

# Does the Format of Preclass Reading Quizzes Matter? An Evaluation of Traditional and Gamified, Adaptive Preclass Reading Quizzes

Elena P. Kolpikova,<sup>†</sup> Derek C. Chen,<sup>†</sup> and Jennifer H. Doherty\*

Department of Biology, University of Washington, Seattle, WA 98105

## ABSTRACT

Preclass reading quizzes (RQs) have been shown to enhance student performance. Many instructors implementing evidence-based teaching assign preclass RQs to ensure their students are prepared to engage in class activities. Textbook companies now offer a gamified, adaptive-learning RQ format. In these RQs, students answer point-valued questions until they reach a threshold. If students answer incorrectly, the question decreases in point value on the next attempt. These RQs also give students who answer questions incorrectly more questions on that topic and direct students to sections of a textbook they need to review. We assessed the impact of gamified, adaptive preclass RQs compared with more traditional preclass RQs on in-class RQs and course exam performance as well as students' perceptions of RQs. Students in the gamified, adaptive treatment performed equally compared with students in the traditional, static treatment on in-class RQs and course exams. While students in the gamified, adaptive treatment did have a more positive perception of preclass RQs, this factor explained less than 3% of the variation in RQ perception. Our findings suggest that instructors should verify that gamified, adaptive technologies impact student learning in their course before integrating them into their course and asking students to pay for them.

## INTRODUCTION

Active-learning teaching strategies are more effective than traditional lecture-style teaching at improving student learning (Freeman *et al.*, 2014). Active-learning instruction has also been demonstrated to enhance student's self-reported engagement and satisfaction with a course (Armbruster *et al.*, 2009). One particularly effective active-learning course design is "high structure," in which students obtain basic course content before coming to class to prepare them for the more cognitively demanding in-class active-learning exercises (Freeman *et al.*, 2011; Eddy and Hogan, 2014; O'Flaherty and Phillips, 2015). A common method instructors use to encourage students to read before coming to class is to assign preclass reading quizzes (RQs). A common format for preclass RQs is open-book, multiple choice quizzes based on a textbook reading. The goal of preclass RQs is to motivate students to read their textbook in order to learn basic vocabulary and to explore the upcoming course topic (Heiner *et al.*, 2014). Studies have shown that implementing preclass RQs result in increased exam performance (Johnson and Kiviniemi, 2009; Moravec *et al.*, 2010; Freeman *et al.*, 2011; Pape-Lindstrom *et al.*, 2018).

The first generation of discipline-based education research has shown that preclass RQs can improve student performance, but preclass RQs come in many formats (e.g., multiple choice, short answer, adaptive). At this time, there is little research on whether a particular preclass RQ format is more effective at preparing students to successfully engage in a highly active-learning environment (for an example, see Moravec *et al.*, 2010). Furthermore, students are not a monolithic group of learners, and different RQ

Cynthia Brame, *Monitoring Editor*

Submitted Jun 3, 2019; Revised Aug 5, 2019;

Accepted Aug 23, 2019

CBE Life Sci Educ December 1, 2019 18:ar52

DOI:10.1187/cbe.19-05-0098

Conflict of Interest: A Macmillan Learning employee assisted the course instructor (J.H.D.) in setting up the online RQ assignments, as she does for many faculty piloting electronic resources associated with Macmillan Learning biology textbooks. She also assisted in data collection by downloading student answers to RQ questions, the number of questions answered by students in the gamified, adaptive treatment, and the time each student spent on each RQ. These data were used to set the number of questions students in the trad-RQ treatment received, but in no other analyses. Macmillan employees had no involvement in the design of the study, analysis or interpretation of the data, or in the preparation, review, or approval of the article.

<sup>†</sup>Co-first authors.

\*Address correspondence to: Jennifer H. Doherty (doherty2@uw.edu); ORCID: 0000-0002-2333-1692.

© 2019 E. P. Kolpikova, D. C. Chen, and J. H. Doherty. CBE—Life Sciences Education © 2019 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.

formats may have a differential impact on preparedness for the diversity of students in our classrooms. We consider each of these issues, preclass RQ format and disaggregating students, to be finer-grained research questions and examples of second-generation discipline-based education research that will contribute to fully maximizing the learning experience for all of our students (Freeman *et al.*, 2014; Dolan, 2015).

Given the demonstrated value of preclass RQs, many textbook companies are now selling educational supplementary material that incorporates software that uses adaptive learning (e.g., PrepU NCLEX-RN 10000, McGraw Hill LearnSmart, Macmillan LearningCurve). Adaptive-learning assignments give students who answer questions on a course topic incorrectly more questions on that topic. This software may also direct students who answer questions incorrectly to specific pages or figures within the textbook to gain the information required to correctly answer the question.

Some preclass RQ formats produced by textbook companies have also incorporated an additional feature known as gamification to encourage students to complete preclass RQ assignments (e.g., McGraw-Hill LearnSmart, Macmillan LearningCurve). Gamification uses game elements in a non-game context to increase student enjoyment and thus motivate students to better engage with the course material (Cheong *et al.*, 2013). These types of assignments require students to continuously answer point-valued questions until they reach or surpass the point-value threshold defined by the particular homework assignment. However, if a student initially answers a question incorrectly, the question decreases in point value over multiple attempts. Including the adaptive elements in conjunction with the gamification elements is theorized to increase student affect and performance within the classroom (Gordon *et al.*, 2013).

Currently, the literature on the effectiveness of adaptive or gamified, adaptive learning within higher education presents an inconclusive assessment of the impact of these activities on student learning. In a study of nursing students using the adaptive-learning PrepU NCLEX-RN 10,000 software to prepare for their National Council Licensure Examination (NCLEX-RN), a positive correlation between adaptive quizzing-system usage and content mastery was found. However, in a study on students in an introductory psychology course, students who used LearningCurve gamified, adaptive software learned the same as those who did not (Becker-Blease and Bostwick, 2016). These findings are supported by those of James (2012), who showed that, in an introductory biology course at a 4-year university, exam performance did not differ significantly between students who used LearnSmart gamified, adaptive software and those who did not. However, findings by Gurung (2015) showed that psychology students at a 4-year university who used more adaptive technologies (McGraw-Hill LearnSmart, Worth PsychPortal Worth, or Cengage Alpia) performed better on exams and assignments.

