Title: A discursive question: Supporting student-authored multiple-choice questions through peer-learning software in non-STEMM disciplines

Authors: Stephen Snow[1a], Adriana Wilde[1,2], Paul Denny[3], m.c. schraefel[1]

1. *s.snow@soton.ac.uk, agw106@soton.ac.uk, mc@ecs.soton.ac.uk*. Electronics and Computer Science, University of Southampton, SO17 1BJ, United Kingdom

2. *agw5@st-andrews.ac.uk*, School of Computer Science, University of St Andrews, Jack Cole Building, North Haugh, St Andrews KY16 9SX, United Kingdom

3. *paul@cs.auckland.ac.nz*, Department of Computer Science, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

a. Corresponding author

Abstract

Peer-learning that engages students in multiple-choice question (MCQ) formulation promotes higher task engagement and deeper learning than simply answering MCQ's in summative assessment. Yet presently, the literature detailing deployments of student-authored MCQ software is biased towards accounts from Science, Technology, Engineering, Maths and Medicine (STEMM) subjects, rather than discursive subjects or disciplines where content may contain fewer absolute facts and objective metrics and more nuance. We report on qualitative and quantitative findings from a semester-long deployment of a peer-learning software package (PeerWise) in a 140-student course on Interaction Design. PeerWise enables students to author, rate and comment upon their peers' MCQ questions. The platform was enthusiastically adopted as a revision aid, yet overall question quality was poor and students expressed difficulty in translating the discursive nature of the course content into MCQs with only one correct answer. In addressing these shortcomings, this paper offers specific recommendations to instructors of more discursive subjects using student-led MCQ authoring platforms, and further, how platforms such as PeerWise may be adapted to better suit disciplines characterised by discursive content. We propose alternative approaches to moderation and two suggestions for potential amendments to the software itself.

Introduction

The multiple-choice question (MCQ) format is a ubiquitous form of assessment in higher education. The use of MCQ in assessment has increased over time, due to larger student cohorts (with often diminishing resources), and due to developments in computer-assisted assessment and the ease of implementing MCQs in online settings (New, Clarke, Lindsay and McKenna 2004, Nicol 2007). Yet the use of MCQs is also widely criticised, for promoting rote learning, fact memorisation and guessing rather than higher-order skills (McCoubrie 2004). MCQs are used less and considered less suitable in discursive subjects (McKenna 2001, Heron and Lerpiniere 2013) such as arts and humanities compared to their widespread use in Science, Technology, Engineering, Maths and Medicine (STEMM) subjects. Yet literature also suggests that the MCQ format can effectively nurture critical thinking and deeper learning (Brady 2005, Draper 2009, Nicol 2007). Achieving these higher-order learning outcomes, however, is contingent upon: (a) careful MCQ question design requiring logical thinking to answer correctly (Brady 2005), (b) using MCQs as an introduction to a topic rather than only as summative exam questions (Draper 2009), or (c) if students take part in the construction of MCQs (Draper 2009, Nicol 2007).

Peer-learning technologies have applied MCQs as a discussion point (rather than summative assessment only) by using Audience Response Systems (ARS) enabling students to vote on questions posed in class (Hunsu, Adesope and Bayly 2016) and by facilitating student-led question construction, using online educational software such as PeerWise (Duret, Christley, Denny and Senior 2018, Howe, McKague, Lodge, Blunden and Saw 2018, Rhodes 2013). PeerWise has been credited with desirable cognitive outcomes (Bates, Galloway and McBride 2012, Denny, McDonald, Empson, Kelly, and Petersen 2018, Hancock, Hare, Denny and Denyer 2018, Howe *et al.* 2018), increased engagement in course content (Biggins, Crowley, Bolat, Dupac, and Dogan 2015, Denny 2013) and positive student feedback (Mac Raighne, Casey, Howard and Ryan 2015). Yet similar to the usage of MCQs in tertiary education more broadly, the current literature PeerWise and peer-learning applications of MCQs more generally, remains biased towards experiences from STEMM-related subjects (Bates *et al.* 2012, Biggins *et al.* 2015, Bottomley and Denny 2011, Denny 2013, Kay, Hardy and Galloway 2018, Mac Raighne *et al.* 2015, Rea and McClure 2012) rather than discursive subjects (Humpage 2014, Renzo 2014). STEMM subjects are typically characterised by more exam-based assessment and problems for which a given solution or set of solutions exist, compared to more discursive subjects where course content may be characterised by a higher level of interpretation and nuance (Humpage 2014). Few peer-reviewed case studies document practical experience from deployments of peer-learning technology utilising MCQ in discursive subjects or disciplines (Humpage 2014, Renzo 2014). Consequently, there is insufficient evidence as to whether the exhibited benefits for peer-learning applications of MCQs are likely to be replicable in more discursive subjects and how instructors of such subjects may be best supported in using such tools.

