EDUCATION BRIEF

Student Learning with Generated and Answered Peer-written Questions

Shelby L. Hudson, PharmD, MS,a Michael B. Jarstfer, PhD,a Adam M. Persky, PhDa,b

a The University of North Carolina at Chapel Hill, Eshelman School of Pharmacy, Chapel Hill, North Carolina
b Associate Editor, American Journal of Pharmaceutical Education, Alexandria, Virginia

Submitted February 1, 2017; accepted August 15, 2017; published March 2018.

Objective. To investigate the degree to which student-generated questions or answering student-generated multiple-choice questions predicts course performance in medicinal chemistry.

Methods. Students enrolled in Medicinal Chemistry III over a 3-year period were asked to create at least one question per exam period using PeerWise; within the software, they were also asked to answer and rate one peer question per class session. Students’ total reputation scores and its components (question authoring, answering, and rating) and total answer scores (correctness of answers submitted indicating agreement with the author’s chosen answer) were analyzed relative to final course grades.

Results. Students at the non-satellite campus and those who generated more highly rated questions performed better overall in the course accounting for 12% of the variability in course grades. The most notable differences were between the top third and bottom third performing students within the course. The number of questions answered by students was not a significant predictor of course performance.

Conclusion. Student generation of more highly rated questions (referred to as more thoughtful in nature by the software program) is predictive of course performance but it only explained a small variability in course grades. The correctness of answers submitted, however, did not relate to student performance.

Keywords: multiple choice questions, course performance, question generation, peer review, medicinal chemistry

INTRODUCTION

During the learning process, students retain knowledge and skills longer when they receive feedback and retrieve information from memory.1-4 Normally, the instructor provides these opportunities in the learning environment. As such, students often rely on external sources (eg, instructor homework or practice tests) for these opportunities, and as a result, develop inabilities to self-assess their knowledge. Accordingly, we used a technology that allows students to generate questions and make the questions available to all students in the course to answer and provide feedback on the question quality. The goal of this study is to assess if student performance in the course is influenced by the extent they engage in this technology, ie, does answering more questions or generating more questions improve their grade?

PeerWise (Auckland, New Zealand) allows learners to generate assessment questions, have these questions peer-evaluated for quality and correctness, and use these questions to self-test. Several studies have used PeerWise within their courses, mostly at the undergraduate level, and a few within pharmacy and medicine. These latter studies have found increases in student satisfaction and course performance.5-7 Unfortunately, student perceptions of efficacy do not necessarily correlate with learning outcomes.5,8 This study differs than previous studies because it was conducted in a foundational science class over several years where students could vary their engagement in PeerWise in generating and answering questions. This study’s objectives are to investigate to what extent learner engagement with PeerWise relates to his/her course performance and what part of the process may be linked to its benefit. These findings can help in course design to provide guidance and evidence of what strategies may facilitate student learning.

METHODS

Study participants consisted of 487 student pharmacists enrolled in the PharmD program over a three-year period at a large, public university in the southeastern United States. Students occupied two campuses that were synchronously video-conferenced and were enrolled in Medicinal Chemistry III. Upon admission to the program, the average age of students was 23 years (range 19-50),
and approximately 65% were female. The mean grade point average upon admission was 3.5 (out of 4.0), the mean Pharmacy College Admission Test (PCAT) score was approximately 87%, and approximately 80% of students over the three-year period had a prior degree.

Medicinal Chemistry III was a required 2.5 credit course during the fall semester of the third year. The course format was lecture-based with some active learning. As a course requirement, students authored at least one question per exam period (three questions over the course of the semester), answered at least one student-generated question per class session, and evaluated at least one peer written question per each class session. Credit was given based on completion of activities. The course teaching assistants and instructor provided some oversight of the questions to reduce inaccurate answers and reduce the number of superficial questions. If students wrote superficial questions or wrote questions that reflected poor effort, they would receive only half credit for the assignment. There was no limit put on the number of questions written, answered or rated, only a minimum requirement of one question answered and evaluated per class session and one written per each of three exam periods during the semester. When selecting questions to answer or review, students could view the overall rating and difficulty rating of questions, and the question’s overall reputation score; they could not see who authored the question, however. Students were not asked to do anything outside of the normal progression of the course for participation in this study.

