Received: 12 May 2021 | Accepted: 18 June 2021

DOI: 10.1002/prp2.833

ORIGINAL ARTICLE

Answering questions in a co-created formative exam question bank improves summative exam performance, while students perceive benefits from answering, authoring, and peer discussion: A mixed methods analysis of PeerWise

Clare Guilding[1] | Rachel Emma Pye[2,3] | Stephanie Butler[4] | Michael Atkinson[1] | Eimear Field[1]

[1]Faculty of Medical Sciences, School of Medical Education, Newcastle University, Newcastle Upon Tyne, UK

[2]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[3]Psychology Section, University of Reading Malaysia, Malaysia

[4]Department of English, University of Toronto, Toronto, Ontario, Canada

Correspondence
Clare Guilding, Faculty of Medical Sciences, School of Medical Education, Newcastle University, Newcastle Upon Tyne NE2 4HH, UK.
Email: clare.guilding@newcastle.ac.uk

Funding information
Faculty of Medical Sciences, Educational Research and Development Practice Grant

Abstract

Multiple choice questions (MCQs) are a common form of assessment in medical schools and students seek opportunities to engage with formative assessment that reflects their summative exams. Formative assessment with feedback and active learning strategies improve student learning outcomes, but a challenge for educators, particularly those with large class sizes, is how to provide students with such opportunities without overburdening faculty. To address this, we enrolled medical students in the online learning platform PeerWise, which enables students to author and answer MCQs, rate the quality of other students' contributions as well as discuss content. A quasi-experimental mixed methods research design was used to explore PeerWise use and its impact on the learning experience and exam results of fourth year medical students who were studying courses in clinical sciences and pharmacology. Most students chose to engage with PeerWise following its introduction as a noncompulsory learning opportunity. While students perceived benefits in authoring and peer discussion, students engaged most highly with answering questions, noting that this helped them identify gaps in knowledge, test their learning and improve exam technique. Detailed analysis of the 2015 cohort (n = 444) with hierarchical regression models revealed a significant positive predictive relationship between answering PeerWise questions and exam results, even after controlling for previous academic performance, which was further confirmed with a follow-up multi-year analysis (2015–2018, n = 1693). These 4 years of quantitative data corroborated students' belief in the benefit of answering peer-authored questions for learning.

KEYWORDS
assessment for learning, collaborative learning, formative assessment, gamification, medical education, peer learning, PeerWise, MCQ, single best answer

Abbreviations: MCQ, multiple choice question; NUMed, Newcastle University Medicine Malaysia; SBA, Single Best Answer; STEM, science, engineering, technology, and maths.

1 | INTRODUCTION

1.1 | Background and rationale for the educational activity

Active and collaborative learning strategies, as well as frequent formative assessment with feedback, improve student learning outcomes.[1–5] Students are particularly keen to engage with formative assessment that reflects their summative exams and which they can do in their own time via e-assessment.[6] A challenge for educators, particularly those with large class sizes is how to provide students with such learning opportunities without overburdening faculty.

Multiple choice questions (MCQs) are a common form of assessment in medical degree programs. Faced with high-stake exams, medical students revise strategically, seeking opportunities to practice MCQs to test knowledge and recall of course content.[7,8] The Newcastle University medical (MBBS) degree uses MCQ/Single Best Answer (SBA; a form of MCQ[9]) exams in each year of the program but does not release past exam papers. Consequently, students seek practice exam-style questions from faculty or online question banks. Faculty have limited capacity to produce exam items for formative learning in addition to those produced for summative exams; furthermore, online question banks often charge students and the content is not curriculum specific.[10] To address these challenges, we enrolled medical students in the online learning platform PeerWise, which enables students to author and answer MCQs/SBAs, rate the quality of other students' contributions as well as discuss content. It provides an opportunity for self- and peer assessment as well as incorporating aspects of active learning. As a student generated and moderated resource it requires minimal faculty input once established.

1.2 | Pedagogical principles

Constructivist learning theory posits that active self-construction of meaning and understanding through engagement with authentic learning tasks leads to better learning, with new knowledge built on previous knowledge.[11,12] Incorporating PeerWise into a course of study creates several active learning opportunities rooted in constructivist learning theory. Writing questions for example, engages higher order thinking processes, engages students deeply in course content, and choosing distractors raises awareness of misconceptions and plausible alternative interpretations of content.[13–15] Students writing, explaining, and discussing questions with peers are actively engaged in constructing their own knowledge and understanding of the course learning outcomes; they thereby build a collaborative learning community which studies show can enhance student performance.[16] Practice testing has a range of cognitive, metacognitive, and noncognitive benefits.[17] It reinforces learning, improves long-term retention and can be a key predictor of examination performance.[17–19] Answering questions enables students to check their understanding of course content, and provides formative and corrective feedback which are important factors in promoting learning.[20]

PeerWise has been used in learning institutions worldwide.[21] Studies of PeerWise use and impact come predominately from undergraduate STEM subjects (science, engineering, technology, and maths); many have shown positive correlations of overall PeerWise activity with examination results.[22–26] While it is difficult to disentangle the effect of the intervention from confounding factors such as other simultaneous learning activities and the intrinsic ability of students, studies in pharmacy, political science, accountancy, and STEM courses have shown improvements in assessments results when confounds such as previous academic performance are factored into the analyses.[13,27–29] Attempts to determine which aspects of PeerWise use contributed to this positive correlation indicated that academic benefits arose from engaging across a range of activities.[23,29,30]

Fewer studies investigate the impact of PeerWise use in medical degree programs. Some studies showed correlations between question answering, authoring, and commenting with summative exam performance, with the strongest correlations seen for question authoring.[31–33] However, none of these studies controlled for intervention or baseline academic differences, thus students who engaged with the PeerWise activities may represent the proactive high-achieving students who would do well on summative assessment regardless. Other studies investigating the impact of medical or pharmacy students authoring MCQs on academic achievement, using control groups of students who did not create MCQs, found no statistically significant difference between groups.[34,35]

While the impact of engaging with PeerWise on academic performance in medical students remains unclear across studies, qualitative analyses of students' perceptions of PeerWise use, in particular question answering, were generally favorable.[26,31,32] Results from qualitative studies suggest that students found PeerWise an enjoyable learning experience which improved their learning competency, engagement with the materials, and motivation to learn.[26,31,32] Most interventions used PeerWise as a compulsory component of the curriculum, with engagement manipulated through inclusion of PeerWise exercises in summative assessment.[12,36,37] There remain contradictory reports, however, over whether medical students find compulsory MCQ writing exercises of educational benefit.[38,39] In this paper, we explore medical students' use of PeerWise and their perceptions of the value of PeerWise as a learning tool, and investigate the relationship between PeerWise use and assessment results, controlling for previous academic performance.