Interestingly, in a larger study involving students who used LearnSmart gamified, adaptive software in their introductory anatomy and physiology courses at six 4-year and 2-year institutions, Griff and Matter (2013) found that, among 4-year universities, there was no significant improvement between students who used the LearnSmart and those who did not. However, students who attended 2-year institutions did demonstrate greater learning. Similarly, in a general chemistry course

at a large state university, Richards-Babb and colleagues (2018) found that the use of an adaptive online homework system correlated with improved letter grades only for students at average or below.

Research on students' perceptions of these assignments is also inconclusive. Students' perceptions of instruction are important, because they influence how students engage with the lecture material (Trujillo and Tanner, 2014). Becker-Blease and Bostwick (2016) found that, while students liked and valued the gamified, adaptive format, they did not learn more while using the software. Griff and Matter (2013) found the same pattern at the 4-year universities they investigated. Findings by Zumalt and Williamson (2016) suggest that the existence of this preference may be due to direct links to relevant textbook sections that are embedded within questions in adaptive-learning systems. Conversely, Richards-Babb and colleagues (2018) found that, despite selectively increased performance among students, the students themselves preferred the traditional rather than the adaptive format.

We investigated the effectiveness of two formats of preclass multiple-choice RQ: traditional, static RQs (trad-RQs) and gamified, adaptive RQs (adapt-RQ; i.e., Macmillan LearningCurve). This study addresses three research questions designed to help educators better understand how the format of the preclass RQ may impact student learning and whether there is a differential impact on student groups compared with traditional RQs: 1) Do gamified, adaptive RQs improve students' preparedness for class? 2) Do gamified, adaptive RQs improve students' exam performance? 3) Do gamified, adaptive RQs positively impact students' perceptions of a) RQs and b) the course?

We hypothesized that preclass RQs that used a gamified, adaptive format would improve all students' preparedness for class due to one or more of the following reasons. 1) The gamified, adaptive format could discourage students from rushing through the questions, as erroneous answers would cost them points, which would lengthen the time needed to reach the threshold value to complete the assignment. Taking a more deliberate approach to answering questions could lead to better preparation for class material. 2) The gamified, adaptive format could increase practice with specific course topics that each student is struggling with. As students are given more questions on topics they answer incorrectly, the adapt-RQ focuses students' preparatory efforts on their knowledge gaps. 3) The gamified, adaptive format could increase the time students spend reading the textbook. The adapt-RQ provides links directly to relevant textbook sections; again, this helps focus the students' efforts and gives them additional forms of learning resources. 4) Students could enjoy the gamified, adaptive format more due to the gamified aspects and might be motivated to better engage with the course material due to the gamification aspects (points, hints, and textbook links). We hypothesized that these four elements, unique to gamified, adaptive-learning RQs, could increase student preparedness for class, which in turn increases learning during class activities and contributes to higher exam performance. Finally, we hypothesized that, if students preferred the adapt-RQs or if the adapt-RQs prepared students better to engage with course material, this could collectively translate into increased enjoyment of the course.

## METHODS

### Participants and Setting

This study was conducted with 576 undergraduate students from the University of Washington (UW), a very high research activity and more selective institution (Carnegie Classification of Institutions of Higher Education, n.d.). The students were enrolled in the third quarter (10 weeks) of a three-quarter introductory biology series that introduces plant and animal physiology. To register for the class, the students were required to have a passing grade (at least a 2.0 on a 4.0 scale) in the previous course in the series. The course met for 50-minute lectures five times a week and students attended a 2.5-hour laboratory section weekly. In the quarter this study was completed, there were two offerings for the lecture (276 students in the first offering, 312 in the second) that were taught back-to-back. Each lecture offering was associated with 12 or 13 smaller laboratory sections, with 20–24 students per section. Both lecture offerings had the same instructor (author J.H.D.), curriculum and exams. The instructor implemented frequent active-learning strategies, using a combination of think-pair-shares, in-class polling, in-class worksheets, and random call. Approximately 62% of class activities were at the higher Bloom's taxonomy levels of application, analysis, synthesis, or evaluation (Anderson *et al.*, 2001). In addition to textbook readings and preclass RQs, outside-class assignments included online videos and a weekly online practice exam consisting of three old exam questions (Jackson *et al.*, 2018). Students were also provided with learning objectives and old exam questions to study.

Student demographic information was obtained from the registrar and included binary gender (62% female), grade point average (GPA) at the start of the term, participation in the UW Educational Opportunity Program (EOP; i.e., students identified as economically or educationally disadvantaged; 17%), whether a student was from a race/ethnicity that is traditionally underrepresented in science (underrepresented minority [URM]; i.e., African American, Hispanic, Native American, or Hawaiian-Pacific Islander; 9%), and whether a student's parents had a college degree (i.e., first-generation college students [First Gen]; 14%).

### RQ Treatment Implementation

To investigate whether preclass RQ format impacted student preparedness for class activities, learning on exams, or perceptions of the course, we randomly assigned students, grouped by laboratory section, to take their preclass RQ in either the gamified, adaptive multiple-choice format ( $n = 300$ ) or the traditional, static multiple-choice format ( $n = 276$ ). Each student completed only one format of preclass RQ throughout the entire quarter. Questions for both formats of preclass RQ were selected from the same publisher-provided question bank associated with *Life: The Science of Biology* (Sadava *et al.*, 2017), and both preclass RQ formats were accessed through the LaunchPad portal (Macmillan Learning, 2017a). Preclass RQ questions were designed to be at the Bloom knowledge or comprehension level, as the goal of preclass RQs is to help students gain foundational knowledge about each course topic (Anderson *et al.*, 2001; Crowe *et al.*, 2008). The average weighted Bloom's index for the preclass RQ test bank was  $-0.38$  (Wright *et al.*, 2016). If an assessment has a majority of questions at the knowledge or comprehension levels, it would have a weighted Bloom index of

0.33, while an assessment with a majority of questions at the application or analysis levels would have a weighted Bloom index of 0.66.

