Formative student-authored question bank: perceptions, question quality and association with summative performance

Jason L Walsh,1 Benjamin H L Harris,1 Paul Denny,2 Phil Smith1

ABSTRACT

Purpose of the study There are few studies on the value of authoring questions as a study method, the quality of the questions produced by students and student perceptions of student-authored question banks. Here we evaluate PeerWise, a widely used and free online resource that allows students to author, answer and discuss multiple-choice questions.

Study design We introduced two undergraduate medical student cohorts to PeerWise (n=603). We looked at their patterns of PeerWise usage; identified associations between student engagement and summative exam performance; and used focus groups to assess student perceptions of the value of PeerWise for learning. We undertook item analysis to assess question difficulty and quality.

Results Over two academic years, the two cohorts wrote 4671 questions, answered questions 606,658 times and posted 7735 comments. Question writing frequency correlated most strongly with summative performance (Spearman’s rank: 0.24, p<0.001). Student focus groups found that: (1) students valued curriculum specificity; and (2) students were concerned about student-authored question quality. Only two questions of the 300 ‘most-answered’ questions analysed had an unacceptable discriminatory value (point-biserial correlation <0.2).

Conclusions Item analysis suggested acceptable question quality despite student concerns. Quantitative and qualitative methods indicated that PeerWise is a valuable study tool.

INTRODUCTION

Multiple-choice questions (MCQs) are widely used to assess medical student knowledge, resulting in a demand for formative questions from students. However, faculty members rarely have time or incentives to develop formative questions and instead focus primarily on developing material for high-stakes assessments. Student demand for formative MCQs is reflected by the growing use of commercial question databases among medical students.1

A potential solution is to involve students in creating formative questions. A few small-scale approaches have involved medical students in question writing to produce banks of formative questions, with the assumption that the question writing itself is a valuable learning activity.2–4 PeerWise is a freely and globally available online platform that allows students to write, share, answer, rate and discuss MCQs pertinent to their course. It is a non-commercial product created and maintained by the University of Auckland, New Zealand.

We introduced PeerWise to Cardiff University School of Medicine in October 2013 to first-year medical students (2013–2014, year 1; n=297). Examination of its usage data over the first 6 months suggested it was a popular resource.5 These students continued to use PeerWise during their second year (2014–2015, year 2; n=273).

Subsequently, in October 2014, we introduced PeerWise to the new cohort of first-year students (2014–2015, year 1; n=306). A separate PeerWise course was created for each academic year, and each course was only accessible to students within that year group.

There has been no formal evaluation of the use of PeerWise within medicine. Here, we describe the introduction of PeerWise to medical students; present descriptive statistics on its usage; examine if there are associations between question writing, answering and commenting frequency with summative exam performance; and gauge student perceptions of the value of PeerWise, using focus groups and subsequent thematic analysis. We assessed the quality of questions using item analysis.

METHODS

We obtained ethical approval for the project from Cardiff School of Medicine Research Ethics Committee.

Introduction of PeerWise

We delivered a 1-hour session to introduce PeerWise to the entire cohort of first-year Cardiff medical students in 2013 (2013–2014, year 1, n=297). All students were asked to attend with an internet-connected device (eg, smart phone, tablet or laptop).

The session began with a brief 10 min description of PeerWise. We then asked all students to access the PeerWise website and helped them to register on to a PeerWise course that we had previously created. Next, we asked students to write one question each. After allowing approximately 20 min, students were asked to answer, rate and if appropriate comment on the questions written by their peers (20 min). Facilitators circulated offering technical support and question-writing advice. Students were subsequently free to use PeerWise at their discretion. We repeated this introductory session in the following year to the new cohort of first-year students (2014–2015, year 1, n=306).
Faculty input
In the inaugural year, two staff members (clinical academics) administered the course. Principally, they responded to emails related to technical difficulties. In the following year, as the popularity of the resource increased, two medical school academics volunteered to give feedback on the questions for their specific specialty (immunology and biochemistry). This involved reading and commenting on student written questions, specifically commenting on question accuracy, relevance and whether the difficulty was appropriate for the course.

Descriptive statistics of usage
PeerWise automatically collects data on user activity. For both cohorts, we examined:
- number of student-written questions
- number of answers to questions
- number of student comments on questions
- number of students writing questions
- temporal relationship of writing and answering questions in relation to summative examination results

We studied usage data from two cohorts, following one cohort over two academic years (2013–2014, year 1; 2014–2015, year 2) and one cohort over one academic year (2014–2015, year 1).

