

# Learning Gains from a Recurring “Teach and Question” Homework Assignment in a General Biology Course: Using Reciprocal Peer Tutoring Outside Class

E. G. Bailey,<sup>1\*</sup> D. Baek,<sup>2</sup> J. Meiling,<sup>2</sup> C. Morris,<sup>3</sup> N. Nelson,<sup>1</sup> N. S. Rice,<sup>2</sup> S. Rose,<sup>2</sup> and P. Stockdale<sup>2</sup>

<sup>1</sup>Department of Biology, <sup>2</sup>Department of Physiology and Developmental Biology, and <sup>3</sup>Department of Spanish and Portuguese, Brigham Young University, Provo, UT 84602

## ABSTRACT

Providing students with one-on-one interaction with instructors is a big challenge in large courses. One solution is to have students interact with their peers during class. Reciprocal peer tutoring (RPT) is a more involved interaction that requires peers to alternate the roles of “teacher” and “student.” Theoretically, advantages for peer tutoring include the verbalization and questioning of information and the scaffolded exploration of material through social and cognitive interaction. Studies on RPT vary in their execution, but most require elaborate planning and take up valuable class time. We tested the effectiveness of a “teach and question” (TQ) assignment that required student pairs to engage in RPT regularly outside class. A quasi-experimental design was implemented: one section of a general biology course completed TQ assignments, while another section completed a substitute assignment requiring individuals to review course material. The TQ section outperformed the other section by ~6% on exams. Session recordings were coded to investigate correlation between TQ quality and student performance. Asking more questions was the characteristic that best predicted exam performance, and this was more predictive than most aspects of the course. We propose the TQ as an easy assignment to implement with large performance gains.

## INTRODUCTION

One of the biggest challenges in undergraduate biology courses is enabling students to receive one-on-one interaction with the professor. Even if instructors have the luxury of a teaching assistant for every 30 students, this still leaves a high student-to-instructor ratio in large-enrollment courses. Large class sizes and high student-to-teacher ratios encourage a reversion back to traditional lecture-style classes, which are less interactive and promote surface understanding rather than deep student learning. Countless studies have shown that pure lecture courses and passive learning generally do not generate the learning gains that are possible in biology courses in which active learning is encouraged (Handelsman *et al.*, 2004; Wood, 2009; Haak *et al.*, 2011; Freeman *et al.*, 2014).

Because class size is often not under the control of the instructor, pedagogical techniques that increase one-on-one interaction are in high demand. One common solution is the use of personal response systems and think-pair-share activities. Having students discuss content and solve problems together in class, when implemented correctly, has been shown to increase student learning and performance (Smith *et al.*, 2009; Vickrey *et al.*, 2015). Small-group learning during class has also been shown to help students transfer knowledge to novel contexts (Pai *et al.*, 2015).

More formal peer tutoring is also a common practice that has received a lot of attention in education research due to its positive benefits for tutees at many levels of

Jennifer Momsen, *Monitoring Editor*

Submitted December 8, 2017; Revised January

23, 2018; Accepted January 24, 2018

CBE Life Sci Educ June 1, 2018 17:ar23

DOI:10.1187/cbe.17-12-0259

\*Address correspondence to: E. G. Bailey (liz.bailey@georgetown.edu).

© 2018 E. G. Bailey *et al.* CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

education (Cohen *et al.*, 1982; Topping, 1996). Many undergraduate programs that use peer tutors employ students who performed well in the course in previous semesters. One such program in an introductory biology course observed a 1% increase on exams per tutoring session (up to 10) for tutored students compared with those who declined tutoring (Batz *et al.*, 2015). Another study provided evidence for a causal relationship between peer tutoring and final course performance, especially for first-generation college students (Colver and Fry, 2016). A meta-analysis of studies involving peer tutoring (broadly defined) in medical schools showed that students benefited equally from peer teachers and faculty (Bene and Bergus, 2014).

Various learning theories predict the success of peer tutoring. Chi (2009) drew a distinction between active, constructive, and interactive activities. Interactive activities like peer tutoring (in which a student is actually talking to another person about course material) have been shown to result in greater learning gains than activities that are simply constructive or active. From a social constructivist perspective, high-level interactions in which ideas, explanations, justifications, speculations, hypotheses, and inferences are exchanged can actually bring about changes in the cognitive structures of the tutor and tutee (Vygotsky, 1978; King *et al.*, 1998; Tudge and Rogoff, 1999). Specifically, Vygotsky believed that students are able to observe the cognitive skills of a more capable peer and in time internalize and develop them personally. Thus, a more advanced tutor would be necessary to scaffold tutees and help them advance cognitively within their “zone of proximal development” (Vygotsky, 1978). However, there is evidence that the tutor does not need to be more advanced than the tutee. In one study, students learned just as effectively when tutors were discouraged from giving explanations and feedback to the student (Chi *et al.*, 2001). Furthermore, other studies have shown that same-ability students can successfully tutor one another (King *et al.*, 1998; Menesses and Gresham, 2009; Jensen and Lawson, 2011). These observations are supported by equilibration theory, in which learning is stimulated when students encounter disequilibrating experiences in which prior knowledge can only accommodate new information if it is reorganized and contradictions are resolved (Piaget, 1985). In support of this, research shows that tutees must come to an impasse during tutoring sessions for learning to take place (VanLehn *et al.*, 2003), something that could occur with a tutor at the same ability level.

Peer tutoring has been shown to benefit tutors in addition to tutees, with the magnitude of this benefit varying from study to study (Roscoe and Chi, 2007; Bene and Bergus, 2014). In some instances, tutors may actually benefit more than the students they tutor (Fantuzzo *et al.*, 1989). The learning benefits of tutors likely stem from the opportunity that tutors have to participate in metacognitive reflection about their own understanding and expertise. As they help the tutee, they may be required to build upon their previous knowledge, generate inferences, and correct errors, or in other words, participate in reflective knowledge-building. Not all tutors will have this constructive learning experience. Peer tutors may simply lecture to students and summarize what they already know, termed “knowledge-telling” (Roscoe and Chi, 2007). Interestingly, tutee behavior can affect tutor learning. The more high-level questions tutees asked, the more self-monitoring tutors experienced, and

the more inferences they made. If tutees asked shallower questions, tutors tended to engage in shallower knowledge-telling (Roscoe and Chi, 2004). Roscoe (2014) later reported that even untrained student tutors were found to engage in self-monitoring and knowledge-building at times and that this was aided when tutees asked questions.