2 | METHODOLOGY

2.1 | Learning environment

The Newcastle University MBBS program is 5 years in duration and delivered on two campuses: at Newcastle upon Tyne in the United Kingdom, and at Newcastle University Medicine Malaysia

(NUMed) in Johor, Malaysia. Students from the United Kingdom and Malaysia are considered one cohort; the same curriculum is delivered on both campuses, students sit the same exams and they graduate with a Newcastle University MBBS degree. In the 2015 target cohort for this project, 444 fourth year students from Newcastle University (350 from United Kingdom and 94 from Malaysia) were enrolled in PeerWise at the start of Stage 4 (September 2015). In Stages 1–2 students had studied the essentials of biomedical sciences and clinical skills, Stage 3 was clinically based learning. Semester 1 of Stage 4 was primarily campus-based learning within two main courses: Clinical Sciences and Investigative Medicine, and Clinical Pharmacology, Therapeutics, and Prescribing (hereafter referred to as Clinical Sciences, and Clinical Pharmacology, respectively). In December 2015, students took an integrated 300-question MCQ/SBA exam assessing the content of the semester. This exam was the only knowledge-based exam in Stage 4, and students who failed this exam failed the year and proceeded to resit in summer. Due to the integrated nature of the assessment, it was not possible to determine performance in an individual course or subject area. Students' PeerWise use in Semester 1 and performance in the end of semester "Stage 4 exam" was investigated for this project.

2.1.1 | PeerWise use

Students were introduced to the Stage 4 Clinical Pharmacology and Clinical Sciences PeerWise courses through an email to the year group which provided information on how to use the site, the pedagogic rationale and potential benefits, together with detailed written instruction on how to write MCQ/SBA questions. Students were asked to write a question each, include explanations for the correct answer and distractors, tag their question with a topic for question sorting; and they were able to answer, comment on and rate other contributors' questions. Engagement with PeerWise was not a compulsory part of the curriculum; PeerWise was not included within set teaching and learning activities and participation was independent of summative Stage 4 assessment. Thus, students' use of PeerWise was self-directed and interaction with the site was for formative rather than summative assessment and feedback. The PeerWise courses were established partway through Semester 1 for the 2014 cohort. At the start of the 2015 academic year, the Clinical Pharmacology course contained 65 questions and the Clinical Sciences course contained 123 questions written by the previous cohort and available to the target cohort of students to answer.

2.2 | Research design

This study aimed to explore PeerWise use and its impact on the learning experience and exam results of fourth year medical students. The research questions were:

(i) How do students engage with authoring, answering, and commenting on formative assessment questions in PeerWise?
(ii) Does student engagement with PeerWise predict performance in Stage 4 exams, when taking Stage 3 performance into account, and are particular elements of PeerWise, such as authoring questions, driving this relationship?
(iii) What are students' perceptions of the benefits and challenges associated with engaging in authoring, answering, and commenting on formative assessment questions in PeerWise?

A quasi-experimental, convergent parallel mixed methods research design was used.[40] Data collection were concurrent: at the start of Semester 2 following completion of the Stage 4 exam, quantitative data were generated automatically and downloaded directly from the PeerWise system, and a qualitative survey was distributed to students assessing their perceptions and use of PeerWise. Qualitative and quantitative data were analyzed separately, and the strands mixed during interpretation (see Section 4). This mixed methods approach was chosen for (a) triangulation––qualitative survey statements of perceived benefits of PeerWise use compared with quantitative analysis of impact on exam performance, and (b) expansion––extending the breadth of the enquiry by using the qualitative and quantitative methods for the different research questions.[41]

The research was approved by the School of Medical Education Research Management Group and ethical approval gained from the Faculty of Medical Sciences Ethics Committee at Newcastle University (reference #1608/2015). All participants were advised in writing of the aims of the research, that their consent to participate was voluntary, that the data may be used for publication and they could withdraw their consent and contributions at any time.

2.3 | Data collection and analysis

2.3.1 | Student engagement measures

Measures of student engagement were automatically downloaded from the PeerWise website. To measure question complexity of student-generated MCQs, two researchers categorized a random sample of 100 questions written by the 2015 cohort, as follows: knowledge, comprehension, application, analysis, evaluation, and creation, based on Anderson and Krathwohl's revised version of Bloom's taxonomy.[42] Questions were categorized independently, and any discrepancies were discussed until agreed.

2.3.2 | Quantitative data analysis

Five measures are available from PeerWise. The authoring, answering, and rating scores incorporate the number of contributions together with how a student's questions are rated by the cohort and how often the cohort subsequently agreed with their ratings.[43] The reputation score is used as a marker for overall engagement with PeerWise and

TABLE 1 Engagement of students with question answering, authoring, commenting, and rating in PeerWise during Stage 4 Semester 1 (September–December 2015). The "number engaged" column is the number of students who participated in each element of PeerWise

PeerWise course	Clinical sciences					Clinical pharmacology				
# Students (%)[a]	359 (81%)					310 (70%)				
Activity measure (number)	Number engaged	Mean	Median	Max	Total	Number engaged	Mean	Median	Max	Total
Questions answered	359	162	137	343	59,920	308	99	99	184	31,124
Questions authored	47	0.5	0	49	203	28	0.3	0	23	104
Question comments	97	1.4	0	48	535	75	0.8	0	21	260
Questions rated	272	107	62	338	39,749	242	66	48.5	180	20,753

Note: The number in brackets shows this value as a % of the total class of 444.

[a]Only includes students who activated their PeerWise account.

