

Article

Helping Struggling Students in Introductory Biology: A Peer-Tutoring Approach That Improves Performance, Perception, and Retention

Zachary Batz,^{*†‡} Brian J. Olsen,^{*} Jonathan Dumont,^{*†} Farahad Dastoor,^{*} and Michelle K. Smith^{*†}

^{*}School of Biology and Ecology and [†]Maine Center for Research in STEM Education, University of Maine at Orono, Orono, ME 04469-5751

Submitted August 7, 2014; Revised January 7, 2015; Accepted January 8, 2015
Monitoring Editor: Janet Batzli

The high attrition rate among science, technology, engineering, and mathematics (STEM) majors has long been an area of concern for institutions and educational researchers. The transition from introductory to advanced courses has been identified as a particularly “leaky” point along the STEM pipeline, and students who struggle early in an introductory STEM course are predominantly at risk. Peer-tutoring programs offered to all students in a course have been widely found to help STEM students during this critical transition, but hiring a sufficient number of tutors may not be an option for some institutions. As an alternative, this study examines the viability of an optional peer-tutoring program offered to students who are struggling in a large-enrollment, introductory biology course. Struggling students who regularly attended peer tutoring increased exam performance, expert-like perceptions of biology, and course persistence relative to their struggling peers who were not attending the peer-tutoring sessions. The results of this study provide information to instructors who want to design targeted academic assistance for students who are struggling in introductory courses.

INTRODUCTION

The poor retention of undergraduates in science, technology, engineering, and mathematics (STEM) fields has long been an area of concern for U.S. educators and has

spurred broad calls to reform undergraduate STEM courses (American Association for the Advancement of Science, 2011; President’s Council of Advisors on Science and Technology, 2012). Nationwide, nearly half (48%) of the students who enter a bachelor’s program seeking a STEM degree switch into a non-STEM field or leave college altogether (Chen and Soldner, 2013). The majority of students who leave STEM fields decide to do so within the first 2 yr of their program (Watkins and Mazur, 2013). Low grades in introductory STEM courses are often a contributing factor, pushing students out of STEM majors, while higher grades in non-STEM courses pull them toward non-STEM majors (Ost, 2010).

To help retain students in STEM majors, institutions have invested resources trying to improve the performance and retention of students in introductory STEM courses. One commonly used strategy is peer tutoring, in which undergraduates help one another learn within clearly defined tutor and tutee roles (Topping, 2000). A number of national programs have been designed to formalize peer tutoring, including the peer-led team learning (PLTL) and learning assistant (LA)

CBE Life Sci Educ June 1, 2015 14:ar16

DOI:10.1187/cbe.14-08-0120

[†]Present address: Department of Biology, Georgetown University, Washington, DC 20057-1229.

Address correspondence to: Michelle K. Smith (michelle.k.smith@maine.edu).

© 2015 Z. Batz *et al.* CBE—Life Sciences Education © 2015 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.

programs (Gosser *et al.*, 2001; Otero *et al.*, 2010). Both programs train undergraduates who previously succeeded in a course to assist all students currently enrolled in that same course. Peer tutors reduce the student-to-instructor ratio in a course, thereby facilitating additional group work and active-learning opportunities for the students. The introduction of peer tutors into biology, chemistry, engineering, physics, and computer science courses has resulted in increased student grades (Preszler, 2009; Gosser, 2011; Hooker, 2011) and content mastery (Pollock, 2009; University of Colorado Learning Assistant Program, 2011; Chasteen *et al.*, 2012).

While there are many opportunities for peer tutors to assist students in class, they can also help students who are participating in study groups outside class. Typically, the opportunity to participate in these study groups is offered to all students, and the effectiveness of these programs is monitored by comparing course performance metrics for students who participate in the study groups versus those who do not (e.g., Born *et al.*, 2002; Hockings *et al.*, 2008; Stanger-Hall *et al.*, 2010). Because students who join these study groups can have a wide range of abilities, these studies often use demographic information such as incoming grade point average (GPA) or Scholastic Aptitude Test (SAT) score to control for differences in the two populations. After controlling for these differences, several studies have shown that peer-tutoring study groups improve course grades and positively impact persistence (e.g., Born *et al.*, 2002; Hockings *et al.*, 2008; Stanger-Hall *et al.*, 2010). For example, a study offering a PLTL program outside of an introductory chemistry course found that, after controlling for students' background and other characteristics, students who attended earned a third of a letter grade higher overall in the course than students who declined to attend (Hockings *et al.*, 2008).

The success of peer-tutoring programs in STEM courses suggests that wider implementation could improve student performance and persistence of STEM undergraduates. However, not all institutions have the time and money to provide extra study group programs for every student in a class, especially in high-enrollment introductory STEM courses. Therefore, recent efforts have focused on providing targeted help to students who are struggling early in STEM courses. For example, in a study by Deslauriers *et al.* (2012), students who performed poorly on the first exam in an introductory physics or introductory oceanography course received an email invitation to meet with the course professor to discuss their current study strategies and reflect on how they might improve them. The struggling students who met with the professor improved their next exam grade nearly 5% more than their struggling peers who had not.

Similarly, a study in an introductory psychology course offered students who failed the first exam an intervention focused on developing metacognitive skills (Lizzio and Wilson, 2013). These struggling students were contacted by a tutor and invited to complete a workbook that was designed to have them reflect on their performance and identify deficiencies in their preparation. Following this self-reflection, students met with the trained tutor for approximately an hour to discuss the contents of their workbook and specific steps they could take to improve. Approximately two-thirds (64%) of students who took part in this intervention passed the course compared with just 27% of the struggling students who declined the intervention opportunity.

In this paper, we build on work demonstrating the benefits of peer tutoring and interventions that target struggling students and ask: Can weekly peer tutoring still be effective in an introductory biology course when it is offered to the lowest-performing students? Specifically, we offered a peer-tutoring program to students who scored 50% or lower on the first introductory biology exam and assessed the impact of this program on exam performance; student perception of biology as a discipline, using the Colorado Learning Attitudes about Science Survey in Bio (CLASS-Bio; Semsar *et al.*, 2011); and persistence. Taken together, this work helps instructors understand how peer tutoring impacts students who are having the most difficult time in a course and how these outcomes can be used to repair STEM's leaky pipeline.

METHODS

Course Description

This study was conducted in the Introductory Biology (BIO 100) course at the University of Maine (UMaine) during the Fall of 2013. Seven hundred sixty full-time students were enrolled in the course across three sections. Students in all three sections were cotaught by two faculty members (B.J.O. and F.D.), covered the same materials, and took identical exams. Each section met for 50 min, 3 d/wk.

The course is graded out of 1100 points—300 from exams (exams 1–4), 200 from a comprehensive final exam (exam 5), and 100 from weekly online homework assignments and in-class clicker questions. The remaining 500 points come from a required lab component of the course. For this study, exam scores are used as the measure of performance, because they are the primary method of assessment in this course.

Course Demography

The BIO 100 course is the largest course taught at UMaine, and students are primarily first-year, nonminority students from rural counties (Table 1). Most students were enrolled in majors that require additional biology courses. Women outnumbered men almost two to one.