If both tutors and tutees benefit from tutoring, then reciprocal peer tutoring (RPT) should be an effective and efficient model to help students learn. In this model, students take turns filling the role of tutor and tutee, thus making a one-to-one student-to-instructor ratio possible in a class of any size. The reciprocal nature of RPT allows each student to experience both roles, but it also can result in more collaboration and cocreation of understanding as roles become blurred (Duran and Monereo, 2005). Many researchers have reported increased learning gains and strengthened metacognitive regulation in their students after implementing RPT in their classrooms, but the format of RPT used varies widely from study to study (Pigott *et al.*, 1986; Griffin and Griffin, 1998; Dioso-Henson, 2012; De Backer *et al.*, 2015, 2016; Manyama *et al.*, 2016; Yang *et al.*, 2016). There have been mixed results when RPT has been compared with fixed peer tutoring (in which roles are never reversed), with some studies showing no difference and others showing more benefits with RPT (Bentley and Hill, 2009; Cheng and Ku, 2009; Dioso-Henson, 2012). King reported that elementary-aged children benefited more from RPT when peer-guided questions were included, and this benefit was even higher when tutors were trained on the proper order in which to ask specific types of questions as they guided their peers (King, 1990; King *et al.*, 1998).

Previously, much of the research done on peer tutoring has been completed in grades K–12, medical schools, and undergraduate disciplines such as technology and psychology. Studies done in an undergraduate biology setting have shown learning gains as a result of peer-tutoring, which is encouraging. However, most of these used a tutoring model that involved former students and required extensive coordination and/or training of tutors to be part of official teaching teams (Groccia and Miller, 1996; Micari *et al.*, 2005; Stanger-Hall *et al.*, 2010; Hughes, 2011; Batz *et al.*, 2015).

Because there is evidence that same-ability students can effectively tutor one another, we hypothesized that we could implement an RPT model in an undergraduate biology course that would result in learning gains for students without extensive training, planning, or redesigning of courses by instructors. Furthermore, if only current students participated in the RPT, then such a model would not be reliant on monetary funds to hire former students. With these qualities, the assignment could be widely implemented in courses of many different levels, sizes, and formats.

Thus, we tested the effectiveness of a simple, recurring RPT homework assignment called the “teach and question” (TQ) assignment in a general education biology course. The TQ (described in greater detail later) was designed to be completed regularly by pairs of students enrolled in the course. In each session, the students would alternate filling the roles of *teacher* (explaining their understanding of unit material) and *questioner* (asking questions of the teacher to probe his or her understanding and help the teacher think more deeply about the content). The names of these roles were chosen purposefully to encourage

the students to be actively involved during both halves of the exercise. Several studies done on collaborative interactions showed that the partner who takes the role of speaker learns more than the partner who takes on the role of listener (Coleman *et al.*, 1997; Hausmann *et al.*, 2004; Schwartz and Bransford, 1998), perhaps because the speaker is the only one constructing knowledge and making inferences while the listener is being only passively attentive (Chi, 2009). We wanted to avoid this in the TQ and encourage both teacher and questioner to be making inferences beyond prior understanding and encountering impasses when they do not agree with each other or do not know how to answer a question. Furthermore, we hoped that the questions asked would encourage the teacher to engage in more knowledge-building rather than just knowledge-telling in his or her explanations (Roscoe, 2014).

A similar homework assignment in which students alternated the roles of asking and responding to questions about course material was used in an upper-level cell biology course with positive results. However, that study evaluated the implementation of an entire course structure, of which the homework assignment was just a small part. Because of this, the benefits of that specific element of the course remain unknown (Nelson *et al.*, 2009).

A quasi-experimental design was used here, such that the completion of the TQ assignment regularly was the only pedagogical difference between two sections of the same general biology course. We hypothesized that completing the TQ assignment regularly would increase student learning (as evidenced by student performance on course exams compared with the other section of the course), especially for lower-performing students. We also predicted that students who completed TQ sessions of higher quality would benefit more from this assignment than would students who were passive and unengaged. Thus, we wanted to determine which TQ session qualities would best predict final exam performance. Finally, we wanted to investigate student affect toward this assignment, because some students dislike group work and coordinating schedules with a peer.

## METHODS

### Ethics Statement

Written consent was obtained from all participants, and permission for use of human subjects was obtained from the Brigham Young University Institutional Review Board.

### Course and Participants

Participants in this study were college students enrolled in two sections of a general biology course for nonmajors offered at a large private university in 2015. Each section had 61–63 students enrolled, but only students who gave written consent were included in the study. This course met three times a week for 50-minute periods, was required as part of the general education core, and covered the entire biology curriculum (including molecular and cell biology, genetics, evolution, and ecology). Students ranged from first-years to seniors, and about two-thirds were male in each section. Students completed reading assignments before class at the beginning of each of the nine units, and active-learning exercises were used during class time. For each unit, students completed a writing assignment preceded by either a “teach and question” (TQ) or

“review” (R) session, depending on the section. Thus, the TQ section called this writing assignment the TQW (teach, question, write) and the R section called this writing assignment the RW (review and write). The format of the TQ and R portions are described in the *Experimental Design* section. The writing portion of these assignments required students to explain biological processes in their own words in response to a prompt describing a novel context. Students also completed one homework assignment of practice problems for each unit. Homework practice problems were similar to exam questions, and the majority required high-level cognitive skills (“apply,” “analyze,” and “evaluate” levels of Bloom’s taxonomy; Bloom, 1964; Anderson *et al.*, 2001). At the end of each unit, students were given an exam in class with immediate feedback after the test. All assessments were cumulative and included 12 multiple-choice questions, with the majority (~65%) requiring high-level cognitive skills. The cumulative final exam included a multiple-choice section (48 items, with 62.5% requiring high-level cognitive skills) and a written section with short-essay questions similar to their written assignments. A unique grading policy was used in which a certain number of midterm exams and other scores could be dropped if this would raise the student’s grade. However, the final exam became more heavily weighted as these scores were dropped (see Bailey *et al.*, 2017). Thus, a growth mindset (Dweck, 2010) was explicitly encouraged, as students could learn from their mistakes and improve throughout the semester as they prepared for the final exam.

