Using student-generated content to engage students in upper-division quantum mechanics

Antje Kohnle

School of Physics and Astronomy, University of St Andrews, St Andrews, KY16 9SS, United Kingdom

Peerwise is an online peer learning community in which students can author, answer, and give feedback on each other's multiple-choice questions. We describe the implementation of Peerwise in a junior-level quantum mechanics course over seven iterations, with 1369 student-generated questions in total. We describe measures of student engagement in terms of answering questions to prepare for course assessments and the improvement of questions. We discuss factors in the implementation that may have impacted engagement, including the timing of submissions, support for authoring high-quality questions, instructions on commenting and minimal requirements on authoring and commenting.

I. INTRODUCTION

Peerwise [1–3] is an online peer learning community in which students can create, answer, and discuss each other's multiple-choice questions. Students author questions by creating the question text, distractors (i.e. plausible incorrect answers), the solution and keywords for the question topic. After answering a question, students see the given solution, can rate the question for difficulty and quality and engage in discussions with the author. Students can earn badges for various aspects of their activity, and increase their Peerwise score by authoring, answering, rating and commenting on questions. All activity is anonymous to students (but not to the instructor), and students do not see usernames or other forms of identification.

Having students generate problems themselves as creators of educational content may have multiple benefits in terms of fostering deep learning, promoting conceptual understanding and enhancing engagement [4–6]. When constructing a multiple-choice question, students need to consider alternate responses and explain in their solution why these responses are correct or incorrect, which is cognitively more demanding than just answering a given question [4]. The Peerwise system also allows differentiation and self-regulation of activity, e.g. students can answer rather than author questions on topics they are finding difficult, or create challenging questions if they have a strong understanding of the material [7].

Peerwise has been used widely at the introductory level, e.g. in computer science [8, 9], chemistry [10], biology [11], psychology [12] and physics [5–7, 13, 14]. Studies implementing Peerwise into introductory physics courses have found benefits in terms of engagement and learning [5, 7, 13, 14]. Bates et al. [13] found that questions and solutions were of high quality. Mapping question quality to Bloom's taxonomy, they found that more than three-quarters of questions tested higher-order skills, e.g. application, analysis, evaluation or synthesis. Studies have also found a significant positive relation between students' overall engagement with Peerwise at the introductory physics level and their exam performance, including after controlling for prior academic achievement [5, 7, 14].

This article extends previous work studying Peerwise at the introductory level to consider the implementation in an upper division physics course. Arguably, peer learning plays an equally important role in the upper division as at the introductory level. Peer learning using instructor-generated multiple-choice questions (i.e. clicker questions) has been profitably incorporated into transformed interactive-engagement upper division courses [15]. Peerwise takes this a step further in using student-generated content to promote peer learning.

The Peerwise system has affordances that make it useful for more advanced physics courses, such as a LaTeX-style maths editor to input formulas and the possibility to insert images and animations created in computing environments. It is also possible to insert handwritten scans to reduce the focus on technical aspects of authoring questions, particularly for those of a highly mathematical or graphical nature.

This article describes the implementation of Peerwise in a junior-level quantum mechanics course at the University of St Andrews over seven iterations (with one iteration per year). The course has an enrolment of typically 70 to 100 students, all of whom are physics majors. The course covers standard wave mechanics topics that include the harmonic oscillator, the hydrogen atom, ladder operators and angular momentum. The course makes use of clickers but has a significant lecture component. The author was the course instructor.

Students in the course had not encountered Peerwise previously. An introduction session at the start of the course motivated the use of the system and gave students guidance in authoring high-quality questions, i.e. questions that test higher-order cognitive skills, include plausible incorrect choices and detailed solutions that explain why each of the choices is correct or incorrect. Peerwise counted for 6% of the course credit, and thus was only a small contribution to the course grade. Half of students' Peerwise grade was fulfilling minimal requirements in terms of authoring at least two questions, answering at least ten questions and rating and commenting on at least six questions over the course of two submission deadlines (in the fourth and tenth week of the eleven week teaching period). The first iteration had only a single submission deadline in the tenth week of the course. The other half of the Peerwise grade was students' total Peerwise score (which was visible to students in the system and combines scores for authoring, answering, commenting and rating), with students attaining 100% for a Peerwise score at or above a specified fixed value. The introduction session structure and the total quantity of the submissions were similar to the implementation by Bates et al [13], and remained the same over the iterations. There was only close observation, but no intervention, by the instructor.