In this paper, we present an analysis and discussion of our experience in a tertiary course on Human-Computer Interaction (HCI), where we required our students to author their own MCQs using PeerWise. Through analysis of quantitative data on student participation metrics and qualitative reflections of students' personal experiences with the PeerWise platform, this paper aims to: (1) explore the facilitators and barriers to student-authored MCQ software deployed in discursive subjects realising the potential benefits reported from deployments in STEMM subjects- e.g. task engagement and student satisfaction (Biggins *et al.* 2015, Denny *et al.* 2015, Mac Raighne *et al.* 2015, Tatachar and Kominski 2017); (2) understand how to support instructors of discursive subjects in best realising the potential of such software; and (3) propose design suggestions for how student-authored MCQ software may better meet the needs of discursive subjects.

Background

PeerWise is an online peer-learning platform which allows students (and facilitators) to create and publish their own MCQs which may contain text and/or images. Each question must have one correct answer and up to four additional distractors (PeerWise, n.d.). Students may additionally answer, comment upon, and award a star-rating (up to 5 stars) to the quality of other students' questions. Being continuously accessible online, the platform extends conversations on course content beyond the classroom. Over time, PeerWise becomes a repository of student-generated questions which can be used as a revision aid (PeerWise, n.d.). The premise underpinning the software is that through creating questions, students become more deeply involved in the content as co-creators of it, and develop a deeper understanding than answering questions alone (Denny 2013). PeerWise is currently the most widely used online platform for student-authored MCQs, with a user-base of over 2,000 academic institutions worldwide and hosting more than three and a half million questions (PeerWise, n.d.). The software has been widely reviewed in educational literature (Bates *et al.* 2012, Biggins *et al.* 2015, Denny *et al.* 2008, Denny 2013, Humpage 2014, Renzo 2014, Rhodes 2013) and is credited with generating increased student engagement in course content (Denny 2013), improved exam results (Denny, Tempero, Garbett and Petersen 2017, Howe *et al.* 2018) and receiving positive reviews by students (Renzo 2014).

Published critiques on PeerWise are biased toward experiences from STEMM subjects e.g. physics (Bates *et al.* 2012), astronomy, food science, environmental health (Mac Raighne *et al.* 2015), project management (Biggins *et al.* 2015), medicine (Rea and McClure 2012), biochemistry (Bottomley and Denny 2011, Hancock et al. 2018), nursing (Rhodes 2013) and computing (Devon, Paterson, Moffat and McCraw 2012). In contrast, we have only found two published experiences of PeerWise deployed in discursive subjects, specifically sociology (Humpage 2014) and popular music (Renzo 2014). Humpage (2014) argues that due to the specific focus on MCQs, PeerWise may not be suitable for disciplines such as sociology in which course content may be more nuanced and subject to interpretation than STEMM course content. Renzo (2014) on the other hand reports PeerWise raised useful questions that were not emphasised in the lectures, and that the software was valuable as a building block toward more substantial learning outcomes (Renzo 2014). All existing appraisals of PeerWise to date gather only quantitative data on participation and performance (Rea and McClure 2012, Denny *et al.* 2017), and rely only on closed survey questions to assess student attitudes (Rhodes 2013, Humpage 2014), rather than more in-depth personal reflections.

In contrast, this paper draws from written reflections elicited from students using PeerWise, providing rich written descriptions of their experiences with the software and our specific implementation of it. These qualitative accounts are triangulated against quantitative participation data gathered from the PeerWise software. We conclude with a discussion of the suitability of PeerWise to leverage the MCQ format in discursive non-STEMM subjects. We propose: (1) two key considerations for instructors wishing to leverage student-authored MCQ software in discursive subjects and support students in these contexts, as well as (2) two further considerations for the design of such software itself, toward maximising its applicability outside of traditional STEMM applications.