Preclass RQs were assigned once or twice a week depending on how much reading was necessary to prepare for class in a given week. Preclass RQs were completed online before coming to class. All preclass RQs were available on Friday and closed on variable days in the following week. Due dates throughout the quarter fell on Monday, Tuesday, Wednesday, and/or Thursday depending on how the reading content aligned with class content for the week. Use of textbooks and notes was allowed while students were taking the online preclass RQs, and there was no time limit.

The adapt-RQ format used in this study was LearningCurve (Macmillan Learning, 2017b). In this preclass RQ treatment, students were required to correctly answer a variable number of questions about the assigned textbook reading in order to reach a minimum of 150 points per assigned textbook section (the gamified aspect). If students answered a question correctly on the first try, they earned 20 points toward their total. If they answered correctly on a subsequent try, they earned fewer points. Students were also given the option to ask for "hints" about the answer to a question. If students chose to use the optional hints and proceeded to answer the question correctly, they received fewer points for an answer. In addition to hints, students were given the option to "refer to the text." Choosing to use this option brought students to specific pages of the electronic textbook in LaunchPad where the answer could be found. Students enlisting the "refer to the text" feature were not penalized with fewer points. If students answered a question incorrectly, they were given more questions pertaining to the specific topic (the adaptive aspect). If students answered a question correctly, they were not given more difficult question; they were just not given extra questions. On average, students answered eight questions to reach each 150 points.

For trad-RQs, students were asked to complete a set number of questions randomly pulled from the same LearningCurve question bank. During the first and second weeks of the quarter, the students in the trad-RQ condition received six questions per assigned textbook section based on a prediction of what the average number of questions answered would be in the gamified, adaptive treatment. Starting with the third week of the quarter and continuing to the end of the quarter, students answered eight questions per assigned textbook section, as this was the average number of questions completed by the adapt-RQ treatment students to reach 150 points.

Both formats of preclass RQ were graded only for completion. Students who were in the gamified, adaptive treatment received full credit if they reached the minimum of 150 points per textbook section. The majority of students did not exceed 15 points beyond the required minimum for a full-credit completion grade. Students who were in the trad-RQ treatment received full credit if they answered all of the questions on the preclass RQ, regardless of correctness.

### Data Collection

**Question 1: Do Gamified, Adaptive RQs Improve Students' Preparedness for Class?** To investigate whether preclass RQ format impacted students' preparedness for class activities, we administered an in-class RQ five times during the quarter. Each

in-class RQ was given as a paper-and-pencil quiz, at the beginning of lecture the day an online preclass RQ was due. We interpret the scores on the in-class RQ as a proxy for students' preparedness to participate in and learn from class activities. Students were informed of the in-class RQ at least 2 days before the quiz. The in-class RQs consisted of five multiple-choice questions on course content pulled from a second test bank associated with *Life: The Science of Biology* (Sadava *et al.*, 2017). The questions did not overlap with those in the preclass RQs but were similar in content and Bloom level (Anderson *et al.*, 2001; Crowe *et al.*, 2008). The in-class RQs were graded for correctness, and the points students earned contributed to their final course grades. To minimize cheating, we randomly administered two versions of the in-class RQs to each section of the course, and questions and answer choices were rearranged between lecture sections. All four versions of the in-class RQ had the same mean and SD (unpublished data).

To investigate whether preclass RQ treatment impacted students' perceived preparedness for class activities, we asked students the following question at the start of each in-class RQ:

How prepared do you feel for this quiz? Circle one number:  
not prepared 1 2 3 4 5 6 7 8 9 10 very prepared

We interpret these scores as a proxy for students' perceived preparedness to participate in and learn from class activities.

**Question 2: Do Gamified, Adaptive RQs Improve Students' Exam Performance?** To investigate whether preclass RQ treatment impacted student performance on course exams, we analyzed total exam points. There were five course exams, given every 2 weeks of a 10-week quarter. Exams consisted of short-answer and multiple-choice questions plus short "explain your choice" questions. Each of the first four exams were worth 100 points (7 questions/exam), while the fifth was worth 200 points (14 questions), for a total of 600 possible exam points. Exam questions in this course are similar to the old exam questions used as practice exam questions in our previous research (Jackson *et al.*, 2018). The average weighted Bloom's index of each exam was 0.605 (Wright *et al.*, 2016). A weighted Bloom score of 0.605 indicates that a large percentage of exam questions in this course were at the higher Bloom levels of application or analysis (Anderson *et al.*, 2001).

**Question 3: Do Gamified, Adaptive RQs Positively Impact Students' Perceptions of a) RQs and b) the Course?** To investigate whether students' perceptions of preclass RQs and the course were impacted by the format of the RQ they took, we administered an online survey during the first and last weeks of the quarter (weeks 1 and 10). The questions for the survey were created by the authors and reviewed for clarity by five education researchers at UW who were not associated with the project. The questions were revised and reviewed by 10 undergraduates using a think-aloud approach (Ericsson and Simon, 1993). The questions were then revised for a final time.

The week 10 survey included two types of questions to investigate students' perceptions of RQs. The first questions asked students to evaluate the value of available course resources.

How valuable are each of the following resources for your mastery of the course material in [course name]? No value 1, 2, 3, 4, 5, 6, 7, 8, 9, very valuable 10

Students responded to this question for each of five resources: online practice exams, in-class discussions (including polling/clickers), random call during class, preclass RQs, and the textbook. This question addressed whether the adapt-RQ treatment influenced students to preferentially value preclass RQs or the textbook.

We also administered these resource value questions to students in week 1, with slightly modified wording: "How valuable do you think each of the following resources will be for your mastery of the course material in [course name]?" As this course is the third in a three-quarter sequence and course resources are similar, we thought it likely students would already have developed opinions about how much they would value a given resource, and we wanted to control for this in our analyses.

The second type of question on the week 10 survey asked students to compare the preclass RQs in this course to their experiences with preclass RQs in their previous introductory biology courses, which were traditional, static RQs.

Reading quizzes in [this course] are \_\_\_\_\_ reading quizzes in [previous intro bio courses]. Better than, slightly better than, the same as, slightly worse than, worse than

The final question on the week 10 survey was used to investigate the impact of preclass RQ format on how much students' enjoyed the course.