From PeerWise, four variables were included in the analysis. These variables are described in Table 1 and are generated by the software through proprietary methods. The primary outcome measure was overall course score, which included three examinations (95% of the grade) and PeerWise contributions (5% of the grade). Campus location (main vs satellite) was included as a variable to investigate the impact on the satellite campus on course performance.

A multiple linear regression was performed on the four PeerWise components and campus using the final course score as the dependent variable; all variables were entered in the model simultaneously (SPSS version 24, IBM Corp, Armonk, NY). In addition, all four PeerWise components were compared relative to the final course score tertile using an ANOVA with Tukey post-hoc with significance set at \( p < .05 \). For significant findings, Cohen’s \( d \) was calculated as a measure of effect size.

This study was deemed exempt by the university’s Institutional Review Board.

**RESULTS**

The regression analysis identified two significant predictors of student performance in the course: question authoring component and campus location (Table 2). These factors explained 12% of the variability in grades. There was no statistical difference in any PeerWise components between campuses except that the satellite campus had a lower course average (89 vs 85, \( p < .001, \ d = .67 \)).

PeerWise data was analyzed per final course grade tertiles. The highest performing students had higher question authoring scores compared to the middle tertile (\( p = .012, \ d = .32 \)) and lowest tertile (\( p < .001, \ d = .58 \)).

Table 1. Explanation of PeerWise Components

| PeerWise Component           | Part of Total Reputation Score? | Meaning                                                                                                                                 |
|------------------------------|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| Total Reputation Score       | —                               | Increases with regular contributions, authoring questions that their peers rate highly, answering questions correctly before their peers, and rating questions as they are subsequently rated by their peers. |
| Question Authoring Score     | Yes                             | Increases with earlier submission of more thoughtful contributions as determined by student peer ratings of questions.                   |
| Student Answer Score         | Yes                             | Increases with increased quantity of answers and when actions of other students agree with that student’s earlier choices.               |
| Question Rating Score        | Yes                             | Is a function of the number of ratings submitted by a student and their agreement with other students’ earlier contributions. This score favors early participation and to rate the question fairly. |
| Total Answer Score           | No                              | Increases every time a student submits an answer that is “correct” (in the sense that it agrees with the question author’s suggested answer). |
It is unclear why answering more correctly did not impact performance. One possible explanation is the discrepancy between the level of questions generated by students compared to those generated by the instructors for the examination. For example, if student-generated questions were lower levels of Bloom’s Taxonomy and the examinations were higher level of Bloom’s Taxonomy, then practicing the lower level questions may not transfer to answering higher level questions.\textsuperscript{9} A second reason why answering questions did not predict course grade was a lack of effort. For retrieval to be effective as a memory aid, it has to be challenging and successful.\textsuperscript{1} In other words, answering questions is less fruitful if the success of answering was not associated with challenge or difficulty. The lack of relationship in question answering and course performance also has been noted in the use of audience response systems (ie, clickers).\textsuperscript{10,11} Like PeerWise, audience response systems can provide students opportunities to retrieve information. Both the findings from this study and the audience response data is counter to findings of the testing effect and the impact of instructor-generated question within class.\textsuperscript{12}

This study relied on software-generated metrics to help determine which aspects of the exercise improved student learning. This is a limitation as the calculation of these metrics is unknown and their meaning is only derived from the software developer. The advantage of these metrics is that they triangulate various perspectives of the experience into one metric which, hopefully, reflects the complexities of the process. It also is unclear whether stronger students would naturally participate more in the question writing process, or if by participating in question writing process helped students become stronger on the topic. We did not account for prior medicinal chemistry course achievement but prior success may determine a larger variability in the grades than just the PeerWise intervention. Finally, students could see the ratings for each question. This can have the advantage that students can select better questions to practice but it also may bias their own rating of the question.\textsuperscript{13,14} Students, however, do not see the identity or frequency of choices selected by their peers who have answered their selected question prior to them.