Method

PeerWise was implemented in the 12-week Interaction Design course (COMP2213) at the University of Southampton in 2016. Interaction Design is an introductory course to Human-Computer Interaction with an emphasis on design theory, qualitative methods and elements of cognitive psychology. The course was taken by 140 second year undergraduate students and included 40% assignment-based assessment, 10% PeerWise-related activities and a 50% final exam which included 22 MCQs, short and longer-answer questions. Students were organised into 27 groups of 3-4 students for the assignment-based assessment. In PeerWise, authoring, answering and commenting upon questions was done individually, however participation marks were awarded per group. Students were responsible for ensuring their fellow group members contributed to PeerWise, so the full group participation mark could be awarded.

Our implementation of PeerWise was similar to other STEMM-related deployments described in the literature, in terms of: (a) the participation required of students (i.e. a mark was awarded for meeting minimum authoring and answering targets), (b) the fact that summative marks were not awarded for the quality of the questions created, and (c) that the participation mark had a moderately low contribution towards the final course mark (Denny 2013, Devon *et al.* 2012, Sykes, Hamer and Purchase 2018). We required students to author four and answer four questions on PeerWise over the course of the semester (February-May). The total PeerWise component worth 10% of the total course mark had two components: 5% for *participation* in PeerWise and 5% for the submission of one *reflective essay* per-group. To achieve the full participation mark, each member of a group was required to (individually) author at least four questions, and answer at least four questions by the deadline of 26 April (the 11th week of the 12-week term). At least one of these questions and answers was required by a mid-semester deadline on 18 March (week 5). While the quality of the question content was not marked by the instructors, students were warned that any irrelevant or nonsensical questions or any offensive language would result in a zero-mark. Students were encouraged (but not obliged) to comment on the questions they answered. For the written reflection mark (5%), each

group was required to co-author a 1000-word reflection on their experiences with PeerWise. In the reflection, students were asked to include whether/how PeerWise improved their understanding of course content, comment on the strengths and weaknesses of PeerWise and its implementation in COMP2213, and indicate how learnings from their PeerWise experience may apply to other areas of study.

Data gathered

Our primary goal in this work was to understand whether the benefits of student-authored MCQ software in STEMM subjects, such as task engagement and student satisfaction, may be observed in more discursive subjects. Quantitative data on students' participation in PeerWise was collated from the system. This included statistics gathered from PeerWise on the number of questions authored, answered and commented upon per student, the timestamps of each type of interaction with the system and all question ratings and logins to the PeerWise system during the semester. Qualitative data included the students' MCQ questions submitted to PeerWise and the 27 groups' written reflective essays. Ethical approval was granted by the university to collate, analyse and use de-identified quotes from the written reflections. Gathering qualitative student feedback on PeerWise in this manner is novel in the literature to date.

Analysis of qualitative data

Our analysis of the students' written reflections was split into two parts. All complete essays were combined into a single text file (33 pages). Headings were removed, allowing a thematic analysis to be conducted on the complete set of responses without distraction from context in which reflections were made. The Thematic Analysis methodology was informed by Liamputtong and Ezzy (2005). Reflections were read and re-read, with emergent concepts grouped into themes and sub-themes. A second visitation to the analysis by the same researcher resulted in a consolidation of certain themes, leaving five key themes: (1) Suitability, (2) Collaboration, (3) Assumptions, (4) Improvements and (5) Additional Participation Incentives. Secondly, due to the relatively structured requirements of the reflective essay, responses were separately grouped according to the heading under which they were reported (responses per category analysis) in order to understand the most common strengths and weaknesses cited by students. Owing to our inclusion of quantitative data and space constraints, we discuss only those themes and responses per category which relate specifically to student engagement and satisfaction with the PeerWise deployment, specifically Weaknesses from the responses per category analysis and the themes of Collaboration and Assumptions from the thematic analysis.

The content of questions was monitored throughout the semester, allowing insight into the types of questions authored. We also regularly sort the question pool by most-answered, highest-rated etc, however question content was not formally analysed.