How much did you enjoy [course name] (the course, not the exams)? 1 not at all, 2, 3, 4, 5, 6, 7, 8, 9, 10 a lot

## Modeling Methods

To determine whether preclass RQ treatment had an impact on our response variables, we used multilevel models for analyses of all data (Gelman and Hill, 2007). Linear regression was used to model mean in-class RQ score, mean preparedness score, and total exam points. Ordinal regression was used to model resource value, RQ comparison, and course enjoyment. For each response variable, analysis began with the most complex model, containing every variable of interest plus random effects. Fixed effects for all research questions included all student characteristics (i.e., GPA, gender, EOP status, URM status, First Gen status) and preclass RQ treatment. To determine whether treatment had a differential impact on students with different characteristics, the most complex model also included interactions between preclass RQ treatment and student characteristics. Lecture section was included as a random effect in the most complex model for all linear regressions and as a fixed effect in all ordinal regressions (due to a limitation in the analysis package only allowing random effects with three or more levels). As students were randomly assigned to treatment by laboratory sections, lab section was included as a random effect in the most complex model. Additional factors were added to some models (detailed below).

We used backward selection, guided by Akaike's information criteria (AIC), to determine the best-fit model for the analysis



TABLE 1. Models and  $\Delta AIC^a$  for each best-fit analysis model

Research question	Model <sup>b</sup>	R <sup>2</sup>	$\Delta AIC$
Question 1	Mean in-class RQ score ~ Gender + Preparedness + GPA	0.258	166
Question 2	Total exam points ~ Gender + Number of preclass RQ completed + Mean in-class RQ score + GPA + (1 Lecture section)	0.573	498
Question 3a	Compare RQ ~ Treatment + Preparedness + Treatment*Preparedness	0.025	6
Question 3b	Post resource value ~ Pre resource value + Resource type + Resource type*Pre resource value + Treatment + GPA + Treatment*GPA + Resource type*GPA	0.301	793
	Pre resource value + Resource type + Resource type*Pre resource value	0.290	
	Treatment + GPA + Treatment*GPA + Resource type*GPA	0.011	
Question 3b	Enjoying course ~ Preparedness + Total exam points	0.057	26

<sup>a</sup> $\Delta AIC$  is the difference between the best-fit model and the null model. This difference is a measure of the relative goodness of fit the best-fit model when compared with the null model. The null model is a model that only contains the intercepts and any retained random effects. The null model is similar to the null hypothesis. The null model would be the best-fit model if the proposed factors had no impact on the response variable.

<sup>b</sup>The only models that retained treatment were the compare RQ and post value models. The compare RQ model only contains an interaction with treatment, so we did not decompose the  $R^2$  to partition the variance among factors. For the resource value model, we decomposed  $R^2$  to partition the variance among the treatment interaction and other factors.

(Akaike, 1973). Parameters were sequentially removed, starting with random effects, then moving on to interactions with the highest  $p$  value, and then main fixed effects. AIC values were recorded after each sequential model adjustment. The model with the smallest AIC score was determined to be the best fit to explain the data (see Table 1 for selected models). Models with a difference in AIC score of  $\pm 2$  were considered to be equivalent, and in those cases, to satisfy guidelines of parsimony, the model with the fewest parameters was selected (Burnham and Anderson, 2002). Analyses were carried out using R (R Core Team, 2019) with the lme4 (Bates *et al.*, 2015) and ordinal packages (Christensen, 2019).  $R^2$  values for all models were generated using the sjplot package (Lüdtke, 2019).

In addition to determining whether preclass RQ treatment had an impact on our response variables, we were interested in how strong the impact was (Aguinis *et al.* 2010). To investigate the strength of the impact, we used  $R^2$  as a measure of effect size and decomposed the  $R^2$  to partition the variance among factors retained in each model. Decomposing the  $R^2$  allowed us to estimate how much of the variation explained by a model could be attributed to specific factors. We used Cohen's scale for categorizing the magnitudes of effect sizes calculated by  $R^2$ . An  $R^2$  less than 0.09 is a small effect size, between 0.09 and 0.25 is a medium effect size, and greater than 0.25 is a large effect size (Cohen, 1988). For linear models without random effects, we decomposed the  $R^2$  contributions by averaging over orderings among regressors using the R package relaimpo (Grömping, 2006). This package cannot be used with models that include random effects or with ordinal regression. When a random effect was retained in the best-fit model, we estimated the fixed effects contributions to the marginal  $R^2$  by calculating the change in marginal  $R^2$  when each was removed from the model. Similarly, for the ordinal regressions, we estimated the contributions of each factor or group of factors to Nagelkerke's  $R^2$  by calculating the change in Nagelkerke's  $R^2$  when each was removed from the model.

**Question 1: Do Gamified, Adaptive RQs Improve Students' Preparedness for Class?** Using multilevel linear regression, we modeled student's mean in-class RQ score to investigate whether preclass RQ treatment impacted student's preparedness for class activities. In-class RQ score served as a proxy for

preparedness for class activities. By using the mean, we were able to retain those students who did not take all five quizzes in our data set. In addition to treatment, student characteristics, and interactions, we included the mean of the student's self-reported preparedness for the in-class RQs and the number of preclass RQ assignments the student completed over the quarter as fixed effects. The most complex model was

Mean in-class RQ score ~ Treatment + Gender + URM + First Gen + EOP + GPA + Preparedness + Number of preclass RQs completed + Treatment\*Gender + Treatment\*URM + Treatment\*First Gen + Treatment\*EOP + Treatment\*GPA + Treatment\*Preparedness + (1|Lab section) + (1|Lecture section)

Using multilevel linear regression, we modeled the mean self-reported preparedness for in-class RQs to investigate whether preclass RQ treatment impacted students' perception of their preparedness for class activities. Perceived preparedness for in-class RQ served as a proxy for perceived preparedness for class activities. In addition to treatment, student characteristics, and interactions, we included the number of preclass RQ assignments the student completed over the quarter as a fixed effect. The most complex model was

Mean preparedness ~ Treatment + Gender + URM + First Gen + EOP + GPA + Number of preclass RQs completed + Treatment\*Gender + Treatment\*URM + Treatment\*First Gen + Treatment\*EOP + Treatment\*GPA + (1|Lab section) + (1|Lecture section)