**CONCLUSION**

Incorporating question authoring and answering activities within a foundational science course can potentially have a positive association on students’ course performance, particularly if thoughtfulness and timelyness of question authoring are encouraged. The use of PeerWise has the advantage of the software being freely available and without classroom size limitations while

---

**DISCUSSION**

The regression and ANOVA results of this study suggest that students with higher final course grades created more highly rated questions (ie, more thoughtful questions). There was some indication that students who performed better overall did answer more questions correctly compared to students who performed the least well within the course. Overall, these findings support the idea that the degree of question generation supported student performance but question answering did not or did to a lesser extent. In other words, generating questions and elaborating on current knowledge was more influential than retrieving information for most students. Overall, fewer differences were found between the upper and middle tertile than between the upper and lower tertiles or the middle and lower tertiles, suggesting that students who performed the worst in the course participated less in PeerWise.

An unanticipated and rather surprising finding was the lack of impact of a student’s Total Answer score (correctness of answers) on final course grade performance. This is surprising because students had to answer and rate more questions than they had to generate. This finding also indicates that the correctness of their answers did not predict their overall performance in the course. We would hypothesize that students answering more questions correctly while practicing for an examination would then answer more questions correctly on the examination.

---

**Table 2. Regression Results for the Four PeerWise Component Scores and Campus Location in Predicting Overall Course Score ($r=.24, r^2=.12$)**

| Component                  | Standardized Beta | p value | Cohen’s d |
|----------------------------|-------------------|---------|-----------|
| (Constant)                 | .00               | —       | —         |
| Campus                     | -.22              | .00     | .47       |
| Total Answer Score         | -.097             | .32     | .09       |
| Question Authoring         | .22               | .00     | .43       |
| Question Answering Component | .17             | .29     | .096      |
| Question Rating Component  | -.033             | .77     | .028      |

Compared to the lowest tertile, the highest tertile had higher component scores for the question answering score ($p=.010$, $d=.32$) and question rating score ($p=.012$, $d=.30$); for these factors, there were no differences between the top and middle tertile. There were no significant findings in total answer scores.

---

**American Journal of Pharmaceutical Education 2018; 82 (2) Article 6315.**
providing for an easily implemented tool. This tool can automatically and efficiently generate data to provide feedback the student and faculty on question quality. However, other methods may provide the same or similar benefit. From this study, it may be important to emphasize to students that generating quality questions can help facilitate learning as well as retrieving their knowledge through question answering.

REFERENCES
1. Bjork EL, Little JL, Storm BC. Multiple-choice testing as a desirable difficulty in the classroom. *J Appl Res Mem Cogn.* 2014;3(3):165-170.
2. Hattie J, Timperley H. The power of feedback. *Rev Educ Res.* 2007;77(1):81-112.
3. Hays MJ, Kornell N, Bjork RA. The costs and benefits of providing feedback during learning. *Psychon Bull Rev.* 2010;17(6):797-801.
4. Rowland CA. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull.* 2014;140(6):1432-1463.
5. Pittenger AL, Lounsbery JL. Student-generated questions to assess learning in an online orientation to pharmacy course. *Am J Pharm Educ.* 2011;75(5):Article 94.
6. Tatachar A, Li F, Gibson CM, Kominski C. Pharmacy students’ perception of learning and satisfaction with various active learning exercises. *Curr Pharm Teach Learn.* 2016;8(4):577-583.
7. Walsh JL, Denny P, Smith PE. Encouraging maximal learning with minimal effort using PeerWise. *Med Educ.* 2015;49(5):521-522.
8. Kolluru S. An active-learning assignment requiring pharmacy students to write medicinal chemistry examination questions. *Am J Pharm Educ.* 2012;76(6):Article 112.
9. Persky AM, Wells MA, Sanders KA, Fiordalisi J, Downey C, Anksorus HN. An interdisciplinary approach improving long-term retention of pharmacology knowledge. *J Dent Educ.* In press.
10. Hunsu NJ, Adesope O, Bayly DJ. A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Comp Educ.* 2016;94:102-119.
11. Nelson C, Hartling L, Campbell S, Oswald AE. The effects of audience response systems on learning outcomes in health professions education. A BEME systematic review: BEME Guide No. 21. *Med Teach.* 2012;34(6):e386-e405.
12. Schneider M, Preckel F. Variables associated with achievement in higher education: a systematic review of meta-analyses. *Psychol Bull.* 2017;143(6):565-600.
13. Mussweiler T, Strack F. Numeric judgments under uncertainty: the role of knowledge in anchoring. *J Exp Soc Psychol,* 2000;36(5):495-518.
14. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science.* 1974;185(4157):1124-1131.