Analysis of quantitative data

Descriptive statistics were prepared for the available metrics of students' participation (e.g. number of questions answered or authored, average question ratings, correctly answered questions). Correlations were prepared for (a) number of answers submitted vs number of answers correct, (b) average rating of a question vs number of times that question was answered, and (c) individual PeerWise participation vs performance in the final exam (see below).

Results

In the following sections, we summarise our quantitative analysis of student participation data and describe the types questions generated by students. We then present the results of our qualitative analysis of the students' reflective essays.

Participation in PeerWise

The PeerWise system was open to students throughout the semester, closing only after the final exam on 24 May. The system was kept open beyond the last participation deadline (26 April) such that students could access it for revision if they wished. Of the 140 students in the class, 132 contributed to PeerWise during the semester, authoring a total of 531 separate questions which were collectively answered 8679 times and attracted 312 comments. Despite the PeerWise system being open throughout the semester, more than half of all questions authored were submitted shortly prior to the 18 March (mid-semester) deadline, where each group member was required to have submitted at least one question and one answer (Table 1, Figure 1). A further 167 questions were authored after the first and before the last deadline (Table 1).

Table 1: Questions and answers submitted

	Before first deadline (18 March)	**18 March – 26 April**	**After last deadline (26 April)**
Total number of questions	358	167	6
Students authoring	113	59	2
Average number of questions per student	3.17	2.83	3.00
Total number of answers	1021	1176	6482
Students answering	115	59	65
Average number of answers per student	8.88	19.93	99.72