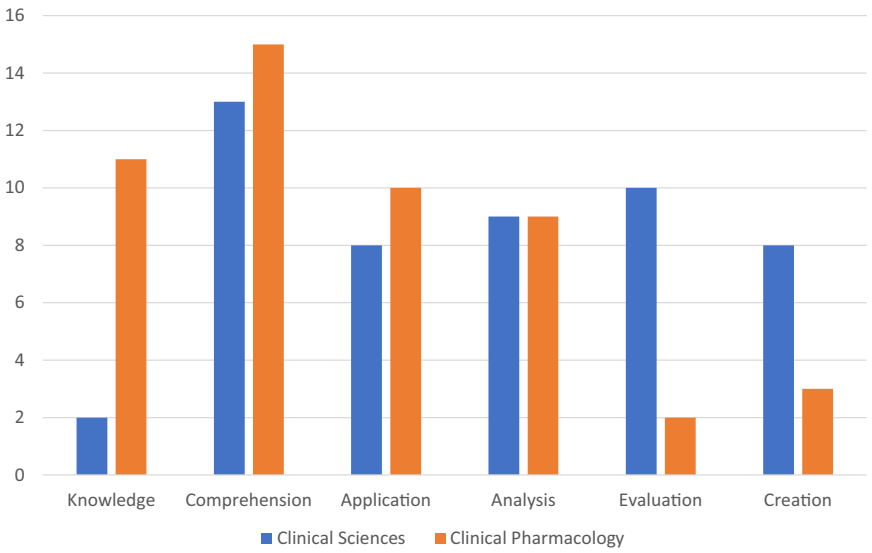

FIGURE 1 Categorization of 50 clinical sciences and 50 clinical pharmacology questions according to Blooms taxonomy. The *Y* axis shows the number of questions within each category

is a combination of the three component scores which favors students participating in all areas. In effect, reputation is a measure of a student's engagement with PeerWise, the timeliness of such engagement (with scores increasing as a result of subsequent student behavior, encouraging early engagement) and how similar their responses were to their peers' later responses. A later addition, the answer score increases each time a student answers a question correctly with a small number of points deducted for incorrect answers, somewhat independent from others' performance.[36] We therefore focused on reputation as a measure of overall engagement, and answer as a measure of knowledge recall.

Our analysis required both Stage 3 and Stage 4 exam scores, therefore we excluded data from students who did not sit both exams. Exam scores were matched to PeerWise usernames and anonymized. All analysis was then carried out on anonymized data. Quantitative data analysis was calculated using IBM SPSS (version 26).

2.3.3 | Survey

A survey was developed by the researchers and students with previous PeerWise experience to explore students' participation with PeerWise, and perceived hindrances and benefits to engagement. The survey was hosted on Bristol Online Survey and underwent content and face validity checks by the members of the team who developed and reviewed it before distribution to the 2015 cohort of Stage 4 students via email. The survey asked the students whether they had authored, answered, or commented on questions, and dependent on their response asked for free text comments on what prevented them from engaging, or about any perceived benefits. A final question asked the students which elements of PeerWise they believed most supported their learning and why.

Free-text comments in the survey were analyzed following the established reflexive thematic analysis approach detailed by Braun and Clarke.[44] Survey responses were downloaded directly into Microsoft Excel and read by researchers to identify "patterns of meaning and issues of potential interest" (p. 15).[44] Researchers independently categorized the responses into broad preliminary codes, then through discussion combined these into the final codes for data analysis. Codes were analyzed, and related codes combined to generate subthemes and overarching themes. Subthemes were quantified to facilitate examination of the magnitude of responses.[45]

3 | RESULTS

3.1 | Student engagement with PeerWise

Engagement of the 2015 target cohort with the various elements of PeerWise is shown in Table 1. We evaluated the cognitive levels the student-authored questions tested.[42] Thirteen percent of the questions were categorized as testing knowledge, 28% comprehension, 18% application, 18% analysis, 12% evaluation, and 11% creation. Cognitive levels tested in the Clinical Pharmacology course were skewed more toward the lower end of Blooms taxonomy than questions in the Clinical Sciences course (see Figure 1).

3.2 | Quantitative PeerWise analysis

3.2.1 | Stage 4 scores pre- and post-PeerWise

To determine if Stage 4 exam performance was higher after the introduction of PeerWise, we used a repeated-measures ANOVA of Stage 3 and 4 results across two cohorts prior to PeerWise introduction (2012, $n = 322$ and 2013, $n = 370$), and following PeerWise introduction (2015, $n = 444$ and 2016, $n = 431$). We did not use 2014 cohort data as PeerWise was established part way through the semester. We first determined that each students' Stage 3 and Stage 4 results significantly correlated within each cohort, using Pearson's correlation (2013 $r = .701$; 2014: $r = .727$; 2015: $r = .734$; 2016: $r = .686$; all values of $p < .001$).

Stage 3 scores remained relatively consistent across cohorts with more variation across Stage 4 (see Figure 2). There was a significant main effect of Stage ($F(1, 1563) = 7603.211$, $p < .001$), with Stage 4 scores significantly lower than Stage 3, and a significant interaction between Stage and Cohort ($F(3, 1563) = 86.8$, $p < .001$). Using Stage 4 minus Stage 3 difference scores to control for cohort differences in the baseline Stage 3 measures, we ran a one-way ANOVA using a priori contrasts to compare pre- and post-PeerWise performance. There was a significant difference, with a smaller score drop post-PeerWise than pre-PeerWise ($t(1563) = 10.46$, $p < .001$), implying a significant benefit of PeerWise on Stage 4 exam performance.

3.2.2 | 2015 cohort analysis

A mean of Clinical Pharmacology and Clinical Science scores for the five PeerWise measures reputation, answering, authoring, rating, and answer was calculated for the 2015 cohort. A linear regression was performed with Stage 4 score as the predicted value, and Stage 3 score entered as the first variable in Model 1, with reputation and answer entered into Model 2. Stage 3 score explained 53.7% of the variance (Model 1: $F(1, 442) = 515.245$, $p < .001$). While Model 2 was significant and predicted 55.3% of the variance, with a significant R^2 change ($F(2, 440) = 8.988$, $p < .001$), the only significant variable was mean answer ($t = 3.751$, $p < .001$), with mean reputation not a significant predictor ($t = 0.407$, $p = .684$). This pattern is also seen when analyzing each PeerWise course separately.

As Authoring has been specifically proposed as a mechanism for better learning,[46] we also regressed mean answering, authoring, and rating as separate variables predicting Stage 4, after entering Stage 3 into the model. The model was significant ($F(4, 439) = 136.548$), explaining 55% of the variance. Only mean answering was a significant predictor ($t = 3.641$, $p < .001$), with mean rating ($t = -1.226$, $p = .221$) and mean authoring ($t = -0.77$, $p = .441$) nonsignificant predictors.