In order to promote a focus on conceptual understanding, students were asked to author at least one interpretive question, focusing on how to interpret quantum behaviour. Students were given guidance on what constitutes an interpretive question in the introduction session.

Prior studies in physics measured engagement with Peerwise in terms of the quantity and quality of submissions for the different components, the number of days of activity and the timeline of use [5–7, 13, 14]. In this study, we operationalized engagement in terms of the quantity of submissions and the timeline of use, and additionally considered revisions to authored questions. This article discusses the following research questions (RQs):

1. How did students engage with Peerwise in this course?
2. What factors in the implementation may have promoted student engagement?

This study is limited to student engagement with the system rather than measures of student learning. Section II discusses measures of student engagement with a focus on RQ1. Section III discusses the results in terms of factors that may have promoted student engagement with a focus on RQ2.
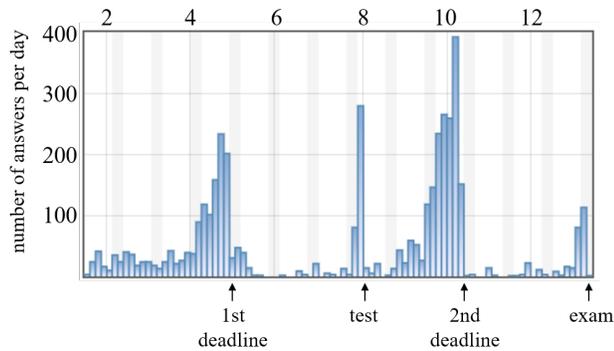
FIG. 1. The number of submitted answers per day for iteration 5 as a function of time (measured in weeks of the semester as the horizontal axis). Also shown are the timings of the two Peerwise submission deadlines and the mid-term test and final exam.

II. MEASURES OF STUDENT ENGAGEMENT

In this section, we describe students' use of the system across the semester (II A), the editing of questions in order to make improvements (II B) and students' preferences in terms of answering, rating, commenting and authoring questions (II C). Results make use of the metrics given in the Peerwise instructor interface as well as an analysis of the content of questions and comments.

A. Timeline of Peerwise use

Figure 1 shows the number of submitted *answers* per day across the semester for the fifth iteration of the course. These histograms look similar for all iterations excepting the first one, where there was only one rather than two Peerwise submission deadlines. One can see low-level activity over the duration of the course (especially in the first weeks prior to multiple deadlines in other courses), with four peaks in activity. Two of the peaks in activity in Fig. 1 precede the two Peerwise submission deadlines. Interestingly, one can see two further peaks in activity that precede the mid-term test and the final exam. Averaged over all iterations, the percent of students in the class using Peerwise in the week prior to the test and after the second submission deadline (prior to the exam) were 26% (standard deviation SD=8%) and 22% (SD=7%) respectively. These were in part different students, so that on average a total of 38% (SD=9%) of students used Peerwise to prepare for the test and/or the exam. It is worth noting that the activity beyond the second submission deadline did not contribute to course credit. These results may indicate that a sizable fraction of students perceived answering the student-generated questions to be useful tools to prepare for the course assessments.

In contrast to Fig. 1, the histogram of the number of submitted *questions* per day (not shown) has just two peaks in activity (both prior to the submissions deadlines), along with

low-level activity across the teaching period. No questions were submitted beyond the second submission deadline. This result most likely reflects the greater time and effort needed to author questions compared with answering questions.

For the first iteration with only a single deadline towards the end of the teaching period, there were only about half the number of questions available at the time of the mid-term test compared with the other iterations with two deadlines (29% of total questions for iteration 1 compared with 50% (SD=3%) for iterations 2 to 7 combined). There was also no marked increase in the use of Peerwise prior to the test.

B. Improvement of questions

Peerwise allows students to edit the questions they have authored to make improvements. The previous version of the question is then archived and is not available anymore for students to answer. Students can also delete their questions.

