially isolated from their classmates. Specifically, we investigated the outcomes of a series of TQ assignments that require students to utilize verbalized studying (the tutoring portion of the assignment) and elaborative interrogation (the questioning portion of the assignment) to build their knowledge. In this study, some sections of the course were assigned to complete the TQ assignments remotely with a peer and others were assigned to complete the assignment independently. All students in the study, regardless of treatment, were required to speak and describe course learning objectives out loud, ask higher-order questions out loud, and upload an audio recording of each assignment to the learning management system. This study also seeks to understand if outcomes from the TQ assignment are different for certain populations of students, based on factors such as extroversion, social anxiety, and scientific reasoning ability.

The research questions guiding this study are as follows.

1. Does reciprocal peer tutoring increase student learning above verbalized independent studying? Do certain student characteristics predict greater benefit from reciprocal peer tutoring (extroversion, scientific reasoning, social anxiety)?
2. Does reciprocal peer tutoring increase student attitudes above verbalized independent studying? Do certain student characteristics predict greater attitudes toward reciprocal peer tutoring (extroversion, scientific reasoning, social anxiety)?

METHODS

Ethics statement

The primary author's Institutional Review Board reviewed and approved the study protocol (IRB2020-467). Written consent was obtained from all participants in this study.

Participants and context

Participants in this study were undergraduate college students enrolled in one of 11 course sections of a 200-level human anatomy course for pre-health science majors at a large private university in 2021. In total, 189 students enrolled in the course, and 167 students gave written consent to participate in this study. Typically, this course would meet once a week for a 2-h class in a large lecture hall. However, due to the public health measures enacted in response to Covid-19 in early 2021, the lectures were prerecorded and students watched the videos asynchronously. All activities included in this study were required of every student in the class regardless of study participation; moreover, no additional activities were given to students involved in this study.

Experimental design

To answer our research questions, we utilized a quasi-experimental design. Students self-selected their section, and then we randomly assigned each of the 11 sections to a treatment. Six sections were assigned to the teach-and-question with a peer (TQ-P) treatment group and five sections were assigned to the teach-and-question independently (TQ-I) treatment group (see Fig. 1). During the semester, students completed one to two TQ assignments each week of the class, for a total of 23 TQ assignments overall. Aside from the difference in assignment type, virtually all other course characteristics were identical among treatment groups (instructor, exams, learning outcomes, textbook, lab assignments).

At the beginning of the semester, students took a presurvey assessing multiple factors, including demographics, interest in anatomy, scientific reasoning abilities, belongingness in the sciences, communication and social anxiety, and extroversion. In the survey, students were asked a single Likert-style question to gauge their interest in anatomy. Because interest was assessed using a single item, the conclusions we can draw about interest changes are more limited than our other variables. To test students' scientific reasoning abilities, students took the 24-item Lawson's classroom test of scientific reasoning (LCTSR) that has previously been validated in college student populations (19, 20).

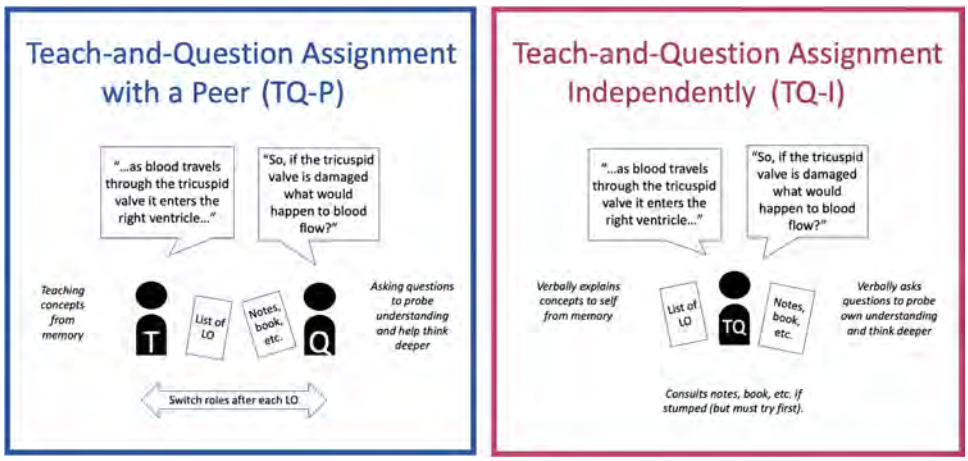

FIG 1. Description of treatment groups, TQ-P and TQ-I.