Identification of Struggling Students

In upper-level courses, college GPA is a good predictor of whether a student will struggle (Murtaugh *et al.*, 1999), but this measure was not available, because students often take the BIO 100 course in their first college semester (Table 1). Therefore, in this study, we identified struggling students based on their first exam scores (Jensen and Moore, 2008). The average score on the first exam was 55.4% (SD = 16.7%). For the purpose of this study, students who earned 50% or lower on the first exam were identified as “struggling,” while students who scored above 50% were identified as “non-struggling” (Figure 1).

Shortly after the first exam, struggling students received both an email and a paper invitation to enroll in peer tutoring (Supplemental Figure 1). These invitations expressed concern about the student's performance on the first exam and contained information about when the sessions would meet, the materials and structure of the sessions, and directions on how to sign up. Approximately one-third of struggling students signed up for the program (Figure 1).

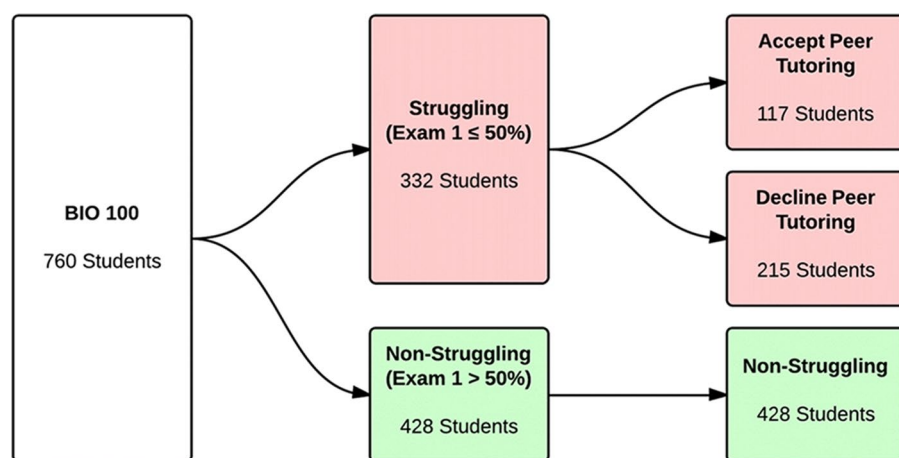


Figure 1. Flowchart of group membership in the study.

A small number ($n = 6$) of non-struggling students independently contacted the course professors and asked to join the peer-tutoring program. These students were permitted to join a peer-tutoring session but are excluded from all data analyses.

The undergraduate peer tutors were hired as part of an LA program (Otero *et al.*, 2010). In accordance with the LA program framework, students who previously scored in the top quartile in BIO 100 and expressed interest in a STEM teaching career were hired. The peer tutors were concurrently enrolled in a course covering practical and theoretical aspects of STEM education to help prepare them both for their roles as peer tutors and their potential futures as STEM educators.

Table 1. Demographic overview of full-time BIO 100 students during Fall 2013

	n^a	% of class
Gender		
Women	499	65.7
Men	261	34.3
Home county		
Urban (population > 250,000)	132	18.3
Rural (population < 250,000)	591	81.7
Ethnicity		
Nonminority	636	91.8
Minority ^b	57	8.2
Additional biology course requirements		
Plan requires additional biology	576	75.8
Plan does not require additional biology	130	17.1
Undeclared	54	7.1
Year in school		
First year	576	75.8
Sophomore	134	17.6
Junior	40	5.3
Senior	9	1.2
Standardized test	Mean	SD
SAT score ^c	1070.3	131.0

^aTotal Bio 100 enrollment $n = 760$ full-time students; class size varies slightly throughout table due to gaps in available demographic information.

^bMinority students include: Hispanic/Latino, American Indian/Alaska Native, black/African American, and Asian.

^cAnd/or SAT score equivalent of ACT score (see *Methods*).

Structure of Peer Tutoring

In this study, peer tutoring included 1) weekly meetings with up to 10 struggling students and the peer tutor and 2) question packets prepared by the researchers and faculty instructors that were given to the students each week. The question packets included five multiple-choice questions selected from exams given in previous years. The selected questions aligned with course learning goals and, when possible, were at the application and analysis levels (Bloom *et al.*, 1956; Supplemental Figure 2). Although the faculty members who teach this course do not post old exams, graded exams are returned each year to students. Consequently, students often have copies of exams from previous years, and the questions have been posted to various websites.

Students worked together in small groups to answer each question and write a justification of why the correct answer was correct along with additional justifications for why each incorrect answer was incorrect. While students worked, peer tutors circulated to monitor their progress and answer questions. Toward the end of the session, the peer tutor would bring the group back together to review their answers and justifications.

To investigate how well the peer tutoring prepared students for exams, three researchers (F.D., Z.B., and J.D.) independently compared the content of each exam question with the content of the questions asked in the peer-tutoring sessions. Questions that did not have 100% agreement were discussed until there was consensus. In total, 55.6% of the exam questions were similar in content to questions covered in the peer-tutoring sections.

To ensure that peer tutors were prepared for these sessions, they attended weekly meetings with the course instructors to go through the questions before their sessions with students. The peer tutors were expected to complete the question packet on their own before coming to the instructor meeting. During each meeting, the peer tutors collectively reviewed their answers to and justifications for each question.

Student Data

This study focuses on comparing the outcomes for struggling students who enrolled in optional peer tutoring with those struggling students who declined to enter peer tutoring. The peer-tutoring program represents one difference between

these two groups, but students may differ in other critical ways. To account for this, we compared these two groups of students using demographic, engagement, performance, and perception metrics. Although it is impossible to rule out significant differences in other, unmeasured aspects, these metrics provide a broad view of the types of students in each group and the factors that impact student performance.

Demographic Information

We considered a student's gender, ethnicity, year in school, and whether his or her home county is urban or rural (U.S. Department of Agriculture, 2003; Table 1). In addition we considered whether a student's major required additional biology courses in the future and his or her standardized test scores.

At UMaine, students can submit scores from the SAT and/or the ACT for admission (Table 2). For comparisons, we converted all ACT scores into SAT scores by first converting ACT scores to *z* scores using the national ACT statistics (mean = 21.1, SD = 5.3) and then converting these *z* scores to equivalent SAT scores using the national SAT statistics (mean = 1010, SD = 164.8). The formula for this conversion is

$$\left[\left(\frac{\text{student ACT score} - \text{national ACT mean}}{\text{national ACT SD}} \right) \times \text{national SAT SD} \right] + \text{national SAT mean} = \text{equivalent SAT score}$$

If a student submitted both SAT and ACT scores, his or her ACT score was converted to an SAT score, and the average of that score and the submitted SAT score was used for analysis.

In total, 53 students submitted both scores, and there was no significant difference between the mean true SAT score and the calculated equivalent SAT score (paired *t* test; mean difference = 12.9, SD = 75.3; *t*(52) = 1.25, *p* = 0.22). Students with neither SAT nor ACT scores available were excluded from any analyses in which SAT is used as a covariate.