Table I compares the number of live questions (those that are available for students to answer) and archived questions at the end of each of the course iterations. Only small numbers of questions were deleted (4 to 13 questions per iteration), and are therefore included in the archived question column.

The right-hand column of Table I shows that for each of the Peerwise iterations a significant fraction of students archived (or deleted in a small number of cases) one or more of their questions in order to make improvements. This result could indicate that students have a sense of ownership in the questions they have authored.

Students were required to give meaningful physics comments on at least six questions they answered over the two submission deadlines, pointing out in a specific and constructive way what is good about a question, what could be improved and how. We determined for each archived and deleted question whether there had been comments pointing out errors or making suggestions for improvement that may have initiated the author to revise their question. Table II sum-

TABLE I. The number of Peerwise questions available to answer ("live questions") and archived (or deleted in a small number of cases) at the end of each of the course iterations. The right-hand column is the percent of students that archived (or deleted) questions in each of the iterations.

course iteration	number of students	live questions	archived questions	percent of students archiving questions
1	69	144	42	36.2
2	86	179	60	39.5
3	83	214	118	66.3
4	74	186	78	48.6
5	92	224	63	44.6
6	77	197	91	55.8
7	98	225	59	40.8

TABLE II. The percent of questions that were archived or deleted following one or more comments pointing out an error, making a suggestion for improvement or without a comment.

course iteration	comments prior to archiving/deleting		
	pointing out error	suggesting improvement	no comment
1	26.2%	30.9%	42.9%
2	33.3%	40.0%	26.7%
3	37.3%	33.9%	28.8%
4	23.1%	29.5%	47.4%
5	30.2%	34.9%	34.9%
6	24.2%	33.0%	42.8%
7	42.4%	30.5%	27.1%

marizes the results. The "no comment" column in Table II includes a small number of cases (0 to 2 cases per iteration) where comments were wholly positive. A small number of comments that both pointed out an error and made a suggestion for improvement are included in the "pointing out error" column.

Table II shows that across the iterations 53% to 73% of questions were archived following one or more comments pointing out errors or suggesting improvements. Common suggestions for improvement included defining terms or the context more clearly, improving graphics, adding detail to solutions or improving choices such as adding another plausible incorrect choice.

Comments pointing out errors included conceptual errors, mathematical errors (both computational and conceptual in nature) and flaws in question design such as missing information. For the 159 questions in total across all iterations that were archived or deleted following a comment pointing out an error, 152 (96%) of them pointed out actual errors. 93% of these questions were revised, and 84% of them successfully with the error removed. Thus, these comments did indeed improve the quality of the live questions.

C. Student preferences

As described in section I, half of the Peerwise course credit was students' total Peerwise score (visible to students in the system), with students attaining 100% for a Peerwise score at or above a specified fixed value. In order to increase their Peerwise score, students could freely engage in authoring, answering, commenting and rating of questions as long as they fulfilled the minimal requirements (answering ten, rating and commenting on six and authoring two questions).

Table III shows the mean, median and maximum values across these components for each of the course iterations. As shown in Table III, students preferred to increase their Peerwise score by answering questions, rating questions, commenting on questions and authoring questions, in this order.

Averaged over all iterations, 92% of students exceeded the minimal requirements in terms of answering, 92% in terms of rating, 74% in terms of commenting and 32% in terms of authoring. Averaged over all iterations, between 3% and 8% of students did not meet the minimal requirements for the individual components, and are included in the values in Table III. Students rated the majority of questions that they answered (77% of questions averaged over all iterations).

Authoring questions thus seemed most challenging to students, in terms of the majority of students authoring exactly two questions, the minimum required. This result is not surprising, given that the time and effort to generate a question is likely to be substantially greater than the time and effort needed to answer or comment on a question. Commenting requires more thought about specific strengths and weaknesses of a question than rating.

When students view lists of questions available to answer, they are shown additional information including each questions' average difficulty rating (from students that had answered this question previously) on a seven-point scale from "very easy" to "very hard". We assessed whether there was a correlation between the difficulty rating and the number of submitted answers by determining Spearman's correlation coefficient r_s. Table IV shows that for each of the iterations excepting the first one, there was a significant negative correlation of moderate strength between difficulty rating and the number of submitted answers. Across iterations 2 to 7, questions rated as hard or very hard were answered on average 10.2 times (SD=5.6), whereas questions rated as very easy on average 18.8 times (SD=8.9). While this is only a correlation and the difficulty ratings are not validated, these results may indicate that students on average preferred answering questions that required less effort or time to answer.