To assess belongingness, the students took a 7-item survey of students' belongingness in the sciences using a 5-point Likert scale adapted from Good et al. (21). Social and communication anxiety was assessed through a 12-item survey using a 5-point Likert scale adapted from McCroskey and from Papanastasiou and Zembylas (22, 23). To assess extroversion, students took a 10-item survey of extroversion, a portion of the big-five personality test, using a 5-point Likert scale from Goldberg (24). Responses for belongingness (7 items), anxiety (12 items), and extroversion (10 items) were summed in their respective categories and treated as continuous variables. At the end of the semester, students took a postsurvey that included Likert-style attitudinal questions about the teach-and-question (TQ) assignment and reanswered the interest in anatomy and belongingness in the sciences questions.

Teach-and-question with a peer treatment group

In the TQ-P assignment (see Fig. 1), students began by watching a segment of a 50-minute prerecorded lecture with a peer through video conferencing software. Approximately every 10 minutes in the lecture video, a slide would appear on the students' screens with learning outcomes. The instructor then prompted students to pause the recorded lecture to participate in the TQ portion of the class. Students began by deciding who would begin as the tutor and who would be the questioner in the assignment. It was the duty of the tutor to instruct their partner on a learning objective from the lecture solely from their own memory, and the duty of the questioner was to investigate the knowledge of the tutor, specifically by inquiring and seeking further information on the content that the tutor was explaining. The purpose of the assignment was for the questioner to help guide the tutor to aspects of the learning objective that they might not fully understand and to discuss those together. After the objective was fully reviewed, the students would then watch the next 10 minutes of the video together until it was time to do the next TQ portion, with the former questioner acting as the tutor.

Teach-and-question independently treatment group

In the TQ-I assignment (see Fig. 1), students watched the same 50-minute prerecorded lecture used in the TQ-P treatment, but unlike those in the TQ-P treatment, students in the TQ-I treatment watched independently. At the same 10-minute intervals of the lecture, a slide would appear on their screen with outcomes that the student was to review. The student would then begin by explaining the content of the learning objectives out loud from their own memory as if there was an audience or peer in the room. To mimic the questioner in TQ-P, the student in TQ-I would inquire and interrogate their own understanding of the content that they just taught aloud. After this sequence was completed, the student would then watch the next 10 minutes of the video until it was time to do the next TQ-I portion and repeat the process over.

TQ assignment grading

For both TQ-I and TQ-P assignments, the entire session was audio-recorded and uploaded to the learning management system for credit. Teaching assistants would listen to portions of each recording and provide feedback to the students to help improve the quality of teaching and questioning. However, the recordings were ultimately graded only for completion. Thus, we did not gather data on the quality of the TQ sessions. We had hoped to record data about the length of the TQ session. However, some students recorded only when they were talking, and others recorded the whole time they were watching the lecture. Thus, the lengths of the audio recordings did not always represent the lengths of the TQ sessions.

Statistical analysis

The following assumptions of each linear regression were met: linearity, normality, equality of variance, and multicollinearity. Due to the nested nature of the data (students were not independent; they were grouped in sections and

TABLE 1
Equivalency of treatment groups, TQ-P and TQ-I

Variable	TQ-P		TQ-I		Test	*P* value	Cohen's d
	Mean ± SD	*n*	Mean ± SD	*n*			
Yr in school	1.930 ± 0.869	68	1.700 ± 0.796	77	Mann-Whitney U	0.114	0.264
Reasoning (LCTSR)	17.00 ± 4.729	73	17.46 ± 3.787	70	Ind. samples T	0.526	0.106
Interest-pre	3.889 ± 0.920	87	3.848 ± 1.014	79	Mann-Whitney U	0.952	0.009
Belongingness-pre	3.957 ± 0.576	80	3.996 ± 0.605	78	Ind. samples T	0.492	0.066
Social anxiety	3.197 ± 0.420	80	3.205 ± 0.446	78	Ind. samples T	0.797	0.019
Extroversion	3.109 ± 0.832	80	3.027 ± 0.909	78	Ind. samples T	0.704	0.094