**Question 2: Do Gamified, Adaptive RQs Improve Students' Exam Performance?** To investigate whether preclass RQ treatment impacted students' performance on course exams, we modeled total exam points as all students took all exams. In addition to preclass RQ treatment, student characteristics, and interactions, we included the student's mean in-class RQ score

and its interaction with treatment and the number of preclass RQ assignments the student completed over the quarter as fixed effects. The most complex model was

Total exam points ~ Treatment + Gender + URM  
 + First Gen + EOP + GPA + In-class RQ score  
 + Number of preclass RQs completed + Treatment\*Gender  
 + Treatment\*URM + Treatment\*First Gen + Treatment\*EOP  
 + Treatment\*GPA + Treatment\*In-class RQ score + (1|Lab section)  
 + (1|Lecture section)

**Question 3: Do Gamified, Adaptive RQs Positively Impact Students' Perceptions of a) RQs and b) the Course?** To investigate whether preclass RQ treatment impacted students' perceptions of preclass RQs, we used ordinal regression to model their RQ comparison and how much they valued course resources. When modeling preclass RQ comparison (compare RQ model), in addition to treatment, student characteristics, and interactions, we included mean preparedness and in-class RQ score and their interaction with treatment as fixed effects. The most complex model was

Compare RQ ~ Treatment + Gender + URM + First Gen + EOP  
 + GPA + Preparedness + In-class RQ score + Treatment\*Gender  
 + Treatment\*URM + Treatment\*First Gen + Treatment\*EOP  
 + Treatment\*GPA + Treatment\*Preparedness  
 + Treatment\*In-class RQ score + Lecture section + (1|Lab section)

When modeling how much students' valued course resources (resource value model), in addition to treatment, we included resource type and its interactions as fixed effects. If one of the interactions was retained in the model, we would conclude students in different groups (e.g., gender or treatment) valued different resources differently. We also included students' values in week 1 ("Pre resource value") as a fixed effect to control for students' incoming valuations of the various course resources. The most complex model was

Post resource value ~ Treatment + Gender + URM + GPA  
 + First Gen + EOP + Resource type + Pre resource value  
 + Resource type\*Pre resource value + Treatment\*Gender  
 + Treatment\*URM + Treatment\*GPA + Treatment\*Resource type  
 + Treatment\*First Gen + Treatment\*EOP + Resource type\*Gender  
 + Resource type\*URM + Resource type\*GPA + Resource type  
 \*First Gen + Resource type\*EOP + Treatment\*Gender  
 \*Resource type + Treatment\*URM\*Resource type + Treatment\*GPA  
 \*Resource type + Treatment\*Resource type\*First Gen + Treatment  
 \*Resource type\*EOP + Lecture section + (1|Lab section)

We also used ordinal regression to investigate whether preclass RQ treatment impacted students' enjoyment of the course. In addition to treatment, student characteristics, and interactions, we included mean preparedness and in-class RQ score and their interaction with treatment as fixed effects. We also included average exam points in the enjoyment of the course model. The most complex model was

Enjoying course ~ Treatment + Gender + URM + First Gen + EOP  
 + Average exam points + Preparedness + In-class RQ score  
 + Treatment\*Gender + Treatment\*URM + Treatment\*First Gen  
 + Treatment\*EOP + Treatment\*Average exam points  
 + Treatment\*Preparedness + Treatment\*In-class RQ score  
 + Lecture section + (1|Lab section)

This research was approved by the Human Subjects Division of the University of Washington Application STUDY00001576. For this experiment, access to the electronic textbook, LaunchPad, and LearningCurve were provided to the students for free by Macmillan Learning.

## RESULTS

### Question 1: Do Gamified, Adaptive RQs Improve Students' Preparedness for Class?

Mean score on the in-class RQ, across all treatments, was  $3.6 \pm 0.7$  (out of 5). The model that best explained mean in-class RQ score included GPA, gender, and mean preparedness score for in-class RQs (Table 1). Therefore, there was no differential impact of the preclass RQ treatment on mean in-class RQ score (Figure 1). See the Supplemental Material for parameter estimates, marginal effects plots, and percent of variation explained by GPA, gender, and mean preparedness score.

The model that best explained mean preparedness for in-class RQ was the intercept-only model, which contained no fixed effects but did retain lab section included as a random effect (see the Supplemental Material). Therefore, there was no differential impact of the preclass RQ treatment on how prepared students felt for in-class RQs.

### Question 2: Do Gamified, Adaptive RQs Improve Students' Exam Performance?

The model that best explained student performance on course exams included gender, the number of preclass RQs completed, mean in-class RQ score, and GPA as fixed effects and lecture section as a random effect (Table 1). Therefore, there was no differential impact of the preclass RQ treatment on students' exam performance (Figure 2). See the Supplemental Material for parameter estimates, marginal effects plots, and percent of variation explained by GPA, number of preclass RQs completed, and mean in-class RQ score.

### Question 3: Do Gamified, Adaptive RQs Positively Impact Students' Perceptions of a) RQs and b) the Course?

The model that best explained students' comparisons of RQs in this course to previous introductory biology courses (compare RQ model) contained preclass RQ treatment, mean