Engagement Metrics

In addition to demographic information, we also collected four metrics from Synapse, a local course management system (CMS). CMS metrics included:

1. Synapse log-ins: The number of times a student logged into the Synapse CMS. This system acts as a hub for all lecture and lab materials and all communications from faculty members, teaching assistants, and peer tutors.
2. Note opening: Skeleton versions of the lecture slideshows were posted online in advance of each lecture. Students were able to access these notes beginning approximately 1 wk before the lecture and then again at any point over the remainder of the semester. Lecture notes were not available via any other method (e.g., they were not

handed out in class). We counted each time a student opened the page containing the lecture notes. Notes could be downloaded or printed, so we could not capture the total number of times during the semester that students viewed the notes outside the course management system.

3. Video and study guide viewing: For most lectures, the faculty members posted a condensed video version of the lecture along with an outline of key topics covered that day. The video could be viewed only within Synapse, but the study guide was available for download. These two items were posted on the same page, so we were unable to determine which resource a student accessed. Furthermore, there is no way to determine how long a student was on the page or what proportion, if any, of the video he or she watched. Thus, we counted the number of times a student opened a page containing these two resources as an index of access to both resources. Furthermore, video and study guide views are positively correlated with lecture attendance (Pearson's *r* = 0.15, *p* < 0.001), suggesting that students typically use these resources to supplement lecture rather than to replace it.
4. Grade book access: Student grades for both the lab and lecture components of the course were posted exclusively on Synapse. We counted the number of times each student accessed the grade book. Although it is difficult to know why some students check their grade book more often, this metric may indicate that students are taking an interest in their performance.

In this study, the term "engagement" refers to these four CMS metrics and lecture attendance. Lecture attendance was tracked using an in-class clicker system. A student was marked as attending as long as he or she answered at least one clicker question. Penalties were outlined in the syllabus if students were caught using multiple clickers to cover for friends who were not attending, and informal observation suggests this was not a widespread problem throughout the semester.

All engagement metrics were divided into two time periods 1) preintervention: from the start of the semester to the first exam; and 2) intervention: after the first exam until the end of the semester.

Course Performance

In addition to the demographic and engagement data, we also analyzed course performance. For the preintervention period, grades on the first exam were compared between struggling students who accepted and declined peer tutoring to assess whether these groups differed. For reference, the performance of these two groups was compared with non-struggling students who performed > 50% on exam 1.

For the intervention period, scores on exams 2–5 were compared between struggling students who accepted and declined peer tutoring. The exam performance of these two groups is then further compared with non-struggling students to place group performance in the context of the entire class.

Perceptions

Student perceptions of biology were compared using results from the CLASS-Bio survey (Semsar *et al.*, 2011). Student

Table 2. Standardized test scores available for BIO 100 students

Type of test taken	% of class
Only SAT	84.2
Only ACT	3.0
Both SAT and ACT	7.2
Neither SAT nor ACT	5.7

perceptions positively correlate with performance, and the perceptions of biology majors tend to improve over 4 yr of study (Hansen and Birol, 2014). Students took the CLASS-Bio online at the beginning (pretest) and end (posttest) of the semester and responded to 31 items such as, "To understand biology, I sometimes think about my personal experiences and relate them to the topic being analyzed," on a five-point Likert scale ranging from "strongly disagree" to "strongly agree." Students received a percent favorable score and were placed along a novice-to-expert continuum based on their alignment with the consensus opinions of biology PhDs.

For the preintervention period, pretest percent favorable scores were compared between the two struggling student groups to determine whether students who accepted peer tutoring held significantly different perceptions of biology than students who declined. As with exam performance, the perceptions of both struggling groups were compared with non-struggling students to provide context.

For tracking the relative changes in student perceptions during the course, percent favorable responses on the CLASS-Bio posttest were compared between struggling students who accepted peer tutoring, struggling students who declined peer tutoring, and non-struggling students. In addition, for students who completed both CLASS-Bio pretest and posttest, paired shifts from pretest to posttest scores were examined for struggling students who accepted and declined peer tutoring. A positive shift indicates that students gained more expert-like perceptions of biology during the semester, while a negative shift indicates they left class with more novice-like views. Previous research has found that students in introductory biology courses typically undergo negative shifts from the beginning to the end of a course (Semsar *et al.*, 2011). While the reasons for these novice-like shifts are still being explored, students in courses that incorporate both active learning and an explicit epistemological curriculum are more likely to show no or expert-like shifts from the beginning to the end of the course (Perkins *et al.*, 2005).

Statistical Models

For determining whether struggling students who accepted peer tutoring were different from students who declined peer tutoring, all demographic (Table 1) and preintervention engagement metrics were input into a single, multivariate, logistic regression. This model answers the question: If all demographic characteristics and measures of engagement are held constant except for factor X, does factor X significantly predict whether a struggling student accepted or declined peer tutoring?

To determine whether peer tutoring improved student performance after exam 1, we modeled a student's mean score on the four exams taken after the intervention began (exams 2–5) as a function of the number of peer-tutoring sessions each struggling student attended, using a multiple linear regression. Only students who attended at least one peer-tutoring session were included in this model. We input the following covariates to isolate the effect of the peer tutoring: demographic characteristics (Table 1) and the five measures of student engagement collected during the intervention period (four CMS metrics and lecture attendance).

This type of model, which has been used to explore similar education research questions (Hockings *et al.*, 2008; Theobald and Freeman, 2014), identifies which factors are associated with increased exam performance when all other factors are held constant.

Finally, to determine whether students who accept peer tutoring improve their perceptions of biology, we modeled CLASS-Bio posttest scores, which measure percent agreement with experts, with the following covariates: whether a student accepted peer tutoring, demographic characteristics (Table 1), five measures of engagement during intervention, and percent favorable scores on the CLASS-Bio pretest.

We tested the independent variables in the logistic and linear models for high collinearity (Pearson's $|r| > 0.70$). Pairs of factors with high correlation are marked in the results. Collinearity may obscure significant input factors in some cases (Dormann *et al.*, 2013), and therefore, in each case, the collinear factors were removed from the model one at a time, and these new models were compared with the old to identify any significant changes in the results.

Persistence

To determine whether persistence in the course was affected by the peer-tutoring intervention, we calculated the percentage of students who dropped the course by determining which students took the mandatory final exam; students who did not take the final exam were marked as dropped. The percentage of students who dropped was compared between struggling students who accepted peer tutoring and struggling students who declined peer tutoring using a Z test for two proportions. Students who dropped out during the course of the semester are excluded from all other analyses to keep groups comparable across the entire study.

Student Surveys

In the final peer-tutoring session, students were given a brief survey to assess their opinions about the peer-tutoring program. These surveys included three Likert-scale items for students to share their opinions on whether the peer-tutoring program helped them develop studying and test-taking skills both in BIO 100 and more generally.

To understand why some students did not participate in peer tutoring, we asked struggling students who declined to participate to fill out a free-response online question that reminded them about the peer tutoring and asked them to describe why they had declined help. Students were told that their answers would help the instructors better design future peer-tutoring programs. Responses were collected from 140 of the 215 struggling students who had declined peer tutoring (65.1%), and the reasons were coded into categories by two independent raters. Some students gave more than one reason in their response and therefore received more than one code. Interrater reliability was found to be acceptable ($\kappa = 0.67$, $p < 0.001$; Landis and Koch, 1977), and the raters discussed discrepancies to reach a consensus coding.