### Experimental Design

This study used a quasi-experimental design in which steps were taken to ensure the two groups were as equivalent as possible. As such, identical course materials, textbooks, assignments (except for the TQ and R portions of the writing assignment, as will be discussed later), instructor, teaching assistants, exams, and expected learning outcomes were used. Because the sections met in different rooms and at different times, students completed the content-independent Lawson Classroom Test of Scientific Reasoning (LCTSR; 2000 version, with 24 items, including four items aimed at postformal reasoning; Lawson, 1978; Lawson *et al.*, 2000) at the beginning of the semester to assess equivalence of students’ entry-level scientific reasoning ability. The LCTSR was scored on a 24-point scale. These scores were also used to categorize students as concrete operational thinkers (scores < 9), transitioning from concrete to formal operations (low transition, scores = 9–14), transitioning from formal to postformal operations (high transition, scores = 15–20), and postformal operational thinkers (scores = 21–24; Lawson *et al.*, 2000).

Students in the TQ treatment section completed a TQ session with a peer before completing their writing assignments individually for every unit. The TQ portion of the TQW required students to meet up with a partner from the class on their own time for at least 30 minutes. Students took turns being the teacher and the questioner. The teacher referred to the learning objectives for the current unit and began teaching his or her partner about these objectives without using notes. The questioner was instructed to ask high-level questions (“How...?” “Why...?” “What would happen if...?” etc.) to probe the teacher’s understanding and help him or her think more deeply

about the material. Students were instructed to ask questions even if they did not know the answer themselves, and they were encouraged to think through difficult questions together and even use outside resources to find answers. Once the teacher had taught all of the learning objectives for that unit or had taught for at least 15 minutes, the students switched roles and repeated the exercise. The entire session was audio-recorded and emailed to the instructor for credit. The instructor and teaching assistants demonstrated quality TQ sessions in class at the beginning of the semester, and training videos portraying both good and bad examples of TQs were available for students to consult throughout the semester. Students were allowed to self-select their partners for each TQ session in order to make the assignment more logistically feasible. Most partnerships were quite consistent throughout the semester, but some students completed their TQs with a variety of peers throughout the semester.

The R treatment section completed a review session on their own (instead of a TQ) before completing the writing assignment for every unit. The R portion of the RW required the students to spend at least 30 minutes reviewing that unit's material individually, the same amount of time required for the TQ assignment in the other section. The instructor did not give these students specific instruction as to how they should study, just that studying should be done individually; however, these students were given the same list of learning objectives as the TQ section. Students self-reported the length of this review time when they turned in their written assignments. The actual time students spent on the R assignment did not differ from the actual time students spent on the TQ assignment (29.2 vs. 29.3 minutes on average,  $p = 0.96$  when compared by unpaired  $t$  test).

Just before the final exam, students completed an attitudinal survey to gather data on student interest, helpfulness of various in-class and at-home course activities, and opinions regarding the TQW and RW assignments. Other than an item regarding the number of hours studied per week, survey questions were Likert style, open response, or involved the ranking of activities.

Throughout this article, error bars on bar graphs represent the standard error of the mean.

### TQ Coding for Quality

Research assistants listened to all TQ audio recordings to code them for quality. Coders noted the total length of the audio file and the time at which the partners switched roles. Every question asked by a questioner was transcribed and categorized as a low-level memory question or a high-level question (including convergent-thinking questions, divergent-thinking questions, and evaluative-thinking questions; Ciardiello, 1998). Research assistants then categorized the interactions that followed each question into one of three categories: 1) the teacher either struggled answering the question or gave a vague answer and then gave up; 2) the teacher either struggled answering the question or gave a vague answer and then showed "grit" (Duckworth *et al.*, 2007; defined here as the teacher worked through it on his or her own, the questioner asked more leading questions to help the teacher come to the answer, or the teacher and questioner looked up the answer together); or 3) the teacher gave a complete, thoughtful

**TABLE 1. Interrater reliability for coding of TQ audio files**

Percent agreement <sup>a</sup>				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1		93.8	94.1	89.1
Coder 2			92.4	94.4
Coder 3				95.0
Coder 4				
Cohen's kappa <sup>b</sup>				
	Coder 1	Coder 2	Coder 3	Coder 4
Coder 1		0.845	0.745	0.740
Coder 2			0.798	0.873
Coder 3				0.871
Coder 4				

<sup>a</sup>Percent agreement was calculated as number of questions with matching categorization/number of total questions coded.

<sup>b</sup>Cohen's kappa was also calculated for each pair in order to take into account chance agreement.

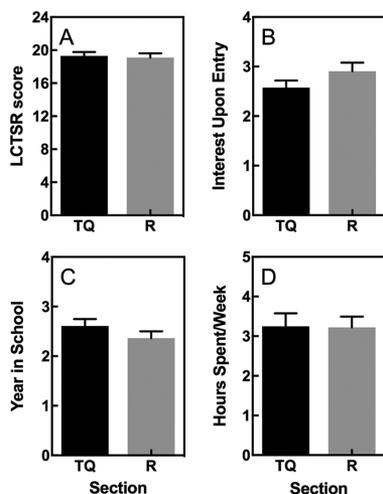
answer to the question. Because this categorization was subjectively determined, interrater reliability was calculated by having the four student researchers code the same five recordings. Each pair of coders was compared separately using both percent agreement and Cohen's kappa (Table 1). Because the agreement was good (Cohen 1960; McHugh, 2012), the rest of the recordings were coded only by a single student researcher due to the large number of hours of audio that needed to be coded. We also wanted to assess the balance between the two students, because there were some partnerships in which one student dominated the conversation and/or showed greater intellectual engagement. For this specific variable, we did not care whether the student was the one dominating or the one being dominated (because this should be captured by other recorded variables). Thus, as a quantitative estimate of the balance of each TQ session, we calculated the absolute value of the difference between the number of questions each student asked. The greater the discrepancy in the partnership, the larger the metric.

In summary, the following variables were recorded for each student for each recording: the LCTSR score of his or her partner; total length of TQ session in minutes; percent time spent teaching (as opposed to questioning); percent of learning outcomes taught; the number of questions asked during the session; the percent of questions asked that were high-level; if the teacher struggled answering a question, the percent of the time that grit was demonstrated (as defined earlier); and the metric to quantify how balanced the session was between partners (with a higher number suggesting more imbalance). For each student, these variables were then averaged for all TQ sessions completed by that student.