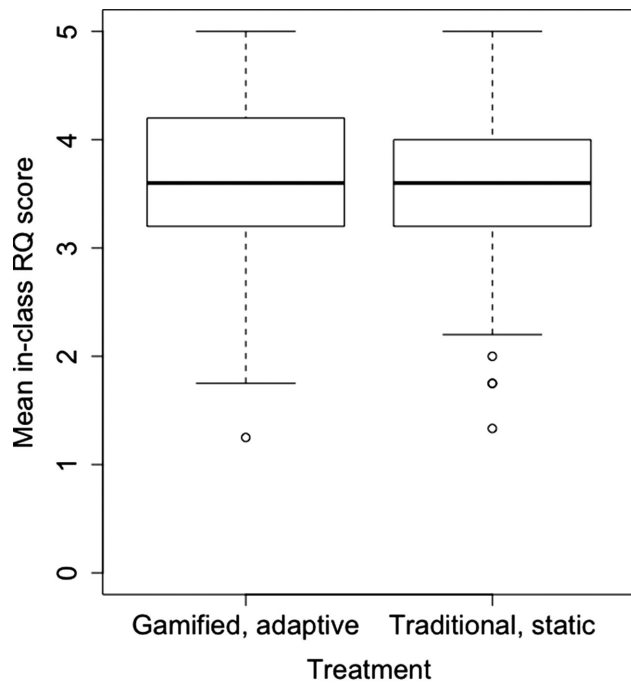


FIGURE 1. Box-and-whisker plot of mean in-class RQ score for students in the gamified, adaptive and traditional, static RQ treatments. After controlling for gender, mean preparedness, and GPA, students perform equally well in both treatment groups. See Table 1 for model and  $R^2$  and the Supplemental Material for tables of parameter estimates, confidence intervals and  $p$  values, and plots of marginal effects for the best-fit models for each response variable.

preparedness for in-class RQs, and their interaction. This model explained only 2.5% of the variation, a small effect size (Table 1). Among students with higher self-reported preparedness for in-class RQs, students in the adapt-RQ treatment viewed the preclass RQ treatment for this course more favorably compared with their previous courses. Students in both preclass RQ treatment groups who self-reported they had a low preparedness for in-class RQs rated the preclass RQ in this and previous courses more similarly (see the Supplemental Material for parameter estimates and marginal effects plots).

Students were also asked to compare the value of five course resources offered in the current class (resource value model). If the interaction between resource type and treatment was retained in the model, we could conclude students in different preclass RQ treatments valued different resources differently. Students' initial value rating taken from week 1 of the quarter for course resources was retained in the best model and explained 29% of the variation in value (the total model explained 30.1%; Table 1; see the Supplemental Material for parameter estimates and marginal effects plots). That is, preconceived ideas of resource value explain most of the variation in students' end of course ratings, regardless of which preclass RQ treatment they received. The other factors retained, which explain a total of 1.1% of the variation (a small effect size), include preclass RQ treatment, GPA, and their interaction and the interaction between GPA and resource type.

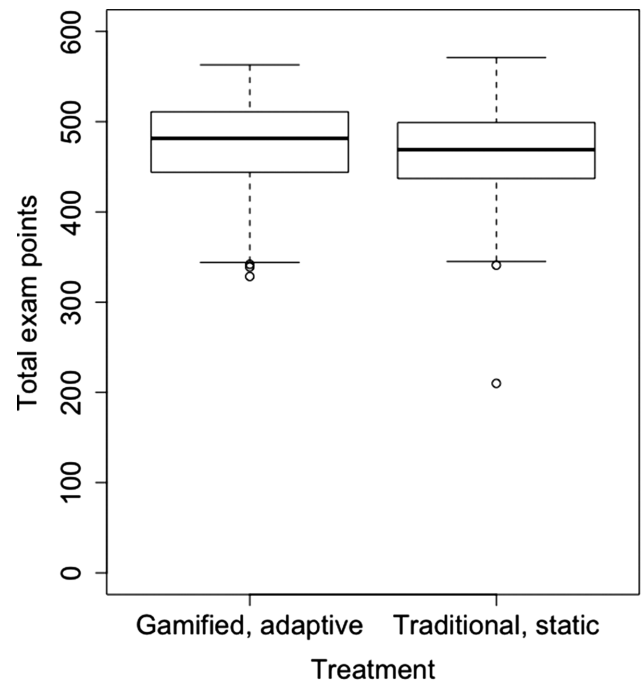


FIGURE 2. Box-and-whisker plot of total exam points for students in the gamified, adaptive and traditional, static RQ treatments. After controlling for GPA, mean in-class RQ score, number of RQs completed, and gender, students perform equally well in both treatment groups. See Table 1 for model and  $R^2$  and the Supplemental Material for tables of parameter estimates, confidence intervals and  $p$  values, and plots of marginal effects for the best-fit models for each response variable.

The model that best explained students' enjoyment of the course contained mean preparedness for in-class RQs and total exam points (Table 1). Therefore, there was no differential impact of the preclass RQ treatment on students' enjoyment of the course. See the Supplemental Material for parameter estimates, marginal effects plots, and percent of variation explained by in-class RQs and total exam points

## DISCUSSION

### Question 1: Do Gamified, Adaptive RQs Improve Students' Preparedness for Class?

Using adapt-RQs had no impact on students' actual or perceived preparedness for class activities as measured by performance and perceived preparedness for in-class RQs. That is, students who did the trad-RQs were just as prepared for class activities as those using the adapt-RQs. There are several possible explanations for this result. It could be that students in this course have already developed the habit of reading the textbook from previous courses in the introductory biology series. Therefore, the format of the preclass RQs provided no additional incentive to complete the assigned reading. It could also be that students in this course are highly capable of learning from reading assignments and do not gain any benefit from the focused hints offered by the gamified, adaptive format. As the students in this course may be more adept at gleaned information from the text, it is also possible that they got many questions right on the first try and therefore did not

use the additional practice the gamified adaptive format provided. The mean critical reading Scholastic Aptitude Test (SAT) score for students in this course was 621, compared with the national average of 497 from students who graduated high school in the same year, 2015 (College Board, 2017). Given the high critical reading SAT scores of the students in this study, our findings align with the results in the study by Griff and Matter (2013). They saw no impact of gamified adaptive preparatory assignments in 4-year colleges but did find a positive impact of gamified, adaptive preparatory assignments in students at 2-year colleges, which generally have open enrollment that allows less prepared students to enter college (Center for Community College Student Engagement, 2016).

### Question 2: Do Gamified, Adaptive RQs Improve Students' Exam Performance?

Student performance on exams was unaffected by preclass RQ format. We reasoned that gamified, adaptive preclass RQs could improve student preparedness to learn from in-class activities, which were at higher Bloom levels. We hypothesized that being more prepared to learn from this in-class practice at higher Bloom levels would in turn prepare students to perform better on higher Bloom-level exam questions. However, our results indicate that preclass RQ format did not impact student preparedness to learn. Therefore, we conclude that both treatment groups were equally prepared to learn from class activities, and we should not expect a difference on exam performance. Furthermore, both the preclass and in-class RQs consisted of questions at the knowledge or comprehension Bloom levels, whereas course exams consisted predominantly of questions at the application or analysis levels. Previous research has shown that homework practice at high and lower Bloom levels improves performance on low Bloom-level exams; however, the corollary is not true: Homework practice at low Bloom levels does not prepare for success at higher Bloom levels (Jensen *et al.*, 2014).