All statistical tests were performed using SPSS Version 21 (SPSS, IBM, Armonk, NY). Approval for this study was obtained from the university's institutional review board (UMaine IRB 2012-12-14).

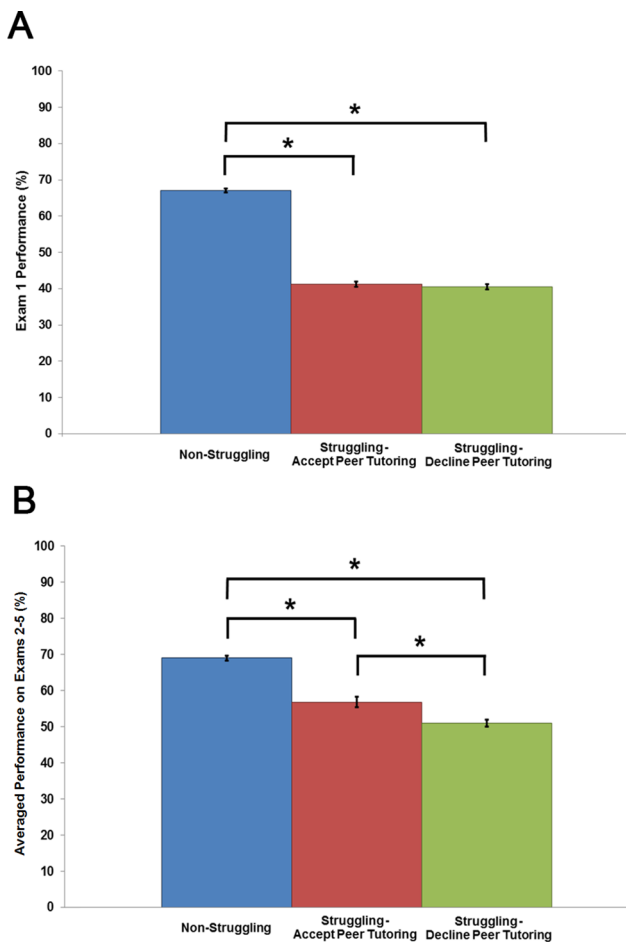


Figure 2. Comparison of exam performance among groups. (A) Mean preintervention exam 1 performance. One-way analysis of variance (ANOVA) $F(2, 681) = 638.6$; brackets with asterisks indicate significant two-tailed differences between groups, Tukey post hoc test, $p < 0.05$. (B) Mean intervention performance for exams 2–5. One-way ANOVA $F(2, 681) = 120.2$, $p < 0.001$; brackets with asterisks mark significant two-tailed differences between groups, Tukey post hoc test, $p < 0.05$. Bars indicate SE.

RESULTS

Do Struggling Students Who Accept Peer Tutoring Differ before Intervention from Those Who Decline?

BIO 100 students who scored $< 50\%$ on the first exam were invited to participate in the peer-tutoring sessions, and approximately one-third of struggling students accepted (Figure 1). To determine whether struggling students who accepted differed from those who declined, we examined the exam 1 performance of different groups before peer tutoring began (Figure 2A). The non-struggling students performed significantly higher when compared with struggling students who accepted and declined peer tutoring. Notably, the exam 1 scores of struggling students who went on to join peer tutoring were not significantly different from the exam 1 scores of students who declined peer tutoring.

In addition to examining exam 1 scores, we also compared CLASS-Bio (Semsar *et al.*, 2011) scores on the pretest

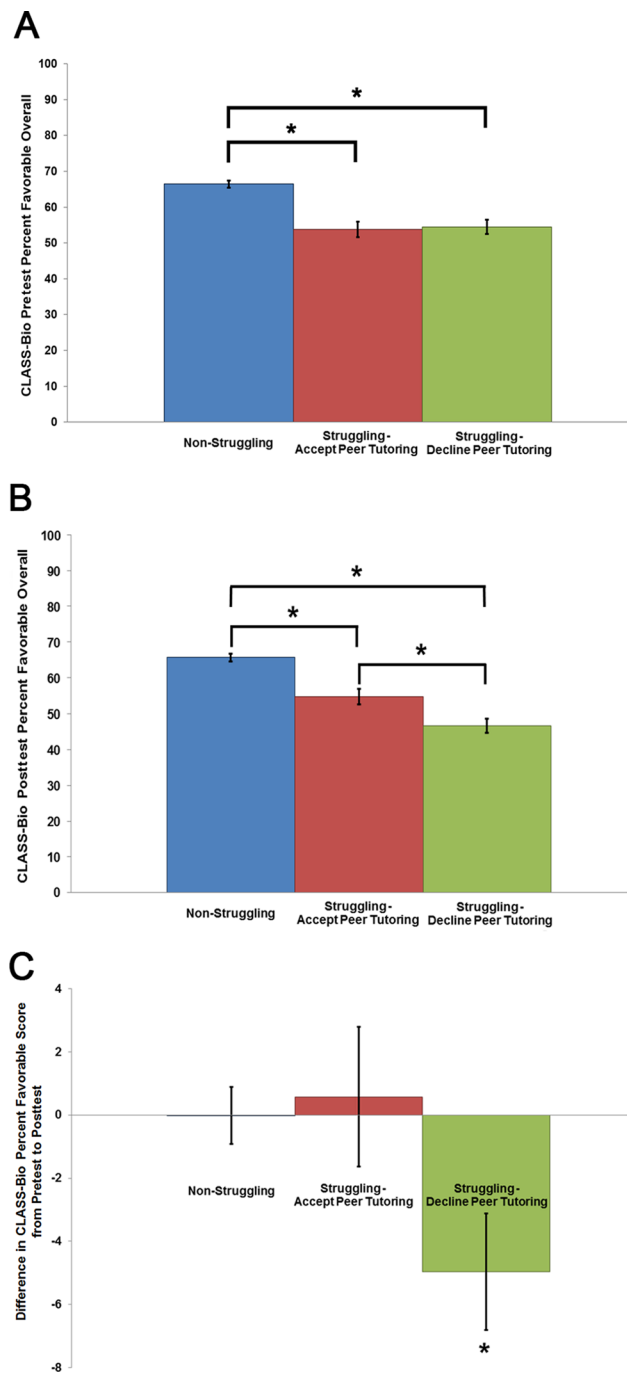


Figure 3. Student CLASS-Bio scores at the beginning and end of the semester. (A) Mean scores overall on CLASS-Bio pretest by group. One-way ANOVA $F(2, 555) = 26.7$, $p < 0.001$; brackets with asterisks mark significant two-tailed differences between groups, Tukey post hoc test, $p < 0.05$. (B) Mean scores on CLASS-Bio posttest overall by group. One-way ANOVA $F(2, 641) = 44.0$, $p < 0.001$; brackets with asterisks mark significant two-tailed differences between groups, Tukey post hoc test, $p < 0.05$. (C) Mean shifts in CLASS-Bio overall score for individual students from pretest to posttest. The significant two-tailed shift is marked with an asterisk, paired t -test, $p < 0.05$. Bars indicate SE.