Quartile splits

Following previous analyses,[24,28] we aimed to identify if certain groups of students benefitted from PeerWise use during their course, so repeated the analyses above for each quartile, removing Stage 3 from Model 1. The students scoring higher at Stage 3 (Q1 and Q2) showed the same pattern as the whole cohort (Q1: $F(2, 119) = 3.566$, $p = .031$; Q2: $F(2, 101) = 5.269$, $p = .007$), with a

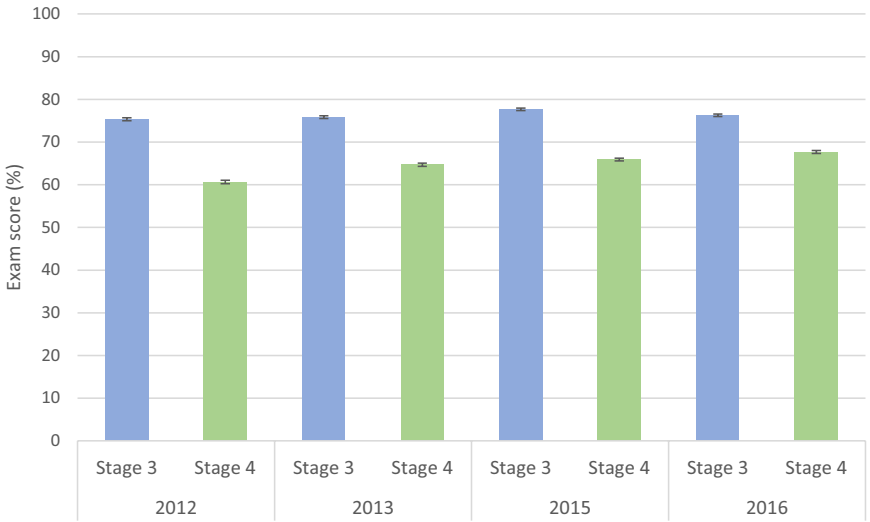

FIGURE 2 Mean Stage 3 and Stage 4 knowledge exam scores over four cohorts: two cohorts before PeerWise introduction (2012, 2013) and two after (2015, 2016). *Y* axis shows the mean exam score for each cohort, error bars are standard error of the mean

significant prediction from mean answer but not mean reputation of Stage 4 results. The models with students who scored lower at Stage 3 (Q3 and Q4) were not significant (Q3: $F(2, 117) = 1.636$, $p = .199$; Q4: $F(2, 95) = 1.871$, $p = .16$), with no significant prediction from either mean reputation or mean answer of Stage 4 results.

3.2.3 | Multi-year analysis

As more cohorts use PeerWise, students' scores on reputation accumulate to reflect the quality of the questions they asked during their enrolment on a particular course. We were interested to explore whether this cumulative indicator of "reputation" was related to Stage 4 performance.

Across four cohorts (2015–2018; $n = 1695$), an hierarchical regression was conducted with Stage 4 results as predicted variable. Stage 3 results were entered first into the model ($F(1, 1694) = 1298.12$, $p < .001$) then mean reputation and mean answer ($F(3, 1691) = 466.06$, $p < .001$). Correlations between the variables were all highly significant, reflecting the large sample size (see Table 2). However, the regression analysis indicated that while mean answer was a significant predictor of Stage 4 results, controlling for Stage 3 ($t = 7.475$, $p < .001$), mean reputation was not ($t = -0.846$, $p = .397$). This corroborates the 2015 analysis, showing over four cohorts, a significant positive relationship between answering PeerWise questions and exam results, even after controlling for previous academic performance.

3.3 | Student perception of the benefits and challenges of PeerWise for learning

There were 174 responses (38.2% of the 2015 cohort) to the survey which assessed students' use of PeerWise and perceived benefits and challenges to participation. The results from the survey question that asked students to rate which elements of PeerWise they found most beneficial are shown in Table 3. Thematic analysis of free text data identified responses in three overarching themes: answering, authoring, and commenting on questions. Themes were composed of subthemes related to what students found beneficial and challenging in these areas (see Table 4).

TABLE 2 Pearson's correlations between Stage 4 and Stage 3 exam scores, and PeerWise mean values of reputation and answering across the two target courses for 2015–2018 cohorts

	Stage 4	Stage 3	Mean reputation
Stage 3	.659***		
Mean reputation	.077**	.083***	
Mean answer	.205***	.106***	.284***

**$p < .01$.; ***$p < .001$.

TABLE 3 Student ratings of perceived relative benefit of different aspects of PeerWise for learning on a 5-point Likert scale, 1 being the least benefit, 5 being the most

	N	Mean	SEM
Answering questions	167	4.40	0.07
Reading explanations of answers	167	4.37	0.06
Writing explanations	108	4.05	0.1
Writing questions	102	3.77	0.12
Evaluating quality of questions	152	2.97	0.1
Commenting on questions	128	2.84	0.1

Abbreviations: *N*, number of student responses; SEM, standard error of the mean.

Below, we describe the themes and subthemes, presenting illustrative data extracts.

3.3.1 | Answering questions theme

93% of survey respondents had answered questions on PeerWise and seven subthemes were identified as perceived benefits of answering (see Table 4). Students commented that answering questions "was hugely beneficial as it showed me the gaps in my knowledge as well as allowing me to see what my peers were revising. It was also very reassuring when I started getting questions right." The testing of knowledge and getting answers right increased students' confidence "it made me feel confident that I was on the right track with my learning." Identification of gaps in knowledge prompted students to do more reading around the subject, and in many cases, completion of questions taught them new material: "Answering questions was a confidence booster if I got the answer correct, and getting them wrong gave me areas to revise again. Also getting them wrong made me learn it, I often found in the real exam I knew the answer because of PeerWise questions."

Many students appreciated the provision of practice exam-style questions, in part because "they were questions—something that are rarely provided to us." Students felt answering questions got them into the mind-set of the exam, helping them think about how questions and the distractors are constructed. Comments included that it was helpful "seeing the possible other options made me more aware of confounders and 'second best' answers to be aware of."

Another subtheme identified as a benefit of answering PeerWise questions was the wide range of curriculum relevant questions, as illustrated by these quotes: "the questions were designed by us and focused specifically on our learning outcomes," "the fact that the questions are written by people on my course most supported my learning, often other online exam tools are very generic and not focused on my outcomes" and "there were many examples of many different questions and the breadth and depth of questions allowed for questioning of how well you knew the subject." Students valued the explanations given for the correct/incorrect answers and used

TABLE 4 Themes and subthemes identified from thematic analysis of student survey responses