Associations of PeerWise activity with summative examination performance
The main aspects of PeerWise activity are question writing, answering and commenting. We recorded the frequency of these three activities for each student over each academic year and correlated the frequency of each activity with summative exam performance. At the end of each academic year, students sat two summative examinations. The mean raw score over these two assessments was converted to a percentage and correlated with question writing, answering and commenting frequency by Spearman’s rank correlation coefficient. We excluded from the respective correlation calculation those students who did not engage with question writing, answering and/or commenting following the introductory sessions.

Additionally, for all academic years, we divided students into categories determined by their level of usage (table 1). Categories were devised after consulting usage data and discussion with students and faculty. We compared the summative performance of students in these groups using one-way analysis of variance (ANOVA) and subsequent independent t-tests. We also compared the summative performance of PeerWise users versus non-users across all 3 years (t-tests).

Item analysis
We examined the 100 most answered questions in each cohort, looking at the discriminatory ability of each question, measured using the Pearson point-biserial correlation (r-pbis), and the difficulty of each question, calculating a ‘P value’. The 100 most answered questions were sampled for analysis as most students had attempted these questions in every cohort. The analyses were carried out using Iteman software (V.4.3. 2013; Assessment Systems, Woodbury, Minnesota, USA). Where students answered an item more than once, we used only the first attempt in the analysis.

Discrimination measure
The Pearson r-pbis was used as a measure of discriminatory ability for each of the 100 most answered questions in each of the three academic years (2013–2014, year 1; 2014–2015, year 2; and 2014–2015, year 1). It is the correlation between item scores and total scores on all questions in the set. The r-pbis can range between – 1.0 and 1.0; the higher the r-pbis, the better the item is discriminating between students; it is typically desired that the r-pbis be as high as possible. Locally (at Cardiff University School of Medicine), an r-pbis of >0.2 is the threshold for which questions are considered appropriately discriminatory to be used/reused in summative medical school examinations.

Question difficulty
To measure the difficulty of each item, we calculated a p value ranging from 0 to 1, representing the proportion of examinees answering correctly. A p value of 1 indicates that all candidates answered the question correctly, and a value of 0 indicates that no candidates answered correctly. Very high or very low values might indicate that an item was too easy or too hard.

Student perceptions
Preliminary usage data indicated that PeerWise is popular. However, these data do not explain the reasons for its popularity or if students perceive it as valuable for learning. We conducted four focus groups to gather student perceptions on the value of PeerWise.

Focus groups and thematic analysis
In order to recruit participants, we sent a circular email to each of the two cohorts. The email invited volunteers to attend focus groups, including those who did not use the resource often. We asked volunteers to reply with their availability and to indicate if they use PeerWise rarely, sometimes, often or very often.

Four semistructured focus groups were held with 3–10 participants in each group. Before focus groups commenced, the purpose of the study was explained, and students were informed about measures to maximise confidentiality and their right to withdraw.

Thematic analysis was used to analyse focus group data, as described by Braun and Clarke.

RESULTS

Descriptive statistics of usage
The high usage of PeerWise was notable. The two cohorts produced a bank of 4671 questions, answered questions 606658 times and posted 7735 comments discussing questions (table 2). Spikes in question writing and answering activity invariably coincided with exam periods (figure 1).

The maximum number of questions written by a single student over a single academic year was 297. In the year groups, 55% (2013–2014, year 1), 40% (2014–2015, year 2) and 57% (2014–2015, year 1) of students that used PeerWise wrote at least one question outside of the introductory sessions. Approximately 20% of students authored 90% of questions across all
year groups. The absolute number of students writing questions in the 2013 cohort dropped from first to second year by 33% (table 2). However, activity on PeerWise increased overall for this cohort. Question writing, answering and commenting activity increased by 32%, 13% and 44%, respectively (table 2).

In all cohorts, there was a clear increase in both question writing and answering activity coinciding with the period of 1–2 weeks immediately before summative examinations. There were also smaller spikes in activity before formative examinations. Figure 1 illustrates this effect.

**Associations of PeerWise activity with summative examination performance**

Mean raw scores over two summative assessments (taken at the end of each academic year) were converted to percentages and correlated with question writing, answering and commenting frequency by Spearman’s Rank correlation coefficient. There were significant correlations between writing, answering and commenting frequency with summative examination performance (p<0.001, R=0.24, 0.13 and 0.15, respectively).

Comparison of summative performance between PeerWise users and non-users showed that users performed significantly better (p<0.001; figure 2A).