given at the start of the semester (Figure 3A). Non-struggling students had significantly higher agreement with experts compared with struggling students who accepted and

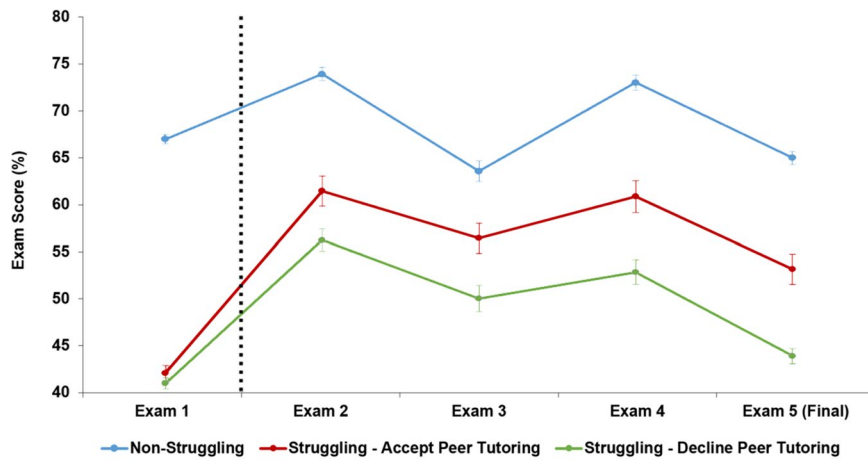


Figure 4. Raw average exam grades throughout the semester by group. Bars indicate SE. The dashed black line indicates the start of the intervention period.

declined peer tutoring. Furthermore, struggling students who accepted peer tutoring began the semester having perceptions of biology equivalent to the perceptions of struggling students who declined peer tutoring.

Next, all of our demographic and preintervention engagement variables for the struggling students were included in a logistic regression to determine whether any of our measured characteristics predicted whether or not a struggling student accepted peer tutoring. The resulting model was not statistically significant compared with an intercept-only model ($\chi^2(11) = 16.77, p = 0.12$), indicating that these input variables do not significantly predict whether or not a student would accept peer tutoring.

Do Struggling Students in the Peer-Tutoring Program Improve Their Course Performance?

Given that students who accepted and declined peer tutoring were similar at the beginning of the course, we next wanted to determine whether the peer-tutoring program helped students improve their performance on subsequent exams. Students who accepted peer tutoring showed improvement in performance relative to struggling students who declined peer tutoring immediately after the intervention, and the differences were consistent throughout the semester (Figure 4). Furthermore, when averaged scores on exams 2–5 were compared, non-struggling students outperformed struggling students; however, the exam performance of struggling students who accepted peer tutoring was significantly better than struggling students who declined peer tutoring (Figure 2B). Moreover, students in the peer-tutoring program with the greatest improvement in their exam performance attended, on average, more peer-tutoring sessions (Supplemental Figure 3).

These results suggest that attending peer tutoring helped struggling students improve their exam performance. However, a number of other factors may explain improved exam performance. Therefore, we ran a multiple regression model to identify the impact of peer-tutoring attendance among these factors. As students attended a variable number of peer-tutoring sessions (Supplemental Figure 4), we designated peer-tutoring attendance as a continuous variable. The model is significant compared with an intercept-only model ($F(12, 76) = 5.28, p < 0.001$) and accounts for 36.9% (adjusted R^2) of the variance in the data (Table 3).

After controlling for differences in demographic characteristics and student engagement, the model predicts that struggling students who accepted peer tutoring would gain 0.99 percentage points on their mean 2–5 exam scores for each tutoring session they attended ($p = 0.03$; Table 3). There were 10 available sessions, so struggling students who attended all weekly sessions would be expected to score approximately one letter grade higher on these last four exams on average than struggling students who attended just one peer-tutoring session.

Table 3. Multiple regression model of factors contributing to mean performance on intervention period exams for students attending at least one peer-tutoring session

	$\beta \pm SE^a$	p
Peer-tutoring program		
Number of sessions attended	0.99 ± 0.44	0.03
Engagement		
Video/study guide views	0.15 ± 0.05	<0.01
Lectures attended	0.53 ± 0.23	0.02
Note openings	-0.05 ± 0.06	0.35
Grade book checks ^b	0.02 ± 0.04	0.64
Synapse log-ins ^b	0.01 ± 0.04	0.79
Demographics		
SAT score	0.04 ± 0.01	<0.01
Year in school (1 → 2 → 3 → 4)	2.59 ± 2.03	0.20
Gender (male → female)	2.49 ± 2.83	0.38
Home county (rural → urban)	2.45 ± 2.82	0.39
Require more biology? (no → yes)	0.92 ± 2.77	0.74
Ethnicity (nonminority → minority)	0.39 ± 3.31	0.91

Factors with p values <0.05 are indicated in bold font.

^a β indicates the expected change in the mean exam score (in percentage points) given a unit change in the parameter of interest. For categorical variables, β indicates the predicted change in score that results from a categorical change in the direction indicated by the arrows in parentheses (e.g., changing the level of “gender” from “male” to “female” increases the expected mean exam score by 2.49 percentage points).

^bThese factors are highly correlated ($r = 0.90$), but removing one of them does not change which input variables are significant or the β values of significant independent variables. Supplemental Figure 5 shows the results from the multiple regression model when the factor Synapse log-ins is removed.

Average performance on the four intervention period exams was also significantly predicted by video and study guide viewing ($\beta = 0.15$, $p < 0.001$) and lecture attendance ($\beta = 0.53$, $p = 0.02$; Table 3). More frequent viewing of videos and study guides and higher lecture attendance are significantly associated with higher exam scores. Additionally, higher SAT scores are associated with better performance on these exams ($\beta = 0.04$, $p < 0.001$). However, it is notable that, after adjusting for the impact of these factors, peer-tutoring attendance still has a significant impact on exam performance.

Several factors do not significantly impact exam performance after adjusting for other factors ($p \geq 0.20$; Table 3). These factors include: differences in note opening, grade book checks, Synapse log-ins, year in school, gender, urban/rural hometown, future biology requirements for major, and minority status.

Do Struggling Students in Peer Tutoring Improve Their Perceptions of Biology?

In addition to their poor performance early in the year, struggling students who accepted and declined peer tutoring entered the class holding more novice-like perceptions of biology than their non-struggling peers (Figure 3A). At the end of the semester, struggling students who accepted and declined peer tutoring still held more novice-like perceptions of biology than their non-struggling peers, but struggling students who accepted peer tutoring ended the year with significantly higher CLASS-Bio scores than struggling students who declined (Figure 3B). In addition, struggling students who declined peer tutoring shifted significantly toward more novice-like perceptions of biology from pretest to posttest, while struggling students in peer tutoring maintain their initial perceptions (Figure 3C).

Next, we ran a multiple regression model among the struggling students to determine whether accepting peer tutoring was significantly associated with higher CLASS-Bio posttest scores after controlling for demographic, intervention period engagement, and CLASS-Bio pretest score factors. The majority of students included in this model declined peer tutoring and accordingly attended zero peer-tutoring sessions. To avoid zero inflation in our previously continuous peer-tutoring attendance variable, we defined peer tutoring in this analysis as a binary variable (accept or decline).