sections were grouped into treatments), we considered using mixed modeling with a random effect for course section to account for nesting. We used two methods to test whether such a random effect was needed in regressions predicting exam performance (as suggested in Theobald 2018) (25). First, we calculated the intraclass correlation coefficient (ICC) of the course section random effect in an empty model with no fixed effects. We found an ICC of 0.085, suggesting that the nesting by section did not explain much variance in exam performance. To confirm this, we compared complete models (including all fixed effects of interest) with and without the random effect of course section. Next, we calculated the Akaike information criterion (AICc). Because adding the random effect (AICc = 927.27) did not improve the model compared to the full model without the random effect (AICc = 925.55), we moved forward with standard multiple linear regression without random effects. We also tested whether accounting for nesting by section was needed in regression models predicting overall assignment attitudes. Similarly, a low ICC (ICC = 0.0) and inability of the random effect to improve the model (AICc = 612.00 compared to AICc = 614.07) led us to use standard multiple linear regression to predict attitudes.

RESULTS

Equivalency of treatment groups

Due to the quasi-experimental nature of our study, we tested whether our treatment groups were equivalent at the beginning of the semester. As shown in Table 1, the two treatment groups, TQ-P and TQ-I, were statistically equivalent at the beginning of the course in terms of year in school, scientific reasoning ability, anatomy interest, sense of belongingness in the sciences, level of communication and social anxiety, and level of extroversion.

Research question 1: assessment outcomes

We used exam scores to compare learning outcomes of TQ-P and TQ-I. Throughout the semester, six exams were administered, and all of the exams were identical for both treatment groups. Figure 2A shows scores for both treatments on each exam given. By split-plot analysis of variance (ANOVA), we saw no significant difference between treatment groups ($P = 0.745$; $\eta_p^2 = 0.001$) and no interaction between treatment and time ($P = 0.343$; $\eta_p^2 = 0.006$). Figure 2B shows the students' average exam percent score, with scores from all six exams averaged together. The mean performance was compared using an independent samples *t* test, and we again found no difference by treatment ($P = 0.505$; Cohen's d = 0.104).

Research question 1: predictors of assessment outcomes

We next wanted to see if the treatment helped specific populations of students. To determine which variables predicted average exam percent score, we performed a multiple linear regression analysis. As shown in Table 2, treatment still had no effect on exam scores when other predictors were also included. Students' scientific reasoning ability (LCTSR score) was the only significant predicter of average exam percent score. We also tested whether adding interactions between treatment and LCTSR, social anxiety, and extroversion improved the model (to see if our treatment helped specific student populations). As shown in Appendix 1, adding in these interactions did not improve the model (Table S1).

Research question 2: attitudinal outcomes

Attitudinal data were gathered at the end of the semester to investigate student perceptions of both TQ assignments, as well as other aspects of the course. Students in both the TQ-P and

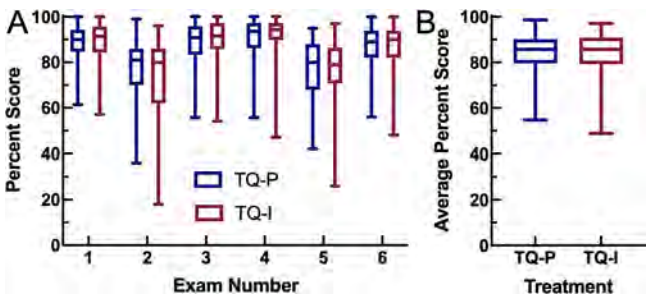

FIG 2. Assessment outcomes. (A) Students' scores on six course exams are compared by treatment groups. (B) Average exam percent scores are compared by treatment. Scores for all six course exams were averaged together for each student.

TABLE 2
Results of multiple linear regression with average exam percent score as target[a]

Variable	Unstandardized B	Coefficient std. error	Standardized coefficient beta	t	Significance
Constant	79.442	12.827		6.193	<0.001
Reasoning (LCTSR)	0.745	0.215	0.310	3.470	<0.001[b]
TQ-I	1.224	1.687	0.064	0.726	0.469
Yr in school	0.760	1.123	0.064	0.677	0.500
Female	−1.523	1.975	−0.075	−0.771	0.442
Social anxiety	−2.135	2.658	−0.095	−0.803	0.423
Extroversion	−0.516	1.244	−0.048	−0.415	0.679

[a]Adjusted $R^2 = 0.061$.
[b]Suggests $P < 0.05$.