The lack of correlation seen for the first iteration ($p = 0.478$, two-tailed) may be due to the single submission deadline, with authoring and answering activity both peaking only prior to this deadline. Thus, there may have been less choice in questions on particular topics compared with the other iterations, where more questions were available early on, and students continued answering them over a longer time period.

III. DISCUSSION AND CONCLUSIONS

This section summarizes the findings from section II in relation to RQ1, and then discusses factors that may have promoted student engagement (RQ2). While there were variations across cohorts, the overall results from section II showed similar trends across the iterations. Activity was peaked prior to the submission deadlines, but in addition 38% of students (averaged over the iterations) used Peerwise prior to at least one of the course assessments. A substantial fraction of questions were archived to make improvements, the majority of these following a comment pointing out an error or making a suggestion for improvement. Students preferred answering, rating, commenting on and authoring questions in

TABLE III. The mean, median and maximum numbers of answered questions, ratings, comments and authored questions for each of the course iterations. The minimal requirements were to answer ten questions, rate and comment on six questions, and author two questions.

course	answered questions			ratings			comments			authored questions		
iteration	mean	median	max	mean	median	max	mean	median	max	mean	median	max
1	31.9	27	117	26.0	22	112	11.7	8	78	2.1	2	7
2	26.5	23	86	20.4	16	82	11.0	8.5	45	2.1	2	4
3	37.7	32	127	29.0	25	96	15.7	13	50	2.6	2	6
4	37.2	28.5	169	29.6	23.5	144	12.6	9.5	42	2.5	2	8
5	40.1	32.5	146	31.5	25	132	10.6	9	29	2.4	2	5
6	38.6	36	163	30.1	26	113	13.8	11	90	2.6	2	6
7	31.9	26	134	26.6	20	126	10.4	8	34	2.3	2	4

TABLE IV. Spearman's correlation coefficient r_s between the difficulty rating of questions and the number of submitted answers. The r_s values with a star (iterations 2 to 7) each have a probability value $p < .0005$ (two-tailed).

iteration	1	2	3	4	5	6	7
r_s	$-.06$	$-.39^*$	$-.40^*$	$-.28^*$	$-.47^*$	$-.50^*$	$-.36^*$

this order to increase their Peerwise score.

We now discuss factors that may have promoted student engagement (RQ2). The results in section II A indicate that two submission deadlines, with the first not too late in the teaching period, are preferable to a single deadline in terms of the timeline of student use. The first iteration of Peerwise had only a single deadline towards the end of the teaching period. This led to fewer questions being available prior to the midterm test and no marked increase in the use of Peerwise prior to the test. The earlier first deadline from the second iteration onwards led to more questions being available early in the semester, and greater use of Peerwise to prepare for the test.

The results in section II B indicate that asking students in their comments to point out strengths and specific ways a question can be improved were important in driving the improvement of questions, as they led to students editing their questions and thus likely contributing to a higher-quality resource. The improvement in quality may also impact students' perceptions of the usefulness of answering questions beyond the minimal requirements (section II C) and the use of Peerwise as a tool to prepare for the course assessments (section II A). These results could indicate that asking students to make specific physics comments on questions had multiple benefits; namely for the students making the comments (in terms of critical assessment of the questions and scientific discourse), the question authors (via revisions to improve questions) and all students on the course (in terms of improving the quality of the live questions).

The results in section II C indicate that the minimal requirements for authoring and commenting on strengths and weaknesses of questions are more important than those for rating and answering questions. Authoring and commenting on questions are more effortful than rating and answering questions. Thus, the minimal requirements on these components more strongly determine the quantity of submitted questions and comments, and are therefore important in ensuring that there are sufficient questions for students to answer, and in ensuring that strengths and weaknesses of questions are fed back to authors.