The summative performance of students in the different writing, answering and commenting frequency groups (table 1) was compared. One-way ANOVA showed that there were significant differences in mean summative examination performance between the writing, answering and commenting frequency groups (p<0.0001). Independent t-tests were subsequently performed.

For question writing, mean summative score increased as question writing frequency increased. There was a significant difference between the mean summative scores of all frequency groups (p<0.05), except between frequent and prolific writer groups. Figure 2B illustrates this trend.

For question answering, the mean summative score of non-users was significantly lower than all other groups (p<0.05). Prolific answerers scored significantly higher than all other groups (p<0.0001). There were no significant differences between the mean summative scores of the rare, occasional and frequent question answering frequency groups (figure 2C).

For question commenting, mean summative score increased as commenting frequency increased. The differences between mean summative scores were significant between all groups (p<0.05), except between occasional and frequent commenters (figure 2D).

**Item analysis**

**Discrimination marker**

The mean r-pbis for the top 100 most answered questions for each academic year were: 0.485 (2013–2014, year 1), 0.446 (2014–2015, year 2) and 0.480 (2014–2015, year 1). The year 2 questions were significantly less discriminatory than those questions generated by year 1 students (p<0.05).

**Question difficulty**

The mean difficulty (p value) in the three groups was 0.370 (2013–2014, year 1) 0.438 (2014–2015, year 2) and 0.362 (2014–2015, year 1). The year 2 questions were significantly easier compared with both year 1 academic years (p=0.001).

Two questions out of the 300 questions analysed had an r-pbis of <0.20. All questions analysed in the year 2 2014–2015 cohort have an r-pbis of >0.20.

**Student perceptions**

**Focus groups and thematic analysis**

Four semistructured focus groups were held gauging student perceptions of PeerWise, containing a total of 23 participants. Focus group duration ranged from 44 to 62 min. Table 3 shows the composition of the focus groups. Two, five, eight and eight of the students reported to using PeerWise rarely, sometimes, often and very often, respectively.

Thematic analysis of focus group transcripts generated 25 initial codes, which were refined into 16 key themes (figure 3).
DISCUSSION
We took a mixed methods approach to evaluate the use of PeerWise at a UK medical school. We looked for associations between PeerWise engagement and summative examination performance and undertook item analysis to investigate student-authored question quality. In addition, we used focus groups to gauge student perceptions of the resource.

The usage data showed that question writing and answering on PeerWise increased prior to formative and summative exams: this was reflected in the focus groups. Students often reported using PeerWise more frequently when closer to exams:

I tend to do a lot of questions during the exam period.—Year 1 student

This could suggest students find PeerWise most useful in the period when they are seeking to reinforce their knowledge.

The use of answering questions for learning is supported by a large body of evidence, suggesting that repeated retrieval practice (testing) is effective for enhancing learning.7–10 This finding might also suggest that there is value in increasing assessment frequency at medical schools to drive learning.11 Surprisingly, question writing frequency also increased around examinations. This suggests a proportion of students found writing questions a worthwhile revision technique, despite the time commitment:

Writing questions is a great way to learn things.—Year 1 student

There were weak but significant correlations between writing, answering and commenting frequency with summative performance. Question writing frequency showed the strongest correlation. This trend was also reflected in the stepwise increase in mean summative score between subsequent question writing groups (figure 2B). In line with the focus group data, this may suggest that question writing is a valuable study method. This supports the emerging literature advocating question writing for learning.12–14 However, in this study, it is difficult to pick apart the impact of question writing on summative exam score from other confounders, for instance question writing frequency could be a marker of student work ethic. Similar to writing frequency, there was a stepwise increase in mean summative exam score with increasing commenting frequency (figure 2C). This may suggest that online discussion of questions supports

### Table 3 Focus group demographics

<table>
<thead>
<tr>
<th>Focus group</th>
<th>Focus group 1 (n=3)</th>
<th>Focus group 2 (n=4)</th>
<th>Focus group 3 (n=6)</th>
<th>Focus group 4 (n=10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Year 1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Year 2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
learning, but again, there may be other potential confounders, such as commenting being a marker of conscientiousness and knowledge. Students frequently reported that they often found discussions on PeerWise informative:

I think you learn more from the discussions than the question sometimes.—Year 2 student

Answering frequency demonstrated the weakest (but significant) correlation with summative performance. Interestingly, there was no stepwise improvement of summative exam score between the rare, occasional and frequent answerer groups. However, prolific question answerers did do better than all other groups. This may suggest there is a threshold effect of answering questions on examination performance. Answering was by far the most common activity on PeerWise, with students reporting it to be particularly useful for reinforcing knowledge and identifying knowledge gaps.