Theme	Subtheme (*n*)
Answering	*Benefits from answering* • Identifying gaps in knowledge (43) • Testing and consolidation of learning (41) • Exam technique and practice (29) • Explanations improve knowledge/understanding (15) • Wide range of curriculum relevant questions (13) • Benchmarking against peers (9) • Novel, active revision method (9) *Challenges of answering* • Questions too niche/difficult (15) • Inconsistent question quality (13) • Questions didn't reflect exam (4) • Difficulty using the site (2)
Authoring	*Benefits of authoring* • Stimulated learning through in-depth research into topic (13) • Consolidation of knowledge/understanding (12) • Insight into how exam questions are composed (9) • Ensured thorough understanding of topic for good quality, error free question (8) • Writing distractors and explanations helps identify confounding information (4) • Identification of gaps on knowledge (4) *Challenges of authoring* • Lack of time (41) • Concern about question writing ability (32) • Concern over negative peer feedback (7) • Used other revision method (6) • Difficult/time-consuming to write questions (5) • Bank already full of questions (5) • Unsure of benefit of authoring (4)
Commenting	*Benefits of commenting* • For clarification of question/answers (24) • Correction of incorrect knowledge/understanding (13) • To help and encourage peers (11) • Generates peer discussion (10) • Explaining answers reinforces learning (7) *Challenges of commenting* • Comments already covered what student would have raised (8) • Did not feel the need (6) • Sought clarification in course materials (3) • Concerned comment may be incorrect (2)

Abbreviation: *n*, number of comments within each subtheme.

these to improve their knowledge and understanding, commenting "the section explaining the answer and other answers was helpful. It helped you remember and understand the condition," and "reading the explanations was often helpful (if they were provided) as they were written in an accessible way, especially if they explained why certain answers were incorrect."

The final beneficial subthemes identified included: benchmarking progress against peers, and that answering questions was a novel, active revision method, highlighted by the comments "I thought it was beneficial in allowing me to see what my peers felt I should know for my exams. Medicine has such width and depth of learning sometimes it can be hard to gauge the level I am aiming for with my revision. This tool allowed me to see what my peers considered reasonable exam questions," and "I was able to see if I was catching up with the rest of the cohort or if I was lacking behind and more work was required of me. It also helps to reactivate things that I had been reading and apply it through some of the brilliant and creatively written questions".

Four subthemes were identified around challenges or hindrances to engagement with answering PeerWise questions. Difficult or poor-quality questions had a negative emotional impact on some students who commented "some questions were demoralizing as they focused in too much on minute details of the curriculum." Quality of questions led one student to query whether the site could be trusted for revision "I felt that some of the answers to the questions were debatable. Also, I didn't know if I could trust the answers that were provided, and I preferred not to take the risk with my revision."

3.3.2 | Authoring of questions theme

32% of respondents said they had authored a question, and a fifth of these said the process had consolidated their knowledge or understanding. A further fifth of the respondents said writing questions stimulated them to do further research which increased their learning: "it made me read to ensure I thoroughly understood the topic before creating a question." Part of the drive to do in-depth research was motivated by the pressure to ensure the question was high quality and error free: "I felt responsible to make a water tight question so that it couldn't be disputed which encouraged me to revise more." Writing questions helped some respondents identify gaps in their knowledge: "it takes a greater understanding to write a question than just to answer one, especially in the MCQ format. By trying to write questions, I found gaps in my knowledge that I wouldn't have recognised otherwise."

Students who authored on PeerWise commented that the process gave them helpful insights into how exam questions are composed and the areas of the curriculum that lend themselves to testing by MCQ/SBA: "it gave me an insight into how question authors might be thinking, particularly with regards to phrasing; this definitely influenced my answers in the exam." Writing distractors (incorrect answers), explanations and getting feedback from peers were highlighted by students as beneficial: "the most helpful parts were thinking of incorrect but plausible answers (good for forming differentials) and especially writing an explanation. I feel like the topics I wrote a question on were much easier to answer in the exam."

68% of respondents said they had not authored a question with a quarter of all respondents citing a lack of time, many of them stating they used PeerWise for last minute revision. Some said they would have written a question had they started using it earlier, and they could see the potential benefits of authoring: "Time constraints because I started using Peerwise too late… If I had accessed Peerwise earlier, I might have because authoring a question would be very beneficial as it stimulates our thinking and links up all our thoughts." A fifth of respondents said they did not author because they were concerned about their ability to write a good quality question, this was often related with concern about what their peers would think and a lack of confidence, for example "I didn't feel that I would be able to produce a question that was detailed enough to be of value to others." Some students were explicit that that they did not author because of concern over negative peer feedback as illustrated by these quotes "Literally peer pressure! The fear of making an error in the question/answer."

The time and difficulty in authoring questions inhibited some students, who chose to stick with established learning techniques: "I wrote separate, more basic, Anki questions and couldn't think of any good 'exam' style questions," "it would take a while to write a question (and check it rigorously to make sure it is correct/good) and time is limited whilst I am revising." Others reflected that they did not author because the PeerWise courses already had many questions or they were unsure of the benefit: "often the topics I would have been confident writing questions in were already very well represented with questions," "a lack of belief that it would be personally beneficial for me."

3.3.3 | Commenting on questions theme

41% of survey respondents said they had engaged in commenting on PeerWise questions. The most common reason given for commenting was to clarify elements of the question or answers: "It helped clarify what the correct answer was and why when it was not clear—it was possible to post in explanations and text from lectures etc to show where the answer could be found etc," and "Commenting allowed me to voice any concerns about either my thought process or the structure of the question. The author was then able to tell me their perspective so I understood the question better and was therefore more able to learn from it." Commenting enabled students to correct errors in others' knowledge or understanding, or have their own errors corrected, students stating they commented "to correct a poor question—helped me understand the subject" and "the author corrected me on a point I had misunderstood." The flagging of an error enabled authors to correct their question: "Reading the comments was useful, I pointed out some information had been missed out which was shortly corrected."

A fifth of students who commented said they did so to help and encourage their peers, through praise for questions, use of humor to lighten the mood while studying or through sharing of their knowledge. Students noted they commented "to motivate authors for contributing questions and express gratitude for their efforts," and "the only comment I made was a jovial, light-hearted jest… Perhaps it offered others some much needed humour enabling them to continue on their slogs for a longer period of time!"

The commenting function of PeerWise provided a platform for peer discussion, and through discussion and explanation of answers students were able to better understand the topics, as illustrated by these quotes: "I discussed some questions I felt inaccurately represented the topic or were misleading and this helped both my and the question author's understanding," and "it was most useful when people were willing to engage in constructive debate about why an answer was appropriate or not appropriate. Commenting required me to articulate my knowledge clearly and that process tested whether I actually understood what I was talking about!"

59% of the respondents said they had not used the commenting function on PeerWise. The most common reason given for not responding was that the comments already covered what student would have raised, "I didn't have much to say as quite a few people had already commented with what I might have said." Students said they did not feel the need to comment, sometimes clarifying that this was because the questions were good enough "I never disagreed with the questions, or felt I had anything else to contribute to them." A couple of students noted they would rather seek clarification from

external sources "if I was unsure I like to refer back to my own notes, textbooks, internet sources."