## RESULTS

### Equivalency of Sections

For all analyses that required comparison of the two course sections, students were included only if they took every midterm exam. As shown in Figure 1, the two sections of the course were also equivalent in terms of reasoning ability upon entry (A), biology interest upon entry (B), class level (C), and hours spent outside class each week (D). Reasoning ability was assessed using the LCTSR. Although the average scores of the treatment

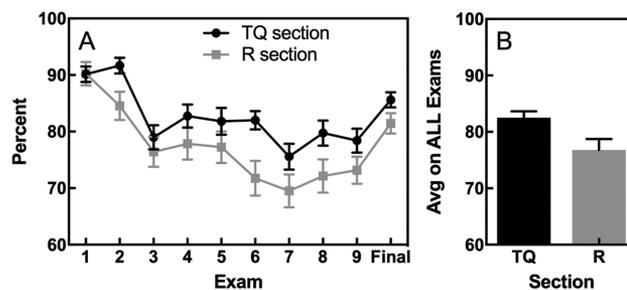


**FIGURE 1.** Equivalency of TQ section and R section. (A) LCTSR assessment was administered at the beginning of the semester and scored on a 24-point scale. Sections were statistically equivalent (unpaired  $t$  test,  $p = 0.78$ ,  $n = 44$ – $46$  students). (B) On an attitudinal survey at the end of the semester, students reported their interest in biology upon course entry using a 5-point Likert scale (1 = not at all interested; 5 = very interested). Sections were compared via Mann-Whitney  $U$ -test ( $p = 0.19$ ,  $n = 40$ – $43$  students). (C) Students' year in school was determined using the official course role from the registrar's office (1 = first-year, 4 = senior). A Mann-Whitney  $U$ -test was performed to compare sections ( $p = 0.22$ ,  $n = 44$ – $46$  students). (D) The number of hours spent on the course per week was self-reported on an attitudinal survey at the end of the course. Sections are indistinguishable (unpaired  $t$  test,  $p = 0.95$ ,  $n = 40$ – $43$  students).

(TQ) and control (R) sections were virtually identical on the LCTSR (19.3 vs. 19.1, respectively,  $p = 0.78$  by unpaired  $t$  test), LCTSR score was used as a covariate in later analyses in order to be conservative. Because the average biology interest of the R section was slightly higher than the TQ section (although  $p = 0.18$ ), students in the TQ section may have had more to gain throughout the semester. Thus, normalized gains in interest are reported later in addition to raw changes, again in an attempt to be conservative.

### Test Scores

The nine cumulative midterm exams and the final exam were identical in both format and content for the TQ and R sections of the course. Figure 2A shows average test scores for each section. The TQ section outperformed the R section an average of 5.7% on every test (95% confidence interval for mean: 3.5–7.9%). Test performance was compared between the TQ and the R sections using a mixed analysis of covariance (ANCOVA) with LCTSR score as a covariate to control for course preparation. As expected, test number and LCTSR were both significant. Class section (TQ vs. R) also significantly explained the variation in test scores ( $p = 0.019$ ) with a medium effect size ( $\eta_p^2 = 0.07$ ). Because the data from some of the midterms violated the assumption of normality, students' average test scores (average for all midterm and final exam scores, not adjusted for course preparation) were also compared using the Mann-Whitney  $U$ -test (see Figure 2B). The average score on all tests was



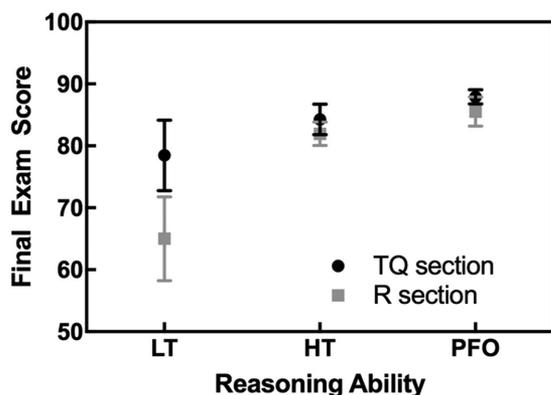
**FIGURE 2.** Exam performance of students in the TQ section of the course compared with students in the R section. Students were included in the analysis if they took every exam. (A) Average percent correct on each exam is shown on the y-axis. The TQ course section ( $n = 44$  students) was compared with the R section ( $n = 41$  students) using a mixed ANCOVA with LCTSR score as a covariate. Score on LCTSR ( $p < 0.0001$ ,  $\eta_p^2 = 0.18$ ), test number ( $p = 0.004$ ,  $\eta_p^2 = 0.03$ ), and course section ( $p = 0.019$ ,  $\eta_p^2 = 0.07$ ) all significantly explained variation in exam scores, but no significant interactions were observed. Graph shows raw exam scores (before adjustment for LCTSR score). (B) Average performance on all course exams was compared using a Mann-Whitney  $U$ -test. Students in the TQ section outperformed those in the R section (because LCTSR was not used as a covariate,  $n = 46$  and 44, respectively;  $p = 0.04$ ).

4.1% higher for students enrolled in the TQ section compared with the R section ( $p = 0.035$ ).

To investigate whether a certain group of students benefited from the TQ most, we separated students into three groups based on their LCTSR scores at the beginning of the course: low transition (LT), high transition (HT), and postformal operational (PFO) reasoners (see *Methods*). As shown in Figure 3, the greatest benefit of the TQ was seen in students with the lowest reasoning scores (LT reasoners). While no interaction between reasoning ability and TQ treatment was seen via two-way analysis of variance (ANOVA),  $t$  tests showed that LT thinkers in the TQ section significantly outperformed LT thinkers in the R section ( $p = 0.0498$ ), while scores of HT and PFO thinkers were indistinguishable between sections ( $p = 0.44$  and 0.43, respectively).

### Effect of TQ Quality

While comparisons of average test scores suggested that completing the TQ assignment as opposed to the R assignment increased student learning, we were interested in the effect of TQ quality on exam performance. To assess this, student research assistants coded all audio files of TQ sessions as described in the *Methods*. To determine whether TQ quality was an important predictor of final exam performance and, more specifically, what characteristics of a TQ best predicted success, we performed a hierarchical multiple linear regression analysis with final exam performance as our target. This method allowed for our independent variable inputs to be categorized into different blocks based on their theoretical importance (see Table 2). Block 1 included one variable to account for scientific reasoning, thus acting as a covariate to control for student ability upon course entry. Block 2 included variables that represent student participation in various course activities and assignments, because we expected overall participation to explain



**FIGURE 3.** Effect of course section on final exam performance when students were categorized by scientific reasoning ability. Students were divided into groups based on their LCTSR score (24-point scale): LT, low transition (transitioning from concrete operational stage to formal operational stage; LCTSR scores: 9–14), HT, high transition (transitioning from formal operational stage to postformal stage; LCTSR scores: 15–20), PFO, post-formal operational reasoning (using formal operational reasoning patterns on theoretical concepts; LCTSR scores: 21–24). Data were analyzed using a two-way ANOVA. Reasoning stage ( $p < 0.0001$ ,  $\eta_p^2 = 0.23$ ) and course section ( $p = 0.03$ ,  $\eta_p^2 = 0.06$ ) both significantly explained variation in final exam scores, but there was no significant interaction between them ( $p = 0.15$ ;  $n = 3-21$ ).

more variation in student performance than would small habits in only one type of assignment (the TQ). Finally, block 3 included variables that represent the quality of each student's TQ sessions. Within each block of independent variables, we used a stepwise method with bidirectional elimination. Independent variables were added to the model if their inclusion significantly improved the fit (probability of  $F < 0.050$ ), and a variable could be eliminated from a later model if it no longer improved the fit (probability of  $F > 0.10$ ).