### Question 3: Do Gamified, Adaptive RQs Positively Impact Students' Perceptions of a) RQs and b) the Course?

Treatment was retained in both models investigating students' perceptions of preclass RQs (i.e., the compare RQ model and the resource value model). This could be interpreted as signifying that the preclass RQ treatment is an important factor in influencing students' perceptions of preclass RQs. However, when we examined the effect size measured by  $R^2$  for the compare RQ model, the whole model explained only 2.6% of the variation, a small effect size. Additionally, when we decomposed the  $R^2$  for the resource value model to remove the variation attributed to the value students had for each resource at the beginning of the course (week 1 survey), we found that preclass RQ treatment explained less than 1% of the variation. This small effect size related to the increase in positive perception of preclass RQs due to the gamified, adaptive format is probably not instructionally meaningful. Given these results, we recommend that researchers consider examining the  $R^2$  more closely when evaluating the impact of interventions on student learning, decomposing the  $R^2$  when possible.

While probably not instructionally meaningful, this slight increase in positive perception of preclass RQs could be due to students' appreciation of the gamified or adaptive nature of preclass RQs or the easy access to aligned textbook readings. It could also be due to the novelty of the gamified, adaptive assignments. If we instituted adapt-RQs in the first quarter of the introductory series, we might not find this positive impact by the third quarter.

The slightly higher value students put on the adapt-RQs did not impact students' overall course enjoyment. Many things go into whether or not a student enjoys a course. In fact, our model for predicting how much a student enjoyed a course only explained 5.7% of the variation; students with higher exam scores and who felt more prepared enjoyed the class more than others. It may have been too ambitious to think that merely changing the format of the preclass RQ assignment could impact course enjoyment when there are so many other contributing factors, such as the use of active-learning strategies (Connell *et al.*, 2016; Corkin *et al.*, 2017) or whether the grade a student is achieving matches or exceeds his or her expectations (Remedios *et al.*, 2000).

### Limitations

While we see no impact of preclass RQ format on student performance, we cannot say whether the preclass RQ assignments, in general, had an impact on learning. This is due to the lack of a no-preclass RQ control. As many authors have shown that adding preclass RQs to a course improves student course performance (Johnson and Kiviniemi, 2009; Moravec *et al.*, 2010; Freeman *et al.*, 2011; Pape-Lindstrom *et al.*, 2018), we felt that having a no-preclass RQ control would be unethical.

We compared the effectiveness of gamified, adaptive-learning preclass RQ to more traditional RQ at a more selective R1 university, in a high-structure active-learning class for biology majors, at the end of the introductory series (Carnegie Classification of Institutions of Higher Education, n.d.). Our results may not be generalizable to other contexts. The extra support provided by gamified, adaptive-learning software may have an impact on students who are less prepared for college biology, such as students enrolled in the first term of introductory biology, nonmajors courses, or less selective institutions. The software may also have a greater impact on students in a more traditional lecture course who have less opportunity to practice in class. Additionally, we only used one version of gamified, adaptive software (i.e., LearningCurve); other versions may impact student learning or affect differently.

### CONCLUSIONS

Our initial hypothesis regarding the effectiveness of the gamified, adaptive assignments was that students would engage in their preclass RQs more and come to class more prepared and ready to learn from class activities. This increased readiness could lead to greater learning in class and translate to higher exam scores. The data do not support this hypothesis for our population of students and the software we used. Given our findings and that the gamified, adaptive software has a financial cost associated with the required subscription fee, the old adage applies: Try it before you buy it.



## ACKNOWLEDGMENTS

We thank all the students who participated in the study and Kyle Loucks, Melissa Mallen, Edith Serna, Dylan Moorleghan, and Naresh Oli from UW and Elaine Palucki of Macmillan Learning for logistical assistance in performing the experiment. We also thank the UW BERG lab, in particular Mary Pat Wenderoth, for helpful discussion and comments on previous drafts of this article.