4 | DISCUSSION

This study found that most students engaged with PeerWise following its introduction as a noncompulsory learning opportunity in Stage 4 (Table 1). While students perceived benefits in authoring and peer discussion, students engaged most highly with answering questions, noting that this helped them identify gaps in knowledge, test their learning, and improve exam technique (Table 4). Quantitative analysis corroborated the perception that answering questions was beneficial to learning. Analysis of the Stage 3 and 4 knowledge exams showed a significantly smaller decrease in exam score following the introduction of PeerWise (Figure 2). Detailed analysis of the 2015 cohort indicated that, after controlling for previous academic performance, PeerWise use significantly predicted Stage 4 exam result underpinned by answering questions but not by authoring, peer discussion, or overall engagement. Multi-year analysis of four cohorts supported the 2015 results, showing a sustained significant positive predictive relationship between answering PeerWise questions and exam performance, after controlling for pre-PeerWise exam results.

Most published studies have integrated PeerWise as a taught, compulsory component of the curriculum with engagement contributing to summative assessment to encourage student participation.[12,36,37] Lower participation has been found when no marks were awarded.[15,47] We introduced PeerWise as a non-taught, noncompulsory component of the curriculum, and in contrast, observed high participation, at least with regards to question answering. Studies that investigated the cognitive level assessed by student-created MCQs found the majority of items in STEM courses assessed at the lower levels of Blooms taxonomy.[30,32,37,48,49] Students in this study have been at university longer than those in most STEM courses, having three years' previous experience passing MCQ/SBA exams. Thus, the questions created by our students ranged more widely in complexity, including analysis and evaluation questions (Figure 1).

In previous studies, students frequently noted answering questions as the biggest benefit of using PeerWise.[13,31,50] The perceived benefits of answering such as improved course knowledge, improved exam taking skills, and increased confidence are similar across this study and others.[13,31,38,51] Frequent formative assessment with explanatory feedback has been shown to be valuable in promoting learning.[4,5] Explanatory feedback aids the learning and retention of information,[52,53] and explanations of MCQs which offered solutions to the questions were highlighted in this study as an important learning tool. Practice testing of MCQs has direct and indirect effects on student learning.[17,19] Indirectly, frequent tests stimulate more frequent studying thus more time-on-task, while directly, the retrieval of information during tests results in deeper learning and greater knowledge retention than more passive forms of learning such as reading notes.[5,18]

Our data on exam performance following PeerWise engagement support the perceived benefits of answering questions, our hierarchical regression models consistently showing a significant relationship between question answering and exam performance. Unlike other studies which indicated that the benefits of PeerWise arose from students overall level of activity,[24,29] we did not find a significant relationship between reputation score and exam performance. Interestingly, when we looked at the impact of PeerWise use in ability quartiles, our data indicated that students in the higher two quartiles of ability were significantly benefited, while the lower quartiles were not. Previous studies have shown a range of results regarding the benefits according to ability quartiles. Hardy et al. showed year 1–2 students in the lower intermediate ability levels on STEM programs may benefit most,[24] while other studies showed students in the highest and lowest quartiles in year 1 may benefit most.[14,54] The overall ability in our cohort of fourth year medical students likely differs from those of early year STEM students, making comparisons of PeerWise impact in ability quartiles across studies difficult.

Only a minority of students engaged with question authoring and peer discussion compared to question answering (see Table 1). This result is consistent with previous studies of medical students which showed that even when engagement is linked to summative assessment, or when authoring and peer discussion exercises are included as a taught component of the curriculum, answering questions remained the most popular element.[31,32,55] In this study, PeerWise was used over one semester, with only 12 weeks of intense teaching before the Stage 4 exam. Key reasons cited by the students for not authoring was lack of time and doubt over ability. Authoring MCQs is challenging and time-consuming. It is the challenging nature of the process which educators perceive drives deep learning and understanding; question writing is suggested to employ higher levels of comprehension than answering.[13,34,46] Indeed, students who had authored in our study believed this stimulated in-depth learning and understanding. Similar results were found in other studies, which showed that question authoring activities drove desirable learning behaviors such as the integration of varying types and sources of information.[36,50]

The feasibility and acceptability of question writing as a study technique for medical students has, however, been questioned.[36,50] A compulsory question writing exercise for a pathology course within a medical degree was recently removed due to low student acceptance.[39] Researchers have suggested that the question writing may not be an efficient learning strategy within time pressured programs,[39,50] and our study did not show a significant association between question writing and improved exam performance. Introducing PeerWise as a noncompulsory learning opportunity allowed our students to engage with the elements they perceived most effective for their learning, and may improve acceptance in other interventions.

A concern over question writing ability and a fear of negative peer feedback further inhibited students from authoring. Walsh et al. identified "PeerWise trolling" in their intervention and suggested that introductory sessions define clear rules for online

interactions.[31] In our correspondence with students it was made clear that while the site was anonymous for students, faculty could identify participants and modify interactions. Since launch, 2143 fourth year students have used the site, yet only one interaction was flagged as inappropriate and removed. Providing face-to-face training in question writing should improve students' confidence in authoring, would result in a larger question bank, and has been requested by other UK medical students.[12] Medical students who received in-person training in question writing reported that it gave them insights into how to strategically approach MCQs which they believed improved their test taking aptitude[50]; these are perceived benefits also noted by authors in this study. Yet, not all students are motivated and have the self-regulation skills to write MCQs, plus given the time constraints of the semester a noncompulsory training session may be more acceptable to the cohort.[48]

4.1 | Practical implications

Our study shows significant benefits of answering student-authored MCQs on exam performance. Building a bank of MCQs for formative assessment is time-consuming and expensive.[56] By engaging students in question authoring, large banks of curriculum relevant questions can be generated easily. Beyond the initial account setups and distribution of instructions and usernames each year, there has been minimal faculty input. Even with minimal input, between 2014 and 2020, Stage 4 students authored 1648 questions for the Clinical Pharmacology and Clinical Sciences courses, which were answered 842,298 times, generating 5402 comments. Arguably, introducing taught question writing sessions would generate a greater number of questions, potentially of higher quality[31]; Gooi and Sommerfeld provide guidance for introducing such question writing sessions into curricula.[57]

4.2 | Strengths and limitations

Limitations of this research include that it is limited to a single course within one institution reducing generalizability of findings, although the large sample size may allow for a level of relatability to other contexts. The study did not incorporate a control group due to ethical considerations of allowing only part of the cohort access to an intervention that may have been beneficial, and the response rate to the survey was <40% so different themes could have arisen if all students had responded. The strengths of this research include the large volume of PeerWise data that was available for analysis, collected from students over 4 years. A significant proportion of students were accessing and using PeerWise during the study adding to the credibility of findings. The hierarchical regression analysis allowed for control of previous academic ability in assessment of the impact of PeerWise on summative exam performance, while the mixed methods approach allowed for a rich exploration and triangulation of data that allowed us to gain a better understanding of the barriers and facilitators to engagement with education tools by undergraduate students.