The results of the hierarchical multiple linear regression are shown in Table 3. Inclusion of LCTSR scores (from block 1) in model 1 resulted in an adjusted  $R^2$  of 0.111. Of all the course activities and assignments in block 2, students' attendance was the only one that significantly improved the regression model. Its inclusion in model 2 significantly increased the adjusted  $R^2$  to 0.331. Finally, as block 3 independent variables were allowed to be included, the average number of questions the students asked during their TQ sessions significantly improved the fit ( $p = 0.006$ ). The final model (model 3) had an adjusted  $R^2$  of 0.410 and included LCTSR, attendance, and number of questions asked per TQ session as significant predictors of final exam performance. Model coefficients for each variable are shown in Table 3.

The only variable described in the *TQ Coding for Quality* section that is not included in this analysis is the partner's LCTSR score. If this variable is included in block 3 (see Table 2), it does not add significantly enough to the model to be included. However, it does reduce the sample size, because a few students did not take the LCTSR at the beginning of the semester. The hierarchical multiple linear regression results with this smaller sample size are very similar to those shown in Table 3 (LCTSR, attendance, and number of questions asked are all still significant predictors), except average score on written assignments (block 2) is added to the model. Because including partner LCTSR is essentially reducing sample size without increasing predictive power, we chose not to include it in the analysis.

Hierarchical multiple linear regression was chosen because of its theoretical validity, but we also analyzed these data using a best subset model to increase our confidence that LCTSR, attendance, and number of questions asked during a TQ session were the best predictors of final exam score. When all the variables shown in Table 2 were available as inputs on equal footing (rather than in blocks) and the best subset was chosen using the Akaike information criterion (corrected for finite sample size), the model that best predicted final exam performance ( $R^2 = 0.432$ ) included LCTSR ( $p = 0.003$ ), attendance ( $p = 0.029$ ), number of questions asked during TQ ( $p = 0.007$ ), and scores

**TABLE 2.** Possible independent variable inputs considered for the hierarchical multiple linear regression to predict final exam score

Block <sup>a</sup>	Theoretical rationale	Variable	Description
1	Measure representing preparation for course (covariate)	LCTSR	Score on LCTSR, a content-independent measure of scientific reasoning ability (entered as number correct with a maximum of 24)
2	Participation in course activities and assignments (overall course habits are expected to have a large impact on performance)	READING	Percent of reading assignments completed
		ATTENDANCE	Percent of classes attended
		HW	Percent of homework practice problems completed
		TQ	Percent of TQ assignments completed
3	Average quality of TQ session (specific habits on a specific type of course assignment are predicted to have smaller effects on course performance)	WRITTEN	Average score on written homework assignments
		LENGTH	Average total length of TQs in minutes
		% TIME TEACH	Average percent of time student spent in role of "teacher"
		GRIT	Average percent of time student demonstrated grit when he or she struggled
		# QUES ASKED	Average number of questions this student asked during TQ
		% QUES HIGH	Average percent of questions asked that were high level
% LO TAUGHT	Average percent of learning outcomes this student covered during TQ		
		IMBALANCE	Larger number suggests greater imbalance (see <i>Methods</i> for calculation)

<sup>a</sup>Variables are organized in blocks according to the theoretical rationale described in the second column. Within each block, variables were considered for model inclusion using a stepwise method with bidirectional elimination.

TABLE 3. Results of hierarchical multiple linear regression with final exam score as target

Model <sup>a</sup>	R <sup>2</sup>	Adjusted R <sup>2</sup>	Significance (change in R <sup>2</sup> )	Variable	B (coefficient)	SE <sub>B</sub>	β (standardized coefficient)	p value
1	0.127	0.111	0.007	(Intercept)	66.846	6.937		<0.001
				LCTSR	1.003	0.355	0.356	0.007
2	0.355	0.331	<0.001	(Intercept)	47.244	7.511		<0.001
				LCTSR	1.072	0.308	0.381	0.001
				ATTENDANCE	0.207	0.048	0.478	<0.001
3	0.442	0.410	0.006	(Intercept)	47.683	7.052		<0.001
				LCTSR	0.988	0.291	0.351	0.001
				ATTENDANCE	0.178	0.046	0.409	<0.001
				# QUES ASKED IN TQ	0.704	0.244	0.305	0.006

<sup>a</sup>Models 1, 2, and 3 show the result of the regression after blocks 1, 2, and 3 were considered, respectively.  $n = 57$ .

on written assignments ( $p = 0.088$ ) as predictors (more detailed results not shown). The similarities between this model and model 3 from Table 3 increase our confidence that these are the best predictors of final exam score. Because the blocks of variables in the hierarchical model allow for ordered variable selection based on theory, we settled on Table 3 as the final result of our regression analysis.

#### Attitudinal Data

Attitudinal data were gathered at the end of the semester to investigate student perceptions of the TQ assignment as well as other aspects of the course. As shown in Figure 4A, both sections reported an increase in biology interest during the semester. This increase was greater for the section that completed TQ assignments compared with the section that completed R assignments (Figure 4B). Because the TQ section started out slightly less interested than the R section (see Figure 1B), they may have had more to gain during the semester. To account for this, we also compared normalized interest gains, and course sections were no longer distinguishable statistically (Figure 4C;  $p = 0.06$ ).

Students were asked to rank six at-home activities according to their helpfulness, with “6” being the most helpful. The TQW and RW assignments were ranked the highest of any at-home activity for both sections, but the TQW assignment was ranked

slightly higher than the RW assignment (Figure 5A). The data in Figure 5, B and C, suggest that the TQW and RW assignments may have been highly ranked for different reasons. Most of the students in the TQ section (87.5%) reported that the TQ portion of the assignment was more helpful than the written portion, while only 51.7% of the RW section preferred the R portion over the written portion. The only other statistically significant difference between the two sections’ ranking of at-home activities was how they ranked studying alone (not as an assignment); the R section ranked it  $3.55 \pm 1.58$  compared with  $2.57 \pm 1.20$  for the TQ section (unpublished data; Mann-Whitney  $U$ -test,  $p = 0.009$ ).