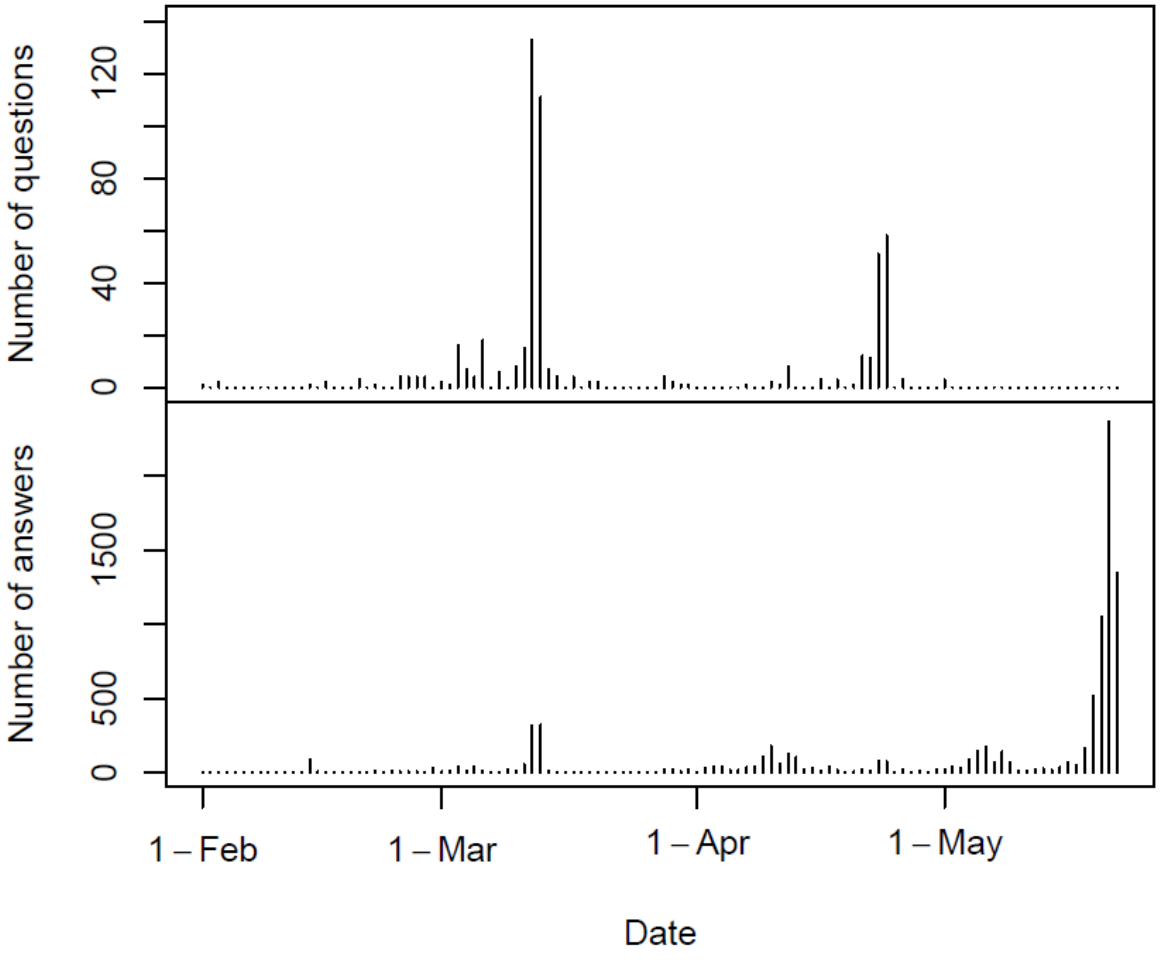

Figure 1: Questions authored and answers submitted over the semester

The temporal distribution of questions authored is in striking contrast to the temporal distribution of the answers submitted, with the majority of answers submitted in the days immediately prior to the final exam (Figure 1). This indicates that PeerWise was used extensively as a revision aid. Because there was no course credit associated with activity after the 26 April deadline, students using PeerWise in this period for exam preparation were doing so voluntarily, which indicates they valued the peer-generated questions for revision and were motivated to use PeerWise well beyond the coursework contribution milestones.

To further understand how students were engaging with PeerWise, we analysed quantitative data relating to answering accuracy to determine if students were taking the task seriously (Figure 2).