TQ-I treatments reported that they would use TQ as a method to study in future courses (73% of all students) and preferred the role of "teacher" over that of "questioner" (87% of all students). Our data hint that students with high social anxiety or low sense of belonging may prefer the questioner role (see Appendix 1, Table S2), but we are unsure how reproducible this difference would be since only 15 students preferred the role of questioner. Both groups reported an increased sense of belongingness ($P = 0.002$; Cohen's d = 0.254) and interest in anatomy ($P < 0.001$; Cohen's d = 0.674) over the course, tested by paired-samples *t* test. Change in belongingness and change in interest in anatomy are shown in Fig. 3A. A split-plot, repeated measures ANOVA was run to see if there was a statistical difference between the treatments in relationship to the pre and post surveys. For sense of belongingness, we saw no significant difference between treatment groups ($P = 0.753$; $\eta_p^2 = 0.001$) and no interaction between treatment and time ($P = 0.472$; $\eta_p^2 = 0.003$). For interest in anatomy, we saw no significant difference between treatment groups ($P = 0.613$; $\eta_p^2 = 0.002$) and no interaction between treatment and time ($P = 0.804$; $\eta_p^2 \leq 0.001$). As seen in Fig. 3B, the students were asked three questions to determine their overall attitude toward the teach-and-question assignment using a 5-point Likert scale. Treatment groups were compared via Mann-Whitney U-test. There was no statistical difference in attitudes toward the assignment between the two treatment groups for any of the three questions: liked assignment ($P = 0.436$; Cohen's d = 0.121), worth time ($P = 0.774$; Cohen's d = 0.044), and comfort with assignment ($P = 0.248$; Cohen's d = 0.180).

Research question 2: predictors of attitudinal outcomes

Again, we were also interested in whether the two treatments affected students differently based on student characteristics such as social anxiety, extroversion, and scientific reasoning ability. To estimate overall attitude toward the TQ assignment, we summed students' answers to the three questions of Fig. 3B, and we used linear regression to predict this overall attitude. We first compared two regression models to predict overall attitude: one without interactions between treatment and student characteristics and the second one with these interactions. As shown in Appendix 1 (Table S3), including interactions with treatment significantly improved the model. Thus, the full model with interactions is shown in Table 3. We found that students with high social anxiety preferred the TQ-I assignment (interaction between treatment and social anxiety: $P = 0.004$), and this interaction is shown visually in Fig. 4A. Students with low scientific reasoning ability at the beginning of the course preferred the TQ-P assignment, and students with high scientific reasoning ability preferred the TQ-I assignment (interaction between treatment and LCTSR: $P = 0.012$), as shown visually in Fig. 4B.

DISCUSSION AND IMPLICATIONS FOR INSTRUCTORS

An aim of this study was to find out whether or not the collaborative and social aspect of TQ is needed for students

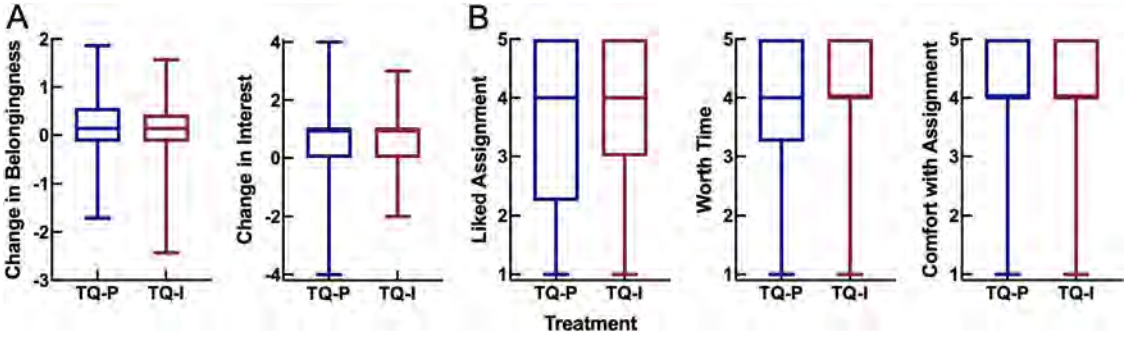

FIG 3. Attitudinal outcomes. (A) Change in belongingness and change in interest, compared by treatment group. (B) Liked assignment, worth time, and comfort with assignment, by treatment group.