While the format of the introduction session was not varied across the iterations, Bates et al. [13] suggest that this session plays a key role in giving students support in authoring high-quality questions, e.g. recognizing the importance of clarity of the question text, plausible distractors and sufficiently detailed solutions. In order to promote Peerwise use early in the semester, we asked for volunteers to create questions on specific topics in the first two classes. High-quality questions were also fed back by showcasing a Peerwise question (without the answer) at the start of most classes. We used the assignment criteria in terms of the types of questions to be submitted to steer the peer learning towards the development of conceptual understanding rather than just mathematical formalism.

In conclusion, the results indicate that peer learning through student-generated content can be a worthwhile addition to upper division physics courses. This article only considered engagement with the system and did not focus on learning gains. From studies at the introductory physics level [5, 7, 14], it seems plausible that various aspects of Peerwise use (e.g. engagement with the course content, scientific discussion with peers, critical thinking, ownership, creativity) could enhance learning outcomes. Future work will relate student engagement with Peerwise to performance in the course.

[1] Peerwise, https://peerwise.cs.auckland.ac.nz, accessed June 23, 2019.

[2] P. Denny, J. Hamer, A. Luxton-Reilly, and H. Purchase, Peerwise: students sharing their multiple choice questions, in *Proceedings of the fourth international workshop on computing education research* (ACM, 2008) p. 51.

[3] P. Denny, A. Luxton-Reilly, and J. Hamer, The peerwise system of student contributed assessment questions, in *Proceedings of the tenth conference on Australasian computing education-Volume 78* (Australian Computer Society, 2008) p. 69.

[4] S. W. Draper, Catalytic assessment: understanding how MCQs and EVS can foster deep learning, British Journal of Educational Technology **40**, 285 (2009).

[5] S. P. Bates, R. K. Galloway, and K. L. McBride, Student-generated content: Using peerwise to enhance engagement and outcomes in introductory physics courses, in *AIP Conference Proceedings*, Vol. 1413 (AIP, 2012) p. 123.

[6] M. M. Casey, S. P. Bates, K. W. Galloway, R. K. Galloway, J. A. Hardy, A. E. Kay, P. Kirsop, and H. A. McQueen, Scaffolding student engagement via online peer learning, Eur. J. Phys. **35**, 045002 (2014).

[7] A. E. Kay, J. Hardy, and R. K. Galloway, Student use of peerwise: A multi-institutional, multidisciplinary evaluation, British Journal of Educational Technology 10.1111/bjet.12754 (2019).

[8] L. Hakulinenand A. Korhonen, Making the most of using peerwise in education, in *Reflektori 2010 Symposium of Engineering Education* (Citeseer, 2010) p. 57.

[9] P. Denny, E. Tempero, D. Garbett, and A. Petersen, Examining a student-generated question activity using random topic assignment, in *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (ACM, 2017) p. 146.

[10] K. W. Galloway and S. Burns, Doing it for themselves: students creating a high quality peer-learning environment, Chemistry Education Research and Practice **16**, 82 (2015).

[11] H. A. McQueen, C. Shields, D. J. Finnegan, J. Higham, and M. W. Simmen, Peerwise provides significant academic benefits to biological science students across diverse learning tasks, but with minimal instructor intervention, Biochemistry and Molecular Biology Education **42**, 371 (2014).

[12] P. D. L. Howe, M. McKague, J. M. Lodge, A. G. Blunden, and G. Saw, Peerwise: Evaluating the effectiveness of a web-based learning aid in a second-year psychology subject, Psychology Learning & Teaching **17**, 166 (2018).

[13] S. P. Bates, R. K. Galloway, J. Riise, and D. Homer, Assessing the quality of a student-generated question repository, Phys. Rev. ST Phys. Educ. Res. **10**, 020105 (2014).

[14] J. Hardy, S. P. Bates, M. M. Casey, K. W. Galloway, R. K. Galloway, A. E. Kay, P. Kirsop, and H. A. McQueen, Student-generated content: Enhancing learning through sharing multiple-choice questions, International Journal of Science Education **36**, 2180 (2014).

[15] S. J. Pollock, S. V. Chasteen, M. Dubson, and K. K. Perkins, The use of concept tests and peer instruction in upper-division physics, in *AIP Conference Proceedings*, Vol. 1289 (AIP, 2010) p. 261.