5 | CONCLUSION

As educators, we look to develop students as lifelong independent learners, so that they become increasingly self-regulating and self-directed as they mature into competent graduates.[58] This study suggests that PeerWise provides such opportunities by encouraging learners to work independently and collaboratively to construct their own learning. Despite its introduction as a noncompulsory learning activity, most students used PeerWise and described a range of perceived benefits, which alongside the positive correlation of PeerWise use and summative exam performance support its introduction within programs and can be used as evidence to encourage future participation within medical education contexts.

ACKNOWLEDGMENTS

Many thanks to Professor Jane Stewart in the School of Medical Education at Newcastle University for her expert guidance in the design of this project. Thanks to Associate Professor Paul Denny for designing and developing PeerWise, making it available for free use and for his responsiveness to queries. This project was supported by a Faculty of Medical Sciences, Educational Research and Development Practice Grant.

DISCLOSURE

The authors have no conflict of interest to declare.

ETHICS STATEMENT

The research was approved by the School of Medical Education Research Management Group and ethical approval gained from the Faculty of Medical Sciences Ethics Committee at Newcastle University (reference #1608/2015). All participants were advised in writing of the aims of the research, that their consent to participate was voluntary, that the data may be used for publication and they could withdraw their consent and contributions at any time.

AUTHOR CONTRIBUTION

CG, REP, and EF designed the study; CG and EF acquired the data; REP performed the quantitative statistical analysis; CG, SB, and MA performed the quantitative data analysis. CG and REP wrote the paper; SB and MA critically revised the paper. All the authors have revised and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Clare Guilding https://orcid.org/0000-0003-2823-1575
Rachel Emma Pye https://orcid.org/0000-0002-1395-7763

Stephanie Butler https://orcid.org/0000-0003-0027-7723
Michael Atkinson https://orcid.org/0000-0003-0009-9234

REFERENCES

1. Biggs JB. *Teaching for Quality Learning at University: What the Student Does*. New York, NY: McGraw-Hill Education; 2011.
2. Fink LD. *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*. San Francisco, CA: John Wiley & Sons; 2013.
3. Hattie J, Timperley H. The power of feedback. *Rev Educ Res*. 2007;77(1):81-112.
4. Draper SW. Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *Br J Edu Technol*. 2009;40(2):285-293.
5. Roediger HL III, Agarwal PK, Kang SH, Marsh EJ. Benefits of testing memory: Best practices and boundary conditions. In: Davies GM, Wright DB, eds. *Current Issues in Applied Memory Research*. New York, NY: Psychology Press; 2010:13-49.
6. Wynter L, Burgess A, Kalman E, et al. Medical students: what educational resources are they using? *BMC Med Educ*. 2019;19(1):36.
7. Burk-Rafel J, Santen SA, Purkiss J. Study Behaviors and USMLE step 1 performance: implications of a student self-directed parallel curriculum. *Acad Med*. 2017;92(11S):S67-S74.
8. Kenwright D, Dai W, Osborne E, et al. Just tell me what I need to know to pass the exam!" Can active flipped learning overcome passivity. *TAPS*. 2017;2(1):1-6.
9. Buttery CM. Assessment in medical education. *N Engl J Med*. 2007;356(20):2109; author reply 2110.
10. Rughani G. Online question banks. *BMJ*. 2013;346:f3500. https://doi.org/10.1136/sbmj.f3500
11. Mann K, MacLeod A. Constructivism: learning theories and approaches to research. In: Cleland JA, Durning SJ, eds. *Researching Medical Education*. Hoboken, NJ: Wiley Blackwell; 2015:51-65.
12. Belshaw A, Mackie A, Phillips HK, et al. Comment on: creating assessments as an active learning strategy: what are students' perceptions? A mixed methods study. *Med Educ Online*. 2020;25(1):1709277.
13. Feeley M, Parris J. An assessment of the PeerWise student-contributed question system's impact on learning outcomes: evidence from a large enrollment political science course. *SSRN Electron J*. 2012. https://doi.org/10.2139/ssrn.2144375.
14. Denny P, Hamer J, Luxton-Reilly A, Purchase H. PeerWise: students sharing their multiple choice questions. In: Lister R, ed. *Proceedings of the Fourth International Workshop on Computing Education Research*. Sydney, Australia: Association for Computing Machinery; 2008:51-58.
15. Luxton-Reilly A, Denny P. Constructive evaluation: a pedagogy of student-contributed assessment. *Comput Sci Educ*. 2010;20(2):145-167.
16. Laal M, Ghodsi SM. Benefits of collaborative learning. *Procedia Soc Behav Sci*. 2012;31:486-490.
17. Adesope OO, Trevisan DA, Sundararajan N. Rethinking the use of tests: a meta-analysis of practice testing. *Rev Educ Res*. 2017;87(3):659-701.
18. Dunlosky J, Rawson KA, Marsh EJ, et al. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychol Sci Public Interest*. 2013;14(1):4-58.
19. Karpicke JD, Roediger HL. The critical importance of retrieval for learning. *Science*. 2008;319(5865):966-968.
20. Nicol D. E-assessment by design: using multiple-choice tests to good effect. *J Further Higher Educ*. 2007;31(1):53-64.
21. Denny P. PeerWise. 2021. https://peerwise.cs.auckland.ac.nz/. Accessed April 28, 2021.
22. Denny P, Hanks B, Simon B. Peerwise: replication study of a student-collaborative self-testing web service in a US setting. In: *Proceedings of the 41st ACM Technical Symposium on Computer Science Education*. March 5, 2010:421–425. https://doi.org/10.1145/1734263.1734407
23. Casey MM, Bates SP, Galloway KW, et al. Scaffolding student engagement via online peer learning. *Eur J Phys*. 2014;35(4):045002.
24. Hardy J, Bates SP, Casey MM, et al. Student-generated content: enhancing learning through sharing multiple-choice questions. *Int J Sci Educ*. 2014;36(13):2180-2194.
25. Hancock D, Hare N, Denny P, et al. Improving large class performance and engagement through student-generated question banks. *Biochem Mol Biol Educ*. 2018;46(4):306-317.
26. Khashaba AS. Evaluation of the effectiveness of online peer-based formative assessments (PeerWise) to enhance student learning in physiology: A systematic review using PRISMA guidelines. *Int J Res Educ Sci*. 2020;6(4):613-https://doi.org/10.46328/ijres.v6i4.1216.
27. Doyle E, Buckley P. The impact of co-creation: an analysis of the effectiveness of student authored multiple choice questions on achievement of learning outcomes. *Interact Learn Environ*. 2020;1-10. https://doi.org/10.1080/10494820.2020.1777166.
28. Galloway KW, Burns S. Doing it for themselves: students creating a high quality peer-learning environment. *Chem Educ Res Pract*. 2015;16(1):82-92.
29. Kay AE, Hardy J, Galloway RK. Student use of PeerWise: a multi-institutional, multidisciplinary evaluation. *Br J Edu Technol*. 2020;51(1):23-35.
30. McQueen HA, Shields C, Finnegan DJ, et al. PeerWise provides significant academic benefits to biological science students across diverse learning tasks, but with minimal instructor intervention. *Biochem Mol Biol Educ*. 2014;42(5):371-381.
31. Walsh JL, Harris BH, Denny P, Smith P. Formative student-authored question bank: perceptions, question quality and association with summative performance. *Postgrad Med J*. 2018;94(1108):97.
32. Kadir FA, Ansari RM, AbManan N, Abdullah MH, Nor HM. The impact of PeerWise approach on the academic performance of medical students. *Malays Online J Educ Technol*. 2014;2(4):37-49.
33. Harris BHL, Walsh JL, Tayyaba S, et al. A novel student-led approach to multiple-choice question generation and online database creation, with targeted clinician input. *Teach Learn Med*. 2015;27(2):182-188.
34. Palmer E, Devitt P. Constructing multiple choice questions as a method for learning. *Ann Acad Med Singap*. 2006;35(9):604.
35. Tatachar A, Kominski C. Assessing a traditional case-based application exercise and a student question creation exercise on student performance and perceptions. *Curr Pharm Teach Learn*. 2017;9(4):689-697.
36. Grainger R, Dai W, Osborne E, et al. Medical students create multiple-choice questions for learning in pathology education: a pilot study. *BMC Med Educ*. 2018;18(1):201.
37. Bottomley S, Denny P. A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochem Mol Biol Educ*. 2011;39(5):352-361.
38. Ritu L, Farah O, Nabil Z. Assessment as learning in medical education: feasibility and perceived impact of student-generated formative assessments. Research Square—Pre-Review. 2021.
39. Smith CD, Dai A, Kenwright D, et al. PeerWise and Pathology: Discontinuing a teaching innovation that did not achieve its potential. *MedEdPublish*. 2020;1. https://doi.org/10.15694/mep.2020.000027.1.
40. Creswell JW, Plano Clark VL. *Designing and Conducting Mixed Methods Research*. 2nd ed. Los Angeles, CA: Sage; 2011.
41. Schifferdecker KE, Reed VA. Using mixed methods research in medical education: basic guidelines for researchers. *Med Educ*. 2009;43(7):637-644.
42. Anderson LW, Krathwohl DR. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman; 2001.