TABLE 3
Results of multiple linear regression with overall attitudinal score as target[a]

Variable	Unstandardized B	Coefficient std. error	Standardized coefficient beta	t	Significance
Constant	31.431	4.874		6.448	<0.001
Reasoning (LCTSR)	−0.207	0.089	−0.295	−2.333	0.021[b]
TQ-I	0.911	0.459	0.162	1.984	0.050[b]
Yr in school	−0.281	0.307	−0.081	−0.916	0.362
Female	−0.290	0.541	−0.049	−0.536	0.593
Social anxiety	−4.350	1.006	−0.666	−4.323	<0.001[b]
Extroversion	−0.526	0.477	−0.168	−1.103	0.272
LCTSR*Tx	0.296	0.116	0.310	2.544	0.012[b]
Social anxiety*Tx	4.117	1.394	0.447	2.953	0.004[b]
Extroversion*Tx	0.647	0.678	0.141	0.954	0.342

[a]Adjusted $R^2 = 0.188$. Tx, treatment.
[b]Suggests $P < 0.05$.

to experience learning benefits or if the learning benefits can be brought about through an independent activity. We found that neither the TQ-P nor the TQ-I participant group performed better on examinations than the other (Fig. 2). In contrast, Bailey et al. found that participants in a TQ treatment had ~6% higher exam scores compared to participants who studied on their own (8). However, in that study, the students who studied individually were allowed to study on their own in any way they chose. Although Bailey et al. did not track what study strategies students used, previous studies suggest that the majority of college students primarily use passive strategies such as rereading or watching material (26). In contrast, we compared the reciprocal peer tutoring treatment (TQ-P) to an individual assignment in which students were required to verbalize and generate questions (TQ-I) and did not find the learning gains reported previously (8). This may imply that the social aspect of the TQ method may not be necessary if students still verbalize and ask questions. This suggests that educators could instruct students to complete TQ assignments independently without decreasing students' learning gains. Alternatively, it is possible that the benefit of a partnered assignment would be more social than academic. However, we saw no difference in students' change in belonging between the treatments (see Fig. 3A). It may be that there were social benefits from the partnered assignment that we did not quantify. Anecdotally, teaching assistants noticed that there were more off-topic conversations aimed at relationship building in the TQ-P group.

Unlike past studies, we did not see decreased performance in the social treatment (TQ-P) for students that reported high social anxiety (Table 2 and Appendix 1: Table S1) (14–17). However, we did find that students with high social anxiety preferred the TQ-I assignment over the TQ-P assignment (see Table 3 and Fig. 4A). Because we found no difference in exam results between the two treatments, it may be helpful for those who experience social anxiety to be given the option to engage independently in active learning methods to help increase their attitude toward the assignment. This may be important, as it has been demonstrated that forcing students

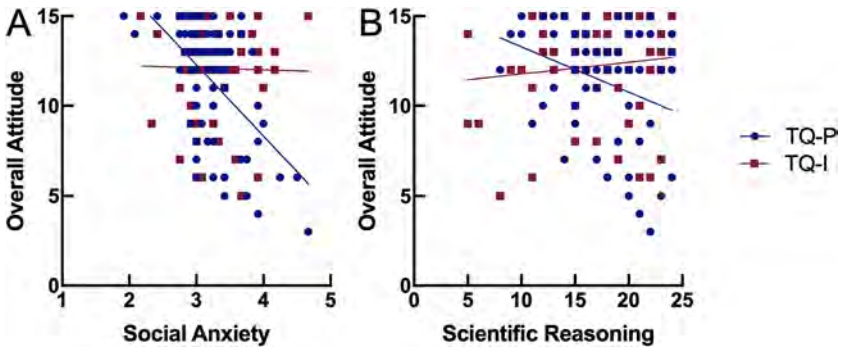

FIG 4. Attitudinal interactions between treatment and social anxiety or scientific reasoning. For both panels, overall attitude was calculated by adding together the three Likert-style attitudinal questions on the postsurvey, which is shown on the *y* axis. The *x* axis shows social anxiety score from pretest (A) or scientific reasoning (B; LCTSR score from pretest). Raw data are shown by treatment, and lines are from simple linear regression to show trends.

with social anxiety to participate in group activities can negatively affect their ability to learn (27–29). However, our study found no interaction between treatment and social anxiety on tests scores.

We also found an interesting interaction between treatment and scientific reasoning ability (Table 3): students with low scientific reasoning abilities preferred the TQ-P assignment, and students with high scientific reasoning abilities preferred the TQ-I assignment (see Fig. 4B). Although this interaction predicted attitudes rather than performance, our results parallel those of Bailey et al., who found that students with lower scientific reasoning ability had greater learning gains from the TQ assignment than those with high scientific reasoning ability (8). Together, these two studies support the idea that implementing a partnered TQ assignment could help increase course scaffolding for students who may need additional learning and studying support.