The model is significant compared with an intercept-only model ($F(13,176) = 13.86$, $p < 0.001$) and accounts for 46.9% (adjusted R^2) of the variance in CLASS-Bio posttest scores (Table 4). After controlling for differences in student engagement, demographic characteristics, and CLASS-Bio pretest score, it was predicted that struggling students who accepted peer tutoring would score 6.05 points on the CLASS-Bio posttest ($p = 0.02$) compared with students who declined peer tutoring.

Also, CLASS-Bio pretest and posttest scores are positively correlated (Table 4). After accounting for all other variables in the model (including peer-tutoring participation) for each point scored on the CLASS-Bio pretest, students scored 0.71 points more favorable on the CLASS-Bio posttest ($p < 0.01$).

In addition, a number of other factors showed interesting but nonsignificant trends ($0.05 < p < 0.2$) on CLASS-Bio

Table 4. Multiple regression model of factors contributing to CLASS-Bio posttest scores for struggling students

	$\beta \pm SE^a$	p
Peer-tutoring program		
Accepted peer tutoring	6.05 ± 2.49	0.02
Engagement		
Video/study guide views	0.09 ± 0.05	0.09
Lectures attended	0.36 ± 0.24	0.13
Synapse log-ins ^b	-0.03 ± 0.05	0.51
Note openings	-0.03 ± 0.05	0.56
Grade book checks ^b	0.02 ± 0.04	0.63
Demographics		
Gender (male \rightarrow female)	5.05 ± 2.85	0.08
SAT score	-0.01 ± 0.01	0.09
Year in school (1 \rightarrow 2 \rightarrow 3 \rightarrow 4)	-0.68 ± 2.15	0.13
Ethnicity (nonminority \rightarrow minority)	-2.61 ± 3.26	0.42
Home county (rural \rightarrow urban)	1.71 ± 3.27	0.60
Require more biology? (no \rightarrow yes)	0.33 ± 2.93	0.91
CLASS-Bio		
Pretest score	0.71 ± 0.06	<0.01

Factors with p values < 0.05 are indicated in bold font.

^a β indicates the expected change in the CLASS-Bio posttest score given a unit change in the parameter of interest. For categorical variables, β indicates the predicted change in score that results from a categorical change in the direction indicated by the arrows in parentheses (e.g., changing the category of "gender" from "male" to "female" increases the expected CLASS-Bio posttest score by 5.05 points).

^bThese factors are highly correlated ($r = 0.90$), but removing one of them does not change which input variables are significant or the β values of significant independent variables. Supplemental Figure 6 shows the results from the multiple regression model when the factor Synapse log-ins is removed.

posttest scores (Table 4). These trends may be too variable or weak, given the statistical power of our test. For example, female students are predicted to end the semester with more expert-like perceptions ($\beta = 5.05$, $p = 0.08$) compared with similar male students. Other factors approaching significance in the model include number of video/study guide views, the number of lectures attended, SAT scores, and year in school.

Do Struggling Students in Peer Tutoring Persist at a Higher Rate?

Ultimately, the goal of this tutoring program is to improve the persistence of struggling students in introductory biology. To evaluate this goal, we compared the overall percent of students dropping out of BIO 100 with the dropout rate for each group and found that struggling students who declined peer tutoring had the highest dropout rate of any group (Table 5). Moreover, the dropout rate of struggling students who declined peer tutoring is significantly higher than struggling students who accepted peer tutoring (Z test for two proportions, $Z = 3.25$, $p < 0.001$).

What Do Struggling Students Think of Peer Tutoring?

During the final peer-tutoring session, a brief survey was given to assess student opinions of the peer-tutoring program

Table 5. Dropout rate by group

	Dropouts	
	<i>n</i>	% of class
Full class	92	12.1
Study group		
Non-struggling	25	5.8
Struggling—accept peer tutoring	14	12.0
Struggling—decline peer tutoring	53	24.7

(Table 6). Students who participated in the peer-tutoring programs overwhelmingly said the peer tutoring was helpful for both the BIO 100 course and their other courses.

Why Do Some Struggling Students Decline Peer Tutoring?

To explore the reasons why struggling students declined help, we gave these students an online free-response survey at the end of the semester. This survey reminded them about the peer-tutoring offer and asked them why they declined. Responses were collected from 65.1% of the struggling students who had declined peer tutoring and coded by two independent raters (Figure 5). Some students gave more than one reason in their response and therefore received more than one code. Struggling students most commonly reported that they declined to participate in the peer-tutoring program because they did not have enough time.

DISCUSSION

This study used peer tutors to provide targeted help to students who did poorly on the first exam in a large-enrollment, introductory biology course. Students who struggle on the first exam in introductory biology typically fail to improve significantly over the remainder of the semester (Jensen and Moore, 2008). While previous research has described the improvements in student performance, perceptions, and attitudes as a result of offering peer tutoring to the entire class (e.g., Born *et al.*, 2002; Hockings *et al.*, 2008; Stanger-Hall *et al.*, 2010), we report here how those benefits can be replicated within a subset of struggling students who are participating in ungraded, extracurricular sessions. Notably in this study, struggling students

Table 6. Struggling student evaluation of peer tutoring

Item	Mean ^a	SD
The peer-tutoring sessions helped me prepare for exams in BIO 100.	4.48	0.55
The peer-tutoring sessions helped me study more effectively for BIO 100.	4.40	0.71
The peer-tutoring sessions helped me develop test-taking skills that are valuable in my other courses.	4.18	0.71

Survey responses *n* = 44

^aOn the survey, 1 was “strongly disagree” and 5 was “strongly agree.”

who regularly attended peer-tutoring sessions improved their exam performance (Figures 2 and 4 and Table 3) and persisted at a higher rate (Table 5). Students who participated in peer tutoring also finished the course with more expert-like perceptions of biology (Figure 3B and Table 4) compared with struggling students who declined peer-tutoring sessions.

Because institutions may not have the resources to hire and train enough peer tutors to support class-wide implementation of peer tutoring, it is important to know that peer tutoring offered to struggling students is effective. Furthermore, the improved performance of struggling students who accepted peer tutoring is not the result of a biased self-selection in the measured characteristics (Figures 2A and 3A). It is possible that our metrics do not document some important variation between our groups. However, it is worth noting that, before the widespread adoption of online CMSs like Synapse, the adjustments for student engagement presented here would not have been possible. Given continued advances in classroom technology, future studies may capture additional variations to better characterize the impact of targeted intervention strategies.

Why Is Peer Tutoring Effective for Struggling Students?

In this study, peer tutoring refers to a system of student support, including weekly meetings with up to 10 students, a trained peer tutor, and prepared question packets given to the students each week. One plausible explanation for the benefits of this program is that the struggling students who accepted peer tutoring improved their exam performance simply because they were spending an extra hour per week studying. However, previous work has found a negligible relationship between the amount of time spent studying and academic performance (Nonis and Hudson, 2010; Masui *et al.*, 2012), suggesting that the quality of studying techniques is also important.