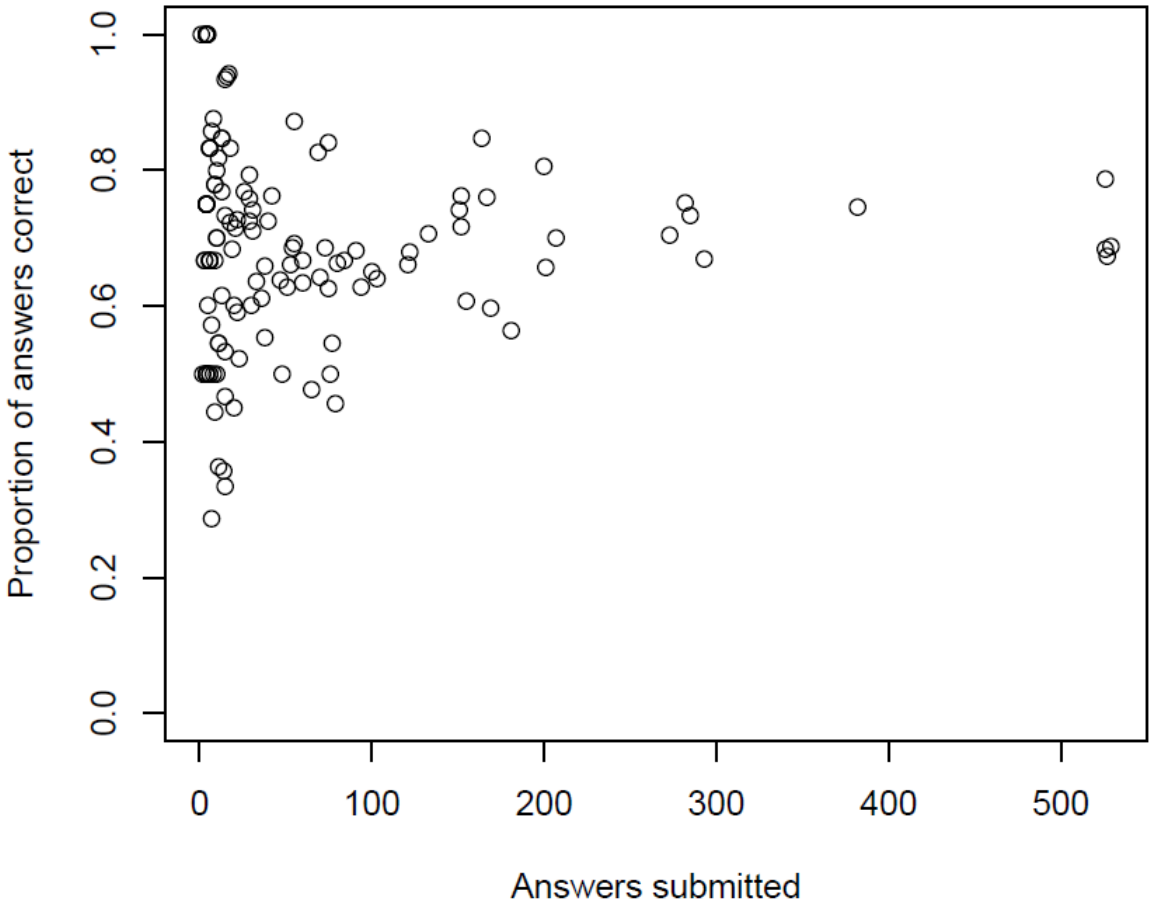

Figure 2: Proportion of correctly answered questions

Figure 2 plots the number of answers submitted by each student against the proportion of the answers that were correct. The average proportion of correctly answered questions across the whole class was 0.69, substantially better than would be expected from random guessing which would fall in the range 0.2-0.25 for questions with 4-5 alternatives. This indicates that the questions were answered seriously and students did not engage in guesswork, even though this practice would have been sufficient to achieve the participation mark for the course.

Question quality

Question quality was not marked by the facilitators. We intended the PeerWise exercise to be peer-led and moderated by the students' ratings of their peers' questions. We did, however, ask students to rate the quality (up to 5 stars) of each question they answered, using the available star-rating feature. The criteria for rating quality was simply what students considered to be an engaging and well-

constructed question with good distractors. We compared the average quality rating of each question with how many times it was answered in order to determine whether (student-assessed) question quality informed revision behaviour (Figure 3). The most frequently answered question was answered 62 times and no question with a rating below 1 was answered more than 15 times.