In summary, we found that an active learning assignment that requires verbalization and questioning leads to similar positive attitudes (Fig. 3B) and exam performance (Fig. 2) whether done in partnerships or as individuals. This makes the individual TQ assignment a simple and effective active learning option for students with high social anxiety who prefer working alone (Fig. 4A) or for all students in virtual learning environments when group work is less feasible. However, partnerships can be formed even when students learn remotely, and this might be especially beneficial for students who enter the classroom with poor scientific reasoning skills (Fig. 4B).

Limitations and future research

It is important to note that the course utilized in this research was not originally intended to be delivered in a virtual learning format and was delivered in this manner only as a safety measure in response to the Covid-19 pandemic. Thus, our results reflect emergency remote teaching rather than classic online learning, and further research on this topic could focus on courses intended to be delivered virtually. Further research should also examine if implementing these learning methods in person, as opposed to virtually, could have any different outcomes. It is possible that partnered assignments completed online, often in the comfort of students' homes, do not elicit the same social anxiety that would be triggered in a classroom with more time constraints and social pressure. Moreover, the need for collaboration in addition to verbalization and questioning should also be investigated with different types of student populations, such as in a university with open enrollment, in different disciplines, or with graduate students. Future studies may also seek to compare three treatments simultaneously: teach and question with a peer, teach and question individually, and individual studying with no mandated teach and question assignment. Finally, it would be interesting to study the effects of homogeneous versus heterogeneous pairs in terms of scientific reasoning. This study could be repeated with purposeful and consistent pairings rather than allowing students to self-select and change their partners throughout the course.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

We thank all the teaching assistants involved in the undergraduate anatomy class in which this research took place. Their work on grading, reviewing, and assisting with the management of the class was imperative for the success of this project, and we commend and sincerely appreciate their diligence and hard work. We also thank the students who participated in this study.

REFERENCES

1. Quintana C. 2020. 'The virus beat us': colleges are increasing going online for fall 2020 semester as COVID-19 cases rise. USA Today, McLean, Virginia.
2. Hodges C, Moore S, Lockee B, Trust T, Bond A. 2020. The difference between emergency remote teaching and online learning. Educause Review, Boulder, CO.
3. Yusuf B. 2020. Are we prepared enough? A case study of challenges in online learning in a private higher learning institution during the Covid-19 outbreaks. Adv Soc Sci Res J 7:205–212. https://doi.org/10.14738/assrj.75.8211.
4. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, Wenderoth MP. 2014. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 111:8410–8415. https://doi.org/10.1073/pnas.1319030111.
5. Haak DC, HilleRisLambers J, Pitre E, Freeman S. 2011. Increased structure and active learning reduce the achievement gap in introductory biology. Science 332:1213–1216. https://doi.org/10.1126/science.1204820.
6. Gazula S, McKenna L, Cooper S, Paliadelis P. 2017. A systematic review of reciprocal peer tutoring within tertiary health profession education programs. Health Prof Educ 3:64–78. https://doi.org/10.1016/j.hpe.2016.12.001.
7. Shenoy A, Petersen KH. 2020. Peer tutoring in preclinical medical education: a review of the literature. Med Sci Educ 30:537–544. https://doi.org/10.1007/s40670-019-00895-y.
8. Bailey EG, Baek D, Meiling J, Morris C, Nelson N, Rice NS, Rose S, Stockdale P. 2018. Learning gains from a recurring "teach and question" homework assignment in a general biology course: using reciprocal peer tutoring outside class. CBE Life Sci Educ 17:ar23. https://doi.org/10.1187/cbe.17-12-0259.
9. Ramachandran A, Huang CM, Gartland E, Scassellati B. 2018. Thinking aloud with a tutoring robot to enhance learning, p 59–68. *In* ACM/IEEE international conference on human-robot interaction. Association for Computing Machinery, New York, NY. https://doi.org/10.1145/3171221.3171250.