High-quality study techniques can be described as an extension of the deliberate practice model of expert development (Plant *et al.*, 2005). The deliberate practice framework proposes that developing expertise in a domain requires:

1. Regularly scheduled training in that domain;
2. Training tasks designed by an expert teacher;
3. Ongoing feedback from an expert teacher during training tasks; and
4. Intense concentration during training with minimal distractions (Ericsson, 2006; Ericsson *et al.*, 1993).

The peer-tutoring sessions described in this study were explicitly designed to follow this framework. Struggling students met the first condition by signing up for weekly tutoring sessions. The second condition was met, because the materials provided in these sessions were designed by the course instructors and research investigators to align with course goals. The third condition was also met, because students received feedback and clarification from their trained peer tutors as they worked. Finally, students met the last condition by meeting with peer tutors who monitored the group and encouraged their students to stay on task throughout the session.

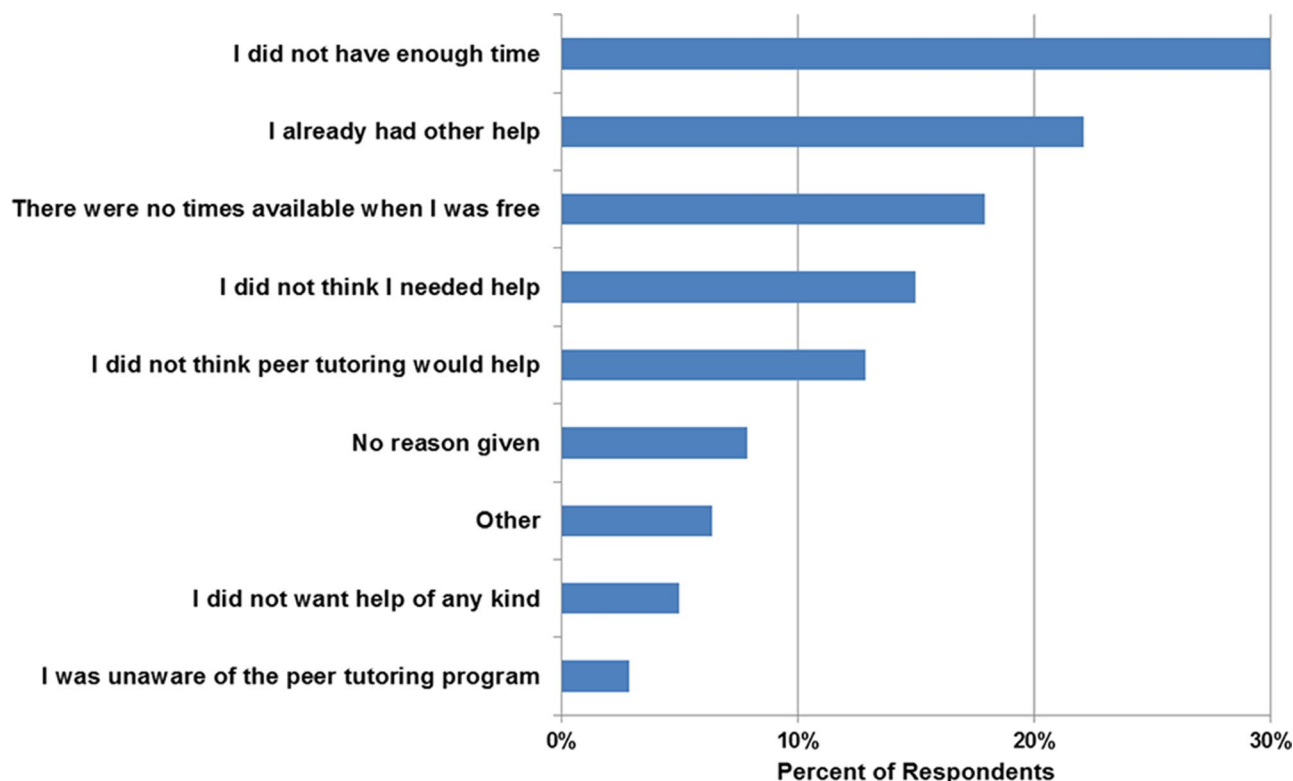


Figure 5. Reasons struggling students give for declining peer tutoring. Some students gave more than one reason in their response and therefore received more than one code.

While the deliberate practice framework stresses that all of these factors are important, it is possible that the observed improvements in performance, perceptions, and persistence among struggling students could be partly or mostly replicated without all the elements currently included in the peer-tutoring program. Additional research isolating or removing components of the peer program will help identify which aspects are most critical to its success. For example, providing the materials used in peer-tutoring sessions to all students would help determine the specific benefits of simply having access to the questions in the format used in the sessions. Alternatively, struggling students could register for self-guided study groups in which they would work through the materials without the guidance of a peer tutor.

Why Use Targeted Peer Tutoring for Struggling Students?

Hiring enough peer tutors for every student in a large-enrollment class can be a resource-intensive process. However, other limited-resource interventions have also been shown to be effective, such as two-part metacognitive interventions (Lizzio and Wilson, 2013) and a single study skills intervention run by faculty members (Deslauriers *et al.*, 2012). Why, then, should institutions with limited resources choose to implement a weekly peer-tutoring program?

One advantage of a weekly, targeted peer-tutoring program over one-time interventions is that a weekly program fosters the development of a learning community

that can help students who feel isolated. Isolation from the educational community can lead to disengagement and low persistence among students in large introductory courses (Tinto, 1993; Evans *et al.*, 2001). Learning communities have been found to combat isolation and improve attitudes, learning outcomes, and motivation for students in STEM fields (Freeman *et al.*, 2008).

Peer tutoring is also unique in that it benefits the students acting as tutors. Outcomes for peer tutors were not specifically investigated in this paper, but previous research has consistently shown that peer tutors improve their content mastery as well as their perceptions of the respective field (Otero *et al.*, 2010; University of Colorado Learning Assistant Program, 2011). In this way, a targeted peer-tutoring program maximizes the benefits of limited resources by significantly improving the performance of struggling students in an introductory course, by increasing retention of STEM majors, and by helping peer tutors develop teaching skills and further mastery in a STEM field.

Reaching Students Who Declined Peer Tutoring

Peer tutoring can improve the performance and persistence of struggling introductory biology students, but this optional program can only help those who choose to attend. This self-selection process means that students who enrolled in the program could have potentially been quite different from struggling students who did not. However, analysis of the performance, perceptions, demographic characteristics, and engagement of the students in these two groups

showed they were similar to the students who declined the assistance (Figures 2A and 3A). These results suggest that, while the background, engagement, and early performance of STEM students are important (Tinto, 1993; Seymour, 2000; Ost, 2010; Watkins and Mazur, 2013), the measures presented here are not sufficient for determining which students are likely to accept help.

Despite the benefits of the peer-tutoring program, not all struggling students participate. For example, in this study, 35.2% of students accepted the offered peer tutoring after struggling on the first exam (Figure 1). Previous studies offering ungraded, extracurricular help to struggling students in introductory psychology courses (Lizzio and Wilson, 2013), struggling students in introductory physics and oceanography courses (Deslauriers *et al.*, 2012), and all students in an introductory chemistry course (Hockings *et al.*, 2008) had acceptance rates of 40.5, 39.1, and 40.0%, respectively. The similarity of these acceptance rates suggests that student interest in this type of assistance may be somewhat constant over a range of classroom settings.