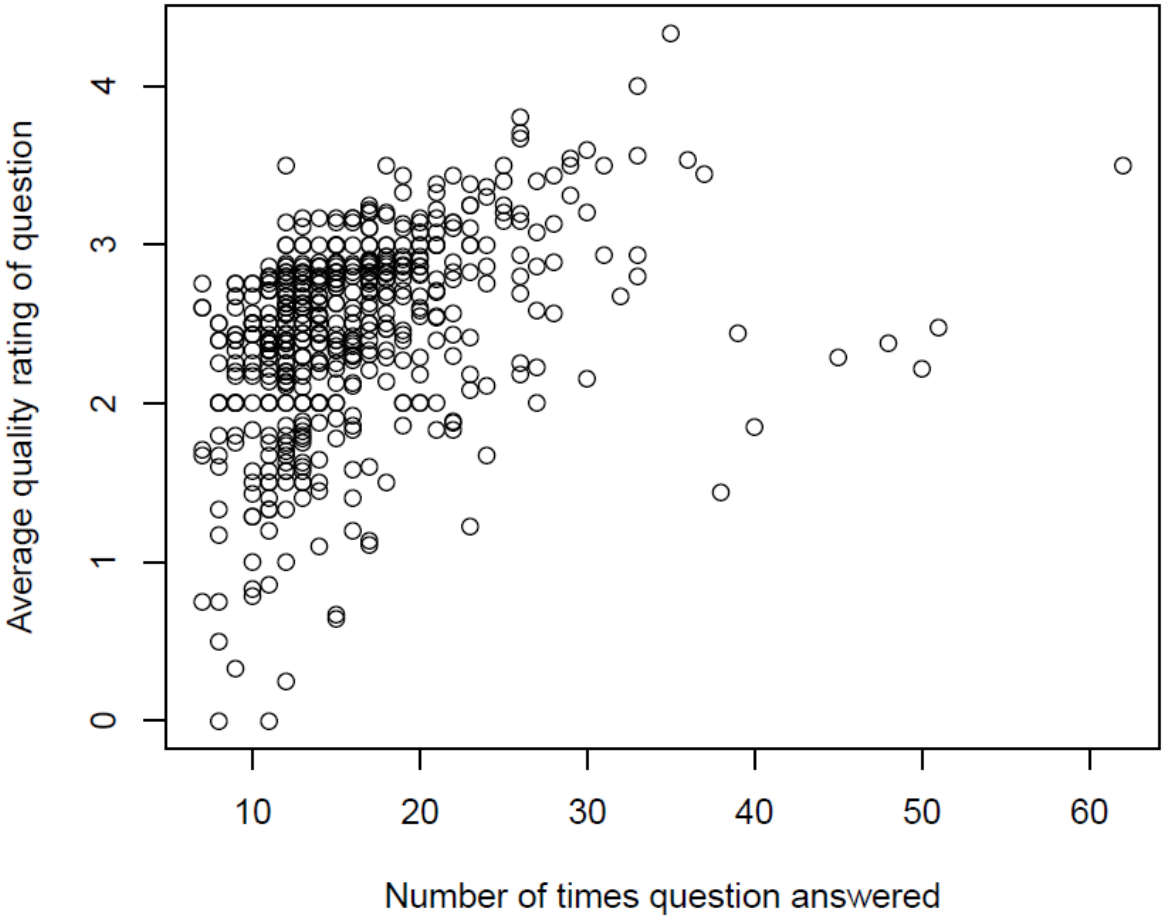

Figure 3: Average quality rating of question versus number of times answered

A weak positive association between the average quality rating of a question and the number of times it was answered is observed (Figure 3). This suggests students used the quality ratings to help them find good questions to answer, a practice which was corroborated in the reflective essays (below).

PeerWise allows students to incorporate visual imagery into authored questions. Two of the three top rated questions incorporated student-generated graphics. One of these was titled The Narcissistic Statistician: *"You recently purchased a new pair of lovely shoes, and you want to find out what other people think. Naturally, you set up a survey. Which one of the following survey questions follows the Likert Scale?"* Four graphical options were presented. Yet despite a number of imaginative and very well designed questions, the majority of questions were definitional in nature, i.e. *"What is the difference between telepresence and co-presence?"* or *"Which of these combinations are the four basic activities of interaction design?"* Repetition was also an issue. Many questions which were technically original, were clearly minimum-effort variations of an existing question.

Overall, students tended towards submitting 'safe' questions based on the easily available facts and definitions available in the textbook, rather than attempting to address the more discursive topics through questions based on positions or arguments.

Correlation to exam performance

Finally, we wished to determine whether engagement with PeerWise was a good predictor of exam performance. To do this we analysed the exam performance of the 54 students who were active in PeerWise within 48 hours of the exam (representing 40% of all students who used PeerWise and who sat the exam) compared to the remaining 77 students who sat the exam but did not participate in PeerWise within the 48 hours (Figure 4).

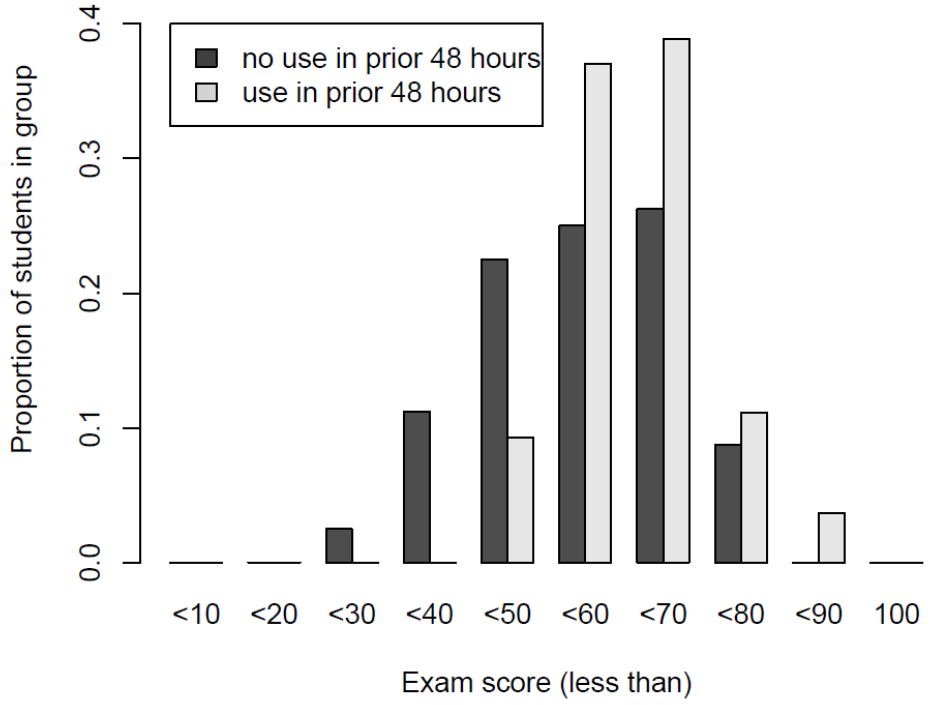

Figure 4- Exam score distributions of students both active and inactive prior to the exam