## REFERENCES

- Aguinis, H., Werner, S., Lanza Abbott, J., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13*(3), 515–539. doi: 10.1177/1094428109333339
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N., & Csaki, F. (Eds.), *Proceedings of the 2nd international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Armbruster, P., Patel, M., Johnson, E., & Weiss, M. (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE—Life Sciences Education, 8*(3), 203–213. doi: 10.1187/cbe.09-03-0025
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi: 10.18637/jss.v067.i01
- Becker-Blease, K. A., & Bostwick, K. C. P. (2016). Adaptive quizzing in introductory psychology: Evidence of limited effectiveness. *Scholarship of Teaching and Learning in Psychology, 2*(1), 75–86. doi: 10.1037/stl0000056
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer Science.
- Carnegie Classification of Institutions of Higher Education. (n.d.). *About Carnegie Classification*. Retrieved May 8, 2019, from <http://carnegieclassifications.iu.edu/>
- Center for Community College Student Engagement. (2016). *Expectations meet reality: The underprepared student and community colleges*. Austin, TX: University of Texas at Austin, College of Education, Department of Educational Administration, Program in Higher Education Leadership.
- Cheong, C., Cheong, F., & Filippou, J. (2013). Quick Quiz: A gamified approach for enhancing learning. *PACIS 2013 Proceedings, 206*. <http://aisel.aisnet.org/pacis2013/206>
- Christensen, R. H. B. (2019). *Ordinal—regression models for ordinal data (R package version 2019.4-25)*. Retrieved May 1, 2019, from [www.cran.r-project.org/package=ordinal/](http://www.cran.r-project.org/package=ordinal/)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- College Board. (2017, January 20). *Class of 2016 SAT Results—2016 SAT Suite Program Results*. Retrieved May 7, 2019, from <https://reports.collegeboard.org/archive/sat-suite-program-results/2016/class-of-2016-results>
- Connell, G. L., Donovan, D. A., & Chambers, T. G. (2016). Increasing the use of student-centered pedagogies from moderate to high improves student learning and attitudes about biology. *CBE—Life Sciences Education, 15*(1), ar3. doi: 10.1187/cbe.15-03-0062
- Corkin, D. M., Horn, C., & Pattison, D. (2017). The effects of an active learning intervention in biology on college students' classroom motivational climate perceptions, motivation, and achievement. *Educational Psychology, 37*(9), 1106–1124. doi: 10.1080/01443410.2017.1324128
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education, 7*(4), 368–381. doi: 10.1187/cbe.08-05-0024
- Dolan, E. L. (2015). Biology Education Research 2.0. *CBE—Life Sciences Education, 14*(4), ed1. doi: 10.1187/cbe.15-11-0229
- Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE—Life Sciences Education, 13*(3), 453–468. doi: 10.1187/cbe.14-03-0050
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.
- 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. doi: 10.1073/pnas.1319030111
- Freeman, S., Haak, D., & Wenderoth, M. P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education, 10*(2), 175–186. doi: 10.1187/cbe.10-08-0105
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, England: Cambridge University Press.
- Gordon, N., Brayshaw, M., & Grey, S. (2013). Maximising gain for minimal pain: Utilising natural game mechanics. *Innovation in Teaching and Learning in Information and Computer Sciences, 12*(1), 27–38. doi: 10.11120/ital.2013.00004
- Griff, E. R., & Matter, S. F. (2013). Evaluation of an adaptive online learning system. *British Journal of Educational Technology, 44*(1), 170–176. doi: 10.1111/j.1467-8535.2012.01300.x
- Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software, 17*, 1–27. doi: 10.18637/jss.v017.i01
- Gurung, R. A. R. (2015). Three investigations of the utility of textbook technology supplements. *Psychology Learning & Teaching, 14*(1), 26–35. doi: 10.1177/1475725714565288
- Heiner, C. E., Banet, A. I., & Wieman, C. (2014). Preparing students for class: How to get 80% of students reading the textbook before class. *American Journal of Physics, 82*(10), 989–996. doi: 10.1119/1.4895008
- Jackson, M. A., Tran, A., Wenderoth, M. P., & Doherty, J. H. (2018). Peer vs. self-grading of practice exams: Which is better? *CBE—Life Sciences Education, 17*(3), es44. doi: 10.1187/cbe.18-04-0052
- James, L. A. (2012). *Evaluation of an adaptive learning technology as a predictor of student performance in undergraduate biology* (MS thesis). Boone, NC: Appalachian State University.
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test...or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review, 26*(2), 307–329. doi: 10.1007/s10648-013-9248-9
- Johnson, B. C., & Kiviniemi, M. T. (2009). The effect of online chapter quizzes on exam performance in an undergraduate social psychology course. *Teaching of Psychology, 36*(1), 33–37. doi: 10.1080/00986280802528972
- Lüdecke, D. (2019). *sjPlot: Data visualization for statistics in social science (R package version 2.6.3)*. Retrieved January 28, 2019, from <https://CRAN.R-project.org/package=sjPlot>
- Macmillan Learning. (2017a). LaunchPad [Material accompanying *Life: The science of biology*]. Retrieved March 27, 2017, from [www.macmillanlearning.com/college/us/digital/launchpad/](http://www.macmillanlearning.com/college/us/digital/launchpad/)
- Macmillan Learning. (2017b). LearningCurve [Material accompanying *Life: The science of biology*]. Retrieved March 27, 2017, from [www.macmillanlearning.com/college/us/digital/achieve-read-and-practice/research-and-metrics](http://www.macmillanlearning.com/college/us/digital/achieve-read-and-practice/research-and-metrics)
- Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before lecture: A strategy that improves learning outcomes in a large introductory biology class. *CBE—Life Sciences Education, 9*(4), 473–481. doi: 10.1187/cbe.10-04-0063
- O'Flaherty, J., & Phillips, C. (2015). The use of flipped classrooms in higher education: A scoping review. *The Internet and Higher Education, 25*, 85–95. doi: 10.1016/j.iheduc.2015.02.002
- Pape-Lindstrom, P., Eddy, S., Freeman, S., & Schinske, Jeff (2018). Reading quizzes improve exam scores for community college students. *CBE—Life Sciences Education, 17*(2), ar21. doi: 10.1187/cbe.17-08-0160
- R Core Team. (2019). *R: A language and environment for statistical computing (Version 3.6.0)*. Retrieved May 1, 2019, from [www.R-project.org](http://www.R-project.org)
- Remedios, R., Lieberman, D. A., & Benton, T. G. (2000). The effects of grades on course enjoyment: Did you get the grade you wanted? *British*

- Journal of Educational Psychology*, 70(3), 353–368. doi: 10.1348/000709900158173
- Richards-Babb, M., Curtis, R., Ratcliff, B., Roy, A., & Mikalik, T. (2018). General chemistry student attitudes and success with use of online homework: Traditional-responsive versus adaptive-responsive. *Journal of Chemical Education*, 95(5), 691–699. doi: 10.1021/acs.jchemed.7b00829
- Sadava, D. E., Hillis, D. M., Heller, H. C., & Hacker, S. D. (2017). *Life: The science of biology* (11th ed.). Sunderland, MA: Sinauer Associates, Macmillan Learning Curriculum Solutions.
- Trujillo, G., & Tanner, K. D. (2014). Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE Life Sciences Education*, 13(1), 6–15. doi: 10.1187/cbe.13-12-0241
- Wright, C. D., Eddy, S. L., Wenderoth, M. P., Abshire, E., Blankenbiller, M., & Brownell, S. E. (2016). Cognitive difficulty and format of exams predicts gender and socioeconomic gaps in exam performance of students in introductory biology courses. *CBE—Life Sciences Education*, 15(2), ar23. doi: 10.1187/cbe.15-12-0246
- Zumalt, C. J., & Williamson, V. M. (2016). Does the arrangement of embedded text versus linked text in homework systems make a difference in students impressions, attitudes, and perceived learning? *Journal of Science Education and Technology*, 25(5), 704–714. doi: 10.1007/s10956-016-9625-5