10. Clinton V, Alibali MW, Nathan MJ. 2016. Learning about posterior probability: do diagrams and elaborative interrogation help? J Exp Educ 84:579–599. https://doi.org/10.1080/00220973.2015.1048847.
11. Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT. 2013. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. Psychol Sci Public Interest 14:4–58. https://doi.org/10.1177/1529100612453266.
12. Forrin ND, MacLeod CM. 2018. This time it's personal: the memory benefit of hearing oneself. Memory 26:574–579. https://doi.org/10.1080/09658211.2017.1383434.
13. Luckie DB, Rivkin AM, Aubry JR, Marengo BJ, Creech LR, Sweeder RD. 2013. Verbal final exam in introductory biology yields gains in student content knowledge and longitudinal performance. Life Sci Educ 12:515–529. https://doi.org/10.1187/cbe.12-04-0050.
14. Cohen M, Buzinski SG, Armstrong-Carter E, Clark J, Buck B, Reuman L. 2019. Think, pair, freeze: the association between social anxiety and student discomfort in the active learning environment. Scholarsh Teach Learn Psychol 5:265–277. https://doi.org/10.1037/stl0000147.
15. England BJ, Brigati JR, Schussler EE. 2017. Student anxiety in introductory biology classrooms: perceptions about active learning and persistence in the major. PLoS One 12:e0182506. https://doi.org/10.1371/journal.pone.0182506.
16. Hood S, Barrickman N, Djerdjian N, Farr M, Magner S, Roychowdhury H, Gerrits R, Lawford H, Ott B, Ross K, Paige O, Stowe S, Jensen M, Hull K. 2021. "I like and prefer to work alone": social anxiety, academic self-efficacy, and students' perceptions of active learning. CBE Life Sci Educ 20:art12. https://doi.org/10.1187/cbe.19-12-0271.
17. England BJ, Brigati JR, Schussler EE, Chen MM. 2019. Student anxiety and perception of difficulty impact performance and persistence in introductory biology courses. CBE Life Sci Educ 18:ar21. https://doi.org/10.1187/cbe.17-12-0284.
18. Cantwell RH, Andrews B. 2002. Cognitive and psychological factors underlying secondary school students' feelings towards group work. Educ Psychol 22:75–91. https://doi.org/10.1080/01443410120101260.
19. Lawson AE. 1978. The development and validation of a classroom test of formal reasoning. J Res Sci Teach 15:11–24. https://doi.org/10.1002/tea.3660150103.
20. Lawson AE, Alkhoury S, Benford R, Clark BR, Falconer KA. 2000. What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. J Res Sci Teach 37:996–1018. https://doi.org/10.1002/1098-2736(200011)37:9%3C996::AID-TEA8%3E3.0.CO;2-J.
21. Good C, Rattan A, Dweck CS. 2012. Why do women opt out? Sense of belonging and women's representation in mathematics. J Pers Soc Psychol 102:700–717. https://doi.org/10.1037/a0026659.
22. McCroskey JC. 1982. An introduction to rhetorical communication, 4th ed. Prentice-Hall, Englewood Cliffs, NJ.
23. Papanastasiou EC, Zembylas M. 2004. Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. Int J Science Education 26:259–280. https://doi.org/10.1080/0950069022000038277.
24. Goldberg LR. 1992. The development of markers for the Big-Five factor structure. Psychol Assess 4:26–42. https://doi.org/10.1037/1040-3590.4.1.26.
25. Theobald E. 2018. Students are rarely independent: when, why, and how to use random effects in discipline-based education research. CBE Life Sci Educ 17:rm2. https://doi.org/10.1187/cbe.17-12-0280.
26. Karpicke JD, Butler AC, Roediger HL, III. 2009. Metacognitive strategies in student learning: do students practise retrieval when they study on their own? Memory 17:471–479. https://doi.org/10.1080/09658210802647009.
27. Masia Warner C, Fisher PH, Shrout PE, Rathor S, Klein RG. 2007. Treating adolescents with social anxiety disorder in school: an attention control trial. J Child Psychol Psychiatry 48:676–686. https://doi.org/10.1111/j.1469-7610.2007.01737.x.
28. Smits JAJ, Rosenfield D, McDonald R, Telch MJ. 2006. Cognitive mechanisms of social anxiety reduction: an examination of specificity and temporality. J Consult Clin Psychol 74:1203–1212. https://doi.org/10.1037/0022-006X.74.6.1203.
29. Hayes SA, Hope DA, Heimberg RG. 2008. The pattern of subjective anxiety during in-session exposures over the course of cognitive-behavioral therapy for clients with social anxiety disorder. Behav Ther 39:286–299. https://doi.org/10.1016/j.beth.2007.09.001.