The participation rates for struggling students leads to this question: Why are some students not interested in taking advantage of available extra help? For this peer-tutoring program, many students reported they were unable to make time for a weekly meeting, did not see that they needed help, or chose other available help in lieu of the peer-tutoring program (Figure 5). Accordingly, we suggest that future recruitment efforts should highlight for students the differential exam performance between struggling students in and out of peer tutoring. Furthermore, students frequently reported scheduling conflicts during available tutoring times; future registration systems should include a wait-list component so program directors are aware of additional students wanting to register for a popular time slot and can shift peer tutors from less popular times to meet demand.

CONCLUSION

This study shows that some students who do poorly on the first exam in a course will participate in a targeted tutoring program, even though the program is ungraded and requires additional time investment. Students who accepted peer-tutoring sessions achieved improved exam performance and ended the year with more expert-like perceptions of biology relative to their struggling peers who declined help. Ultimately, these improvements led to increased persistence in the class and suggest that targeted peer tutoring can be an effective tool for reducing the loss of at-risk students during the critical, early undergraduate portion of the STEM pipeline.

ACKNOWLEDGMENTS

The authors appreciate the participation by the BIO 100 students. We thank Mary Tyler, Karen Pelletreau, and Natasha Speer for helpful feedback on the manuscript. We also thank Erin Vinson for assistance coding student surveys. Thank you to Ryan Cowan, Ron Kozlowski, and the rest of the Synapse staff for assembling data for this analysis. This work is supported by National Science Foundation grant 0962805.

REFERENCES

- American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.
- Bloom BS, Englehard MD, Furst EJ, Hill WH, Krathwohl DR (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I, Cognitive Domain, New York: David McKay.
- Born WK, Revelle W, Pinto LH (2002). Improving biology performance with workshop groups. *J Sci Educ Tech* 11, 347–365.
- Chasteen SV, Pepper RE, Caballero MD, Pollock SJ, Perkins KK (2012). Colorado Upper-Division Electrostatics diagnostic: a conceptual assessment for the junior level. *Phys Rev Spec Top Phys Educ Res* 8, 020108.
- Chen X, Soldner M (2013). STEM Attrition: College Students' Paths into and out of STEM Fields, Washington, DC: National Center for Education Statistics.
- Deslauriers BL, Harris SE, Lane E, Wieman CE (2012). Transforming the lowest-performing students: an intervention that worked. *J Coll Sci Teach* 41, 76–84.
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Marquez JRG, Gruber B, Lafourcade B, Leitaio PJ, *et al.* (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46.
- Ericsson KA (2006). The influence of experience and deliberate practice on the development of superior expert performance. In: The Cambridge Handbook of Expertise and Expert Performance, ed. KA Ericsson, N Charness, RR Hoffman, and FJ Feltovich, Cambridge, UK: Cambridge University Press, 685–706.
- Ericsson KA, Krampe RT, Tesch-Romer C (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 100, 363–406.
- Evans W, Flower J, Holton D (2001). Peer tutoring in first-year undergraduate mathematics. *Int J Math Educ Sci Technol* 32, 161–173.
- Freeman KE, Alston ST, Winborne DG (2008). Do learning communities enhance the quality of students' learning and motivation in STEM? *J Negro Educ* 77, 227–240.
- Gosser DK (2011). The PLTL boost: a critical review of research. *Progressions* 14, (1).
- Gosser DK, Cracolice MS, Kampmeier JA, Roth V, Strozak VS, Varma-Nelson P (2001). Peer-Led Team Learning: A Guidebook, Upper Saddle River, NJ: Prentice Hall.
- Hansen MJ, Birol G (2014). Longitudinal study of student attitudes in a biology program. *CBE Life Sci Educ* 13, 331–337.
- Hockings SC, DeAngelis KJ, Frey RF (2008). Peer-led team learning in general chemistry: implementation and evaluation. *J Chem Educ* 85, 990–996.
- Hooker D (2011). Small peer-led collaborative learning groups in developmental math classes at a tribal community college. *Multicult Perspect* 13, 220–226.
- Jensen PA, Moore R (2008). Students' behaviors, grades and perceptions in an introductory biology course. *Am Biol Teach* 70, 483–487.
- Landis JR, Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Lizzio A, Wilson K (2013). Early intervention to support the academic recovery of first-year students at risk of non-continuation. *Innov Educ Teach Int* 50, 109–120.
- Masui C, Broeckmans J, Doumen S, Groenen A, Molenberghs G (2012). Do diligent students perform better? Complex relations between student and course characteristics, study time, and academic performance in higher education. *Stud High Educ* 39, 621–643.

- Murtaugh PA, Burns LD, Schuster J (1999). Predicting the retention of university students. *Res High Educ* 40, 355–371.
- Nonis SA, Hudson GI (2010). Performance of college students: impact of study time and study habits. *J Educ Bus* 85, 229–238.
- Ost B (2010). The role of peers and grades in determining major persistence in the sciences. *Econ Educ Rev* 29, 923–934.
- Otero V, Pollock S, Finkelstein N (2010). A physics department's role in preparing physics teachers: the Colorado learning assistant model. *Am J Phys* 78, 1218.
- Perkins KK, Adams WK, Pollock SJ, Finkelstein ND, Wieman CE (2005). Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey. In: *AIP Conference Proceedings*, ed. J Marx, P Heron, and S Franklin, Philadelphia, PA: IOP Publishing.
- Plant EA, Ericsson KA, Hill L, Asberg K (2005). Why study time does not predict grade point average across college students: implications of deliberate practice for academic performance. *Contemp Educ Psychol* 30, 96–116.
- Pollock S (2009). Longitudinal study of student conceptual understanding in electricity and magnetism. *Phys Rev Spec Top Phys Educ Res* 5, 020110.
- President's Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Washington, DC: U.S. Government Office of Science and Technology.
- Preszler RW (2009). Replacing lecture with peer-led workshops improves student learning. *CBE Life Sci Educ* 8, 182–192.
- Semsar K, Knight JK, Birol G, Smith MK (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in Biology. *CBE Life Sci Educ* 10, 268–278.
- Seymour E (2000). *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview.
- Stanger-Hall K, Lang S, Maas M (2010). Facilitating learning in large lecture classes: testing the “teaching team” approach to peer learning. *CBE Life Sci Educ* 9, 489–503.
- Theobald R, Freeman S (2014). Is it the intervention or the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE Life Sci Educ* 13, 41–48.
- Tinto V (1993). *Leaving College*, Chicago, IL: University of Chicago Press.
- Topping KJ (2000). *Peer Assisted Learning: A Practical Guide for Teachers*, Newton, MA: Brookline.
- University of Colorado Learning Assistant Program (2011). *LA-TEST, Annual Report to the NSF*. <http://laprogram.colorado.edu/sites/default/files/reports/la-program-10-11.pdf> (accessed 10 March 2015).
- U.S. Department of Agriculture (2003). *Urban–Rural Continuum Code*. www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx#.VEa3rucYV7Y.
- Watkins J, Mazur E (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. *J Coll Sci Teach* 42, 36–41.