Students were asked whether receiving the opposite treatment would have been more or less helpful to their learning (i.e., the TQ section was asked whether reviewing alone would have been more helpful and the R section was asked whether having a partner would have been more helpful). The data are shown in Figure 6, and they suggest that students in all sections view reviewing with a partner as more helpful for their learning.

## DISCUSSION

### Benefits of TQ Assignment

The TQ assignment is a relatively simple one to implement but can have high rewards. It requires no class time to complete and only 30 minutes of student time outside class for each unit.

Instructors do need to provide lists of unit learning objectives to guide student TQ sessions, but creating such lists is also part of good course design and teaching practice. Even when assigned only 10 times during a semester, this simple TQ assignment led to large gains in student performance on assessments (see Figure 2). An increase of ~6% on exams is equivalent to two half-letter grades (e.g., raising a “B-” to a “B+”). This benefit cannot be attributed to increased time required with the material, because a substitute assignment requiring the same time commitment did not bring this benefit. So, what does the TQ assignment require that is not needed when the student studies independently?

The TQ assignment explicitly requires the asking of questions. As summarized in Table 3, the average number of questions a

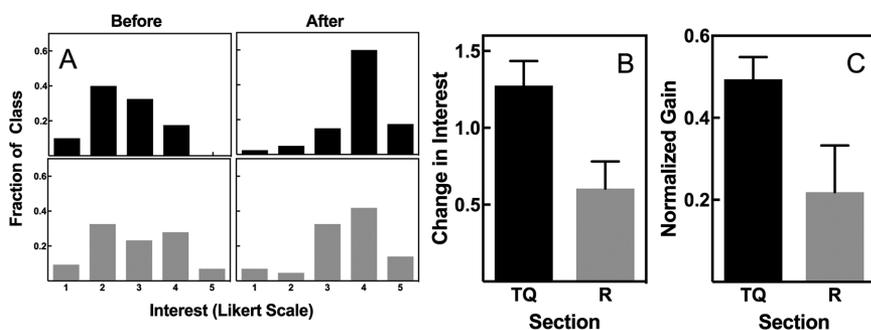


FIGURE 4. Effect of TQ assignment on biology interest. Students self-reported their interest in biology at the beginning and end of the course on a five-point Likert scale (1 = “not at all interested” and 5 = “very interested”;  $n = 40-43$ ). (A) Histograms of interest levels before and after the course for the TQ section (black) and the R section (gray). (B) The average change in interest (interest after – interest before). The TQ section increased more than the R section (Mann-Whitney  $U$ -test,  $p = 0.017$ ). (C) The normalized gain in interest (interest after – interest before/maximum increase possible). Differences between sections were not statistically distinguishable (Mann-Whitney  $U$ -test,  $p = 0.06$ ).

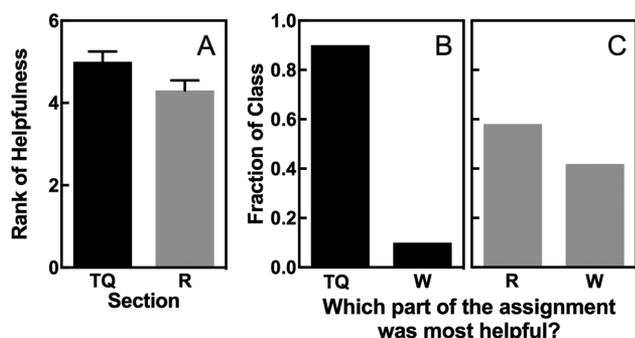


FIGURE 5. Student affect toward TQW vs. RW assignments. (A) On an attitudinal survey at the end of the course, students were asked to rank six at-home activities (TQW or RW, textbook readings, practice homework problems, studying with a peer—not as part of an assignment, studying alone—not as part of an assignment, and reviewing past exams) with 6 being the most helpful and 1 being the least helpful. Average rank of the TQW and RW are shown here. Section distributions were distinguishable by Mann-Whitney  $U$ -test ( $p = 0.015$ ,  $n = 33$ – $35$ ). (B, C) Students were asked to choose which portion of the TQW or RW was most helpful: the TQ/R portion (depending on the section) or the written portion (W). Data are represented as fraction of the class choosing that portion. Distributions for the TQ section (black, B) and R section (gray, C) were compared using a Mann-Whitney  $U$ -test ( $p = 0.0012$ ,  $n = 40$ – $43$ ).

student asked during TQ sessions was the most important aspect of TQ quality when predicting final exam performance. This variable's inclusion in the regression model was able to explain ~8% more of the variation in final exam scores above and beyond the variation already explained by reasoning ability at course entry and class attendance. Interestingly, other general aspects of course participation (see block 2 of Table 2) could not significantly explain variation above and beyond regression model 2 (Table 3). This may be partially due to the lack of sufficient variation in some of these variables, but it still suggests that the number of questions asked during the TQ session may be more important than other general habits of students in biology courses. Interestingly, the percent of questions asked that were high level did not significantly predict performance on the final exam. One possibility is that low-level questions were unexpectedly sufficient to encourage the teacher to engage in knowledge-building (Roscoe and Chi, 2007; Roscoe, 2014) and that even these simple memory questions kept the questioner engaged. On the other hand, perhaps high-level questions are important for constructive learning and we just did not have enough data to see the effects. On average, only ~30% of student questions were high level, so perhaps more training on how to pose high-level questions would have increased their use and effectiveness. Because the number of questions asked was the aspect of TQ quality that was most predictive of success on exams, future research should investigate whether a modified R assignment in which students were required to generate and write down questions related to the course material could increase learning gains in the same manner as the TQ assignment.