43. Denny P. PeerWise. Scoring: for fun and extra credit! nd. https://peerwise.cs.auckland.ac.nz/docs/community/scoring_for_fun_and_extra_credit/#:~:text=Once%20again%2C%20to%20get%20a,%22%20or%20%22excellent%22%20respectively. Accessed April 12, 2020.
44. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101.
45. Boyatzis RE. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks, CA: Sage Publications; 1998.
46. Rosenshine B, Meister C, Chapman S. Teaching students to generate questions: a review of the intervention studies. *Rev Educ Res*. 1996;66(2):181-221.
47. Rhind SM, Pettigrew GW. Peer generation of multiple-choice questions: student engagement and experiences. *J Vet Med Educ*. 2012;39(4):375-379.
48. Poot R, De Kleijn RA, Van Rijen HV, van Tartwijk J. Students generate items for an online formative assessment: is it motivating? *Med Teach*. 2017;39(3):315-320.
49. Bates SP, Galloway RK, Riise J, et al. Assessing the quality of a student-generated question repository. *Phys Rev Spec Top—Phys Edu Res*. 2014;10(2):020105.
50. Kurtz JB, Lourie MA, Holman EE, et al. Creating assessments as an active learning strategy: what are students' perceptions? A mixed methods study. *Med Educ Online*. 2019;24(1):1630239.
51. Shah MP, Lin BR, Lee M, Kahn D, Hernandez E. Student-written multiple-choice questions—a practical and educational approach. *Med Sci Educ*. 2019;29(1):41-43.
52. Moreno R. Decreasing cognitive load for novice students: effects of explanatory versus corrective feedback in discovery-based multimedia. *Instr Sci*. 2004;32(1):99-113.
53. Nelson MM, Schunn CD. The nature of feedback: how different types of peer feedback affect writing performance. *Instr Sci*. 2009;37(4):375-401.
54. Bates SP, Galloway RK, McBride KL. Student-generated content: using PeerWise to enhance engagement and outcomes in introductory physics courses. In: *AIP Conference Proceedings*. American Institute of Physics, 2012;1413(1):123–126.
55. Jobs A, Twesten C, Göbel A, et al. Question-writing as a learning tool for students—outcomes from curricular exams. *BMC Med Educ*. 2013;13(1):89.
56. Tackett S, Raymond M, Desai R, et al. Crowdsourcing for assessment items to support adaptive learning. *Med Teach*. 2018;40(8):838-841.
57. Gooi ACC, Sommerfeld CS. Medical school 2.0: how we developed a student-generated question bank using small group learning. *Med Teach*. 2015;37(10):892-896.
58. Cho KK, Marjadi B, Langendyk V, et al. The self-regulated learning of medical students in the clinical environment—a scoping review. *BMC Med Educ*. 2017;17(1):112.

How to cite this article: Guilding C, Pye RE, Butler S, Atkinson M, Field E. Answering questions in a co-created formative exam question bank improves summative exam performance, while students perceive benefits from answering, authoring, and peer discussion: A mixed methods analysis of PeerWise. *Pharmacol Res Perspect*. 2021;9:e00833. https://doi.org/10.1002/prp2.833