The joint most prevalent theme arising during focus groups was the curriculum specificity of PeerWise content. Students frequently indicated that the questions on PeerWise were relevant to their course and that this was a very positive feature:

Questions are written by people that are in the same [teaching] sessions and they know what is relevant.—Year 1 student

One way this specificity appeared to be manifested is that questions on PeerWise tended to resemble or predict questions in summative assessments:

[in the recent summative exam] there were a number of questions I thought, I have literally answered this on PeerWise.—Year 2 student

This curriculum specificity of PeerWise was also frequently cited as being an advantage over commercially available online question banks aimed at medical students.

We identified two major themes relating to question quality. These were (1) faculty question review is highly valued (joint most prevalent theme) and (2) students had concerns over question quality, for example:

on PeerWise the questions are written by students, so you can’t always trust the answer.—Year 2 student

Students felt strongly that faculty input helped to ensure that questions were relevant (curriculum specific) and factually accurate. However, in practice, the proportion of questions reviewed by faculty was relatively small (<5%). Despite concerns around question quality, item analysis of the top 100 questions from each year indicated that most of the student-authored questions had adequate discriminatory ability and appropriate difficulty for inclusion into local summative examinations. This may indicate that the most answered questions are of high quality. One could posit, although highly valued, faculty review (of the most popular questions at least) may not be necessary. However, a high r-pbis and appropriate difficulty do not necessarily mean the questions are well structured. Further subjective item analysis may be appropriate to assess student-authored question quality. Perhaps incorporating a formal review process of questions and/or question writing training may improve subjective and objective question quality.3 15 16 In addition, the top 100 questions from each year may not represent all 4671 questions available. However, this finding does raise the possibility that incorporating student-authored questions into summative exams may be appropriate.

The fourth most prevalent theme identified was that students felt using PeerWise was a fun/enjoyable experience. This was attributed to the interactivity of PeerWise:

It’s nicer than just going through a textbook as it’s more interactive.
In conclusion, PeerWise is well used and well received by medical students. Some interesting observations arose including: (1) engaging with question writing and a higher frequency of question writing is associated with better summative performance; (2) answering questions was by far the most popular activity on PeerWise, with students invariably reporting that they found it useful for learning. However, the association with answering frequency and summative performance is less clear cut. (3) Commenting frequency was weakly associated with better summative performance, students often finding discussion of questions motivational and informative. However, trolling on PeerWise was identified as a negative aspect of the commenting function. (4) Item analysis indicated acceptable question quality (of the most popular questions) despite student concern; and (5) students valued the curriculum specificity of the generated questions, faculty review, virtual rewards and overall found PeerWise to be an enjoyable study tool. This evaluation justifies the use of student-authored question banks at medical schools.

Acknowledgements We thank Professor David Harris, Dr Lee Coombes and Dr Saadia Tayyaba for their support and advice on the project, and Dr James Matthews for his input as an expert reviewer of student questions and for contributing an undergraduate prize to incentivise student PeerWise engagement in Cardiff.

Contributors All named authors have contributed the following: JWL: contributed substantially to the conception and design of the study, acquisition, analysis and interpretation of the data; wrote the first draft and revised and critically reviewed all subsequent drafts; gives final approval for the submitted version to be published; and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. BHLH contributed to the conception and design of the study, acquisition, analysis and interpretation of the data; covrote the third draft and revised and critically reviewed all subsequent drafts; gives final approval for the submitted version to be published; and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. PD: contributed to the conception and design of the study, acquisition, analysis and interpretation of the data; wrote the first draft and revised and critically reviewed all subsequent drafts; gives final approval for the submitted version to be published; and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests None declared.

Ethics approval Cardiff School of Medicine Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which
permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES
Formative student-authored question bank: perceptions, question quality and association with summative performance

Jason L Walsh, Benjamin H L Harris, Paul Denny and Phil Smith

Postgrad Med J published online September 2, 2017

Updated information and services can be found at:
http://pmj.bmj.com/content/early/2017/09/01/postgradmedj-2017-135018

These include:

References
This article cites 23 articles, 0 of which you can access for free at:
http://pmj.bmj.com/content/early/2017/09/01/postgradmedj-2017-135018#BIBL

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Open access (46)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/