Another big difference between the TQ assignment and the R assignment is the verbal component. Students in the TQ section of the course were actually required to verbally teach

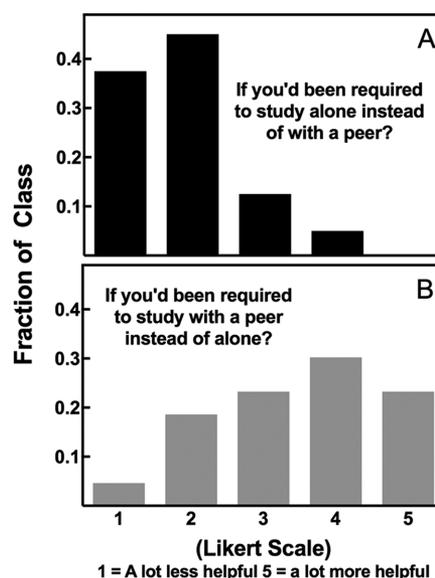


FIGURE 6. Student report of whether or not the opposite treatment would have been more helpful. On an attitudinal survey at the end of the semester, students in the TQ section (black, A) were asked whether being required to study alone would have been more or less helpful than studying with a peer as part of the TQW. Students in the R section (gray, B) were asked whether being required to study with a partner would have been more or less helpful than studying alone as part of the RW. Responses were reported on a five-point Likert scale. Section distributions were compared by Mann-Whitney  $U$ -test ( $p < 0.0001$ ,  $n = 40$ – $43$ ).

course material rather than just reading, thinking, and writing about it. Our results cannot necessarily help us draw any conclusions about this specific aspect of the TQ, because neither the total length of the TQ session nor the percent of that time spent in the role of teacher (our two variables most likely to reflect variation in amount of verbalization) significantly predicted exam performance. This may be because verbalization is not important, or more likely, we may not have had enough variation in these variables to add significantly to the regression model. Because the assignment required a minimum of 30 minutes and students were trained to switch roles halfway through, the majority of the TQ sessions had these characteristics. A future study should be done to investigate whether participating in verbal self-directed explanation of course material could provide the same benefits as the TQ assignment.

The TQ assignment also has a social component that was lacking in the R assignment. Beyond explaining material verbally and generating questions, both of these activities were directed at another person and were thus interactive (Chi, 2009). Occasionally, when students became very engaged with specific questions that were asked, the lines between teacher and questioner became blurred, and the format of the TQ was more of a general discussion. These instances in the recordings were the hardest to code, and we unfortunately did not come up with a good way to quantify this and factor it into our analysis. In these instances, coconstruction of knowledge was occurring (Hausmann *et al.*, 2004), and an R assignment completed as an individual could never replicate this. Because average

partner's LCTSR did not significantly add to our regression models (see *Effect of TQ Quality* section), we cannot draw any conclusions about the effect of partnership type (same-ability peers or not). Because our students self-selected their partners and were not always in consistent partnerships throughout the semester, our experimental design was not optimal to answer this question. Future research should more fully investigate whether same-ability partnerships or partnerships with a higher-performing student and a lower-performing student are more advantageous with this particular RPT model.

One variable we did not have enough information to account for is how much time students spent reviewing material before the TQ session began in an attempt to prepare, even though this was not required or even suggested. We know from student comments to the instructor that some were nervous about teaching a classmate and thus spent time preparing on their own first. While we know that students in the TQ session did not spend more time studying for the class than did the students in the R section (see Figure 1), students in the TQ section potentially spent some of their study time preparing to teach, which is known to increase learning gains, at least in the short term (Fiorella and Mayer, 2013, 2014). Because we have no way of knowing how many students in the TQ section did this, we do not know to what extent this is responsible for the differences between sections.

Another lurking variable is the possibility that students in the TQ section realized how helpful studying with a peer can be because of the TQ assignment, and thus they studied more often in groups than did the students in the R section. Unfortunately, we did not ask the students how often they studied with others. The only item on the attitudinal survey related to studying with others (outside the TQ) was the question that asked them to rank the helpfulness of the six at-home activities. There was no statistical difference in how the TQ section and the R section ranked studying with others (not as part of the TQ), but we cannot ignore the possibility that the extent to which they used that resource could have been different. Interestingly, the students in the R section rated studying alone (not as part of any assignment) as more helpful than did the students in the TQ section (see *Attitudinal Data* section), suggesting that, if students found the R assignment helpful, they may have done more studying on their own.

### Implementing the TQ Assignment

For those interested in implementing the TQ assignment, our course format included the following characteristics that may have contributed to its success. First, students were trained on how to do the assignment properly at the beginning of the semester. The course studied here dedicated about 15 minutes of class time to a TQ demonstration and class discussion about what constitutes a quality TQ session (asking good “Why...?” and “How...?” and “What if...?” questions; drawing as you explain; trying to teach with examples; showing grit when you don't know the answer to the question; guiding your partner to correct his or her explanation if he or she says something incorrect, etc.), and two example videos were available for the students to view online throughout the entire semester (one good example and one bad example with commentary). To help alleviate stress due to the logistics of aligning schedules with a peer, this course also included an optional homework session in which students could come and find a TQ partner. Students were required to

audio-record their TQ sessions so that they could be coded for this study. Anecdotally, this recording seemed to improve the quality of the TQ sessions compared with other semesters when they were not recorded, so we suggest you require audio recordings for credit. The course professor and teaching assistants did not listen to all of the recordings in their entirety during the semester. However, the instructors noted the length of the TQ sessions and estimated the time at which partners switched roles in order to give proper course credit. In addition, for some of the recordings for each assignment, the instructors listened to short snippets chosen at random, one during the first half of the session and one during the second half the session. If instructors had any comments, this feedback was included online with the assignment grade. We found that offering this scattered feedback sent the message that we listened to the recordings, increasing student accountability for the quality of their work. Finally, while this specific course included a written portion of the assignment after the TQ, its inclusion cannot explain the increase in exam scores (see Figure 2), because both sections completed the writing portion. Thus, we believe the TQ could be implemented by itself or in conjunction with a writing prompt as done here.

### Conclusion

Incorporating RPT into a general biology course as a homework assignment increased student exam performance by an average of about two half-letter grades with minimal instructor effort. Asking more questions during these tutoring sessions was a significant predictor of final exam performance, explaining variability in performance above and beyond that explained by student ability at course entry and participation in other elements of the course. In general, students viewed the RPT assignment favorably. This homework assignment could be implemented in any size undergraduate biology class as a great way to bring collaborative learning into the course without extensive planning or major changes in course design.

### REFERENCES

- Anderson, L. W., Krathwohl, D. R., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., ... Wittrock, M. (2001). A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy. *Cognition and Instruction*, 9(2), 137–175.
- Bailey, E., Jensen, J., Nelson, J., Wiberg, H., & Bell, J. (2017). Weekly formative exams and creative grading enhance student learning in an introductory biology course. *CBE—Life Sciences Education*, 16(1), ar2.
- Batz, Z., Olsen, B. J., Dumont, J., Dastoor, F., & Smith, M. K. (2015). Helping struggling students in introductory biology: A peer-tutoring approach that improves performance, perception, and retention. *CBE—Life Sciences Education*, 14(2), ar16.
- Bene, K. L., & Bergus, G. (2014). When learners become teachers: A review of peer teaching in medical student education. *Family Medicine*, 46(10), 783–787.
- Bentley, B. S., & Hill, R. V. (2009). Objective and subjective assessment of reciprocal peer teaching in medical gross anatomy laboratory. *Anatomical Sciences Education*, 2(4), 143–149.
- Bloom, B. S. (1964). *Taxonomy of educational objectives*. New York: Longmans, Green.
- Cheng, Y. C., & Ku, H. Y. (2009). An investigation of the effects of reciprocal peer tutoring. *Computers in Human Behavior*, 25(1), 40–49.
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105.
- Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4), 471–533.

- Cardiello, A. (1998). Did you ask a good question today? Alternative cognitive and metacognitive strategies. *Journal of Adolescent & Adult Literacy*, 42(3), 210–219.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Cohen, P. A., Kulik, J. A., & Kulik, C.-L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19(2), 237–248.
- Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *Journal of the Learning Sciences*, 6(4), 347–365.
- Colver, M., & Fry, T. (2016). Evidence to support peer tutoring programs at the undergraduate level. *Journal of College Reading and Learning*, 46(1), 16–41.
- De Backer, L., Van Keer, H., & Valcke, M. (2015). Promoting university students' metacognitive regulation through peer learning: The potential of reciprocal peer tutoring. *Higher Education*, 70(3), 469–486.
- De Backer, L., Van Keer, H., & Valcke, M. (2016). Eliciting reciprocal peer-tutoring groups' metacognitive regulation through structuring and problematizing scaffolds. *Journal of Experimental Education*, 84(4), 804–828.
- Dioso-Henson, L. (2012). The effect of reciprocal peer tutoring and non-reciprocal peer tutoring on the performance of students in college physics. *Research in Education*, 87(1), 34–49.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087.
- Duran, D., & Monereo, C. (2005). Styles and sequences of cooperative interaction in fixed and reciprocal peer tutoring. *Learning and Instruction*, 15(3), 179–199.
- Dweck, C. S. (2010). Even geniuses work hard. *Educational Leadership*, 68(1), 16–20.
- Fantuzzo, J. W., Dimeff, L. A., & Fox, S. L. (1989). Reciprocal peer tutoring: A multimodal assessment of effectiveness with college students. *Teaching of Psychology*, 16(3), 133–135.
- Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, 38(4), 281–288.
- Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, 39(2), 75–85.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, 111(23), 8410–8415.
- Griffin, M. M., & Griffin, B. W. (1998). An investigation of the effects of reciprocal peer tutoring on achievement, self-efficacy, and test anxiety. *Contemporary Educational Psychology*, 23(3), 298–311.
- Groccia, J. E., & Miller, J. E. (1996). Collegiality in the classroom: The use of peer learning assistants in cooperative learning in introductory biology. *Innovative Higher Education*, 21(2), 87–100.
- Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, 332(6034), 1213–1216.
- Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., ... Tilghman, S. M. (2004). Scientific teaching. *Science*, 304(5670), 521–522.
- Hausmann, R. G. M., Chi, M. T. H., & Roy, M. (2004). Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. *Cognitive Science—COGSCI* (pp. 547–552). London: Psychology Press.
- Hughes, K. S. (2011). Peer-assisted learning strategies in human anatomy & physiology. *American Biology Teacher*, 73(3), 144–147.
- Jensen, J. L., & Lawson, A. (2011). Effects of collaborative group composition and inquiry instruction on reasoning gains and achievement in undergraduate biology. *CBE—Life Sciences Education*, 10(1), 64–73.
- King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27(4), 664–687.
- King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, 90(1), 134.
- Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11–24.
- Lawson, A. E., Alkhoury, S., Benford, R., Clark, B. R., & Falconer, K. A. (2000). What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. *Journal of Research in Science Teaching*, 37(9), 996–1018.
- Manyama, M., Stafford, R., Mazyala, E., Lukanima, A., Magele, N., Kidenya, B. R., ... Kauki, J. (2016). Improving gross anatomy learning using reciprocal peer teaching. *BMC Medical Education*, 16(1), 95.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Menesses, K. F., & Gresham, F. M. (2009). Relative efficacy of reciprocal and nonreciprocal peer tutoring for students at-risk for academic failure. *School Psychology Quarterly*, 24(4), 266.
- Micari, M., Streitwieser, B., & Light, G. (2005). Undergraduates leading undergraduates: Peer facilitation in a science workshop program. *Innovative Higher Education*, 30(4), 269–288.
- Nelson, J., Robison, D. F., Bell, J. D., & Bradshaw, W. S. (2009). Cloning the professor, an alternative to ineffective teaching in a large course. *CBE—Life Sciences Education*, 8(3), 252–263.
- Pai, H.-H., Sears, D. A., & Maeda, Y. (2015). Effects of small-group learning on transfer: A meta-analysis. *Educational Psychology Review*, 27(1), 79–102.
- Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.
- Pigott, H. E., Fantuzzo, J. W., & Clement, P. W. (1986). The effects of reciprocal peer tutoring and group contingencies on the academic performance of elementary school children. *Journal of Applied Behavior Analysis*, 19(1), 93–98.
- Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. *Instructional Science*, 42(3), 327–351.
- Roscoe, R. D., & Chi, M. T. (2004). The influence of the tutee in learning by peer tutoring. *Proceedings of the Cognitive Science Society*, 26(26), 1179–1184.
- Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4), 534–574.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475–5223.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, 323(5910), 122.
- Stanger-Hall, K. F., Lang, S., & Maas, M. (2010). Facilitating learning in large lecture classes: Testing the "teaching team" approach to peer learning. *CBE—Life Sciences Education*, 9(4), 489–503.
- Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education*, 32(3), 321–345.
- Tudge, J., & Rogoff, B. (1999). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. *Lev Vygotsky: Critical Assessments*, 3, 32–56.
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209–249.
- Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-based implementation of peer instruction: A literature review. *CBE—Life Sciences Education*, 14(1), es3.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wood, W. B. (2009). Innovations in teaching undergraduate biology and why we need them. *Annual Review of Cell and Developmental Biology*, 25, 93–112.
- Yang, E. F. Y., Chang, B., Cheng, H. N. H., & Chan, T. W. (2016). Improving pupils' mathematical communication abilities through computer-supported reciprocal peer tutoring. *Educational Technology & Society*, 19(3), 157–169.