The average exam score for the 54 active students was 61.2. This compares to an average exam score of 53.3 for all other students in the course that sat the exam. A comparison of ranks (Wilcoxon rank sum test) shows that the exam score distributions of these two groups of students differ significantly ($W = 1323$; $p < 0.001$). It is possible that much of this difference is due to the fact that the more capable students in the course were more likely to engage in practice testing immediately prior to the exam. However, it may also be that the additional practice had an effect on subsequent exam performance. Exploring this effect in more detail, and controlling for student ability, is an avenue for future work.

Reflective Essay Findings

In their reflective essays, students were asked to discuss the strengths and weaknesses of PeerWise, as well as list how the learnings or affordances provided by the software might be applicable to other areas of their study. As mentioned, here we discuss here the themes and categorised responses relating specifically to student engagement and satisfaction with the PeerWise deployment.

Overall, PeerWise was received well by the students. Analysis of the data from the reflective essays revealed 20 of the 27 groups were positive in their conclusions regarding the implementation of PeerWise in the course, highlighting the tool as beneficial to learning, fun, or both. Five groups were overall negative about PeerWise, considering it an unwanted aside, or irrelevant to the course and two were neutral with no strong arguments for or against the software. The peer-rating system was described as very useful for identifying good questions.

Collaboration

In the thematic analysis, collaboration emerged as a key facilitator of engagement with course content and was typically referred to positively. Sixteen of the 27 groups described how aspects of the collaboration offered by PeerWise enhanced their learning. Collaboration emerged both as a positive attribute of the PeerWise system itself and in descriptions of how groups organised themselves with respect to the activity:

> *"…the ability to share, view, and answer the questions created by everyone else on the course gives it the upper hand over a more individualistic approach [to learning subject content]"* (G17)

> *"In order to avoid influencing each other[…]we decided to create and answer questions independently, and then met to discuss our learning experiences together"* (G2)

Two groups specifically highlighted the collaborative nature of PeerWise being helpful in subjects where course content may be less black-and-white: *"the discussion aspect is particularly beneficial in humanities studies wherein students can share different opinions on a concept raised by another student and provide contrasting views"* (G14). Group 4 highlighted that the ability to comment on questions allowed students to propose an alternative answer to a question if they disagreed, describing this feature as *"especially relevant to a field like HCI (Human-Computer Interaction) where it is beneficial to share our ideas to consider all possible points of view"* (G4). On the other hand, another group mentioned how if a question was authored with an incorrect answer: *"…the topic area of Interaction Design can already be quite vague and any confusions could amplify the problem"* (G22).

Weaknesses

When analysing responses under the "weaknesses" heading of the reflections we found responses fell under either weaknesses of the system caused by the users (i.e. user-related weaknesses) or weaknesses inherent in the software itself (software-related weaknesses). We report briefly on each.

User-related weaknesses: The quality of the questions was the most commonly cited negative aspect of the students' PeerWise experience. While the question rating system was described as a useful means of finding good questions to answer, 21 of the 27 groups highlighted poor quality questions as a weakness of their PeerWise experience:

> *"It was obvious that the majority of questions were completed in a rush before the deadline, making the content of the question, answers and explanation minimal and vague" (G27)*

> *"When answering questions other students had posted, we found a tendency for straightforward questions often with a conspicuous answer. As the question poses no higher level thinking skills such as analysis, evaluation or synthesis then its effectiveness is ultimately negligible"* (G9).

Repetition of questions was also reported as an issue. Despite a lecture covering good MCQ formulation and warnings given mid-term in class that a number of questions were *"getting too similar"*, this did not stop the final question pool comprising 10 variants of one question taken from the first chapter of the textbook: *"Which of these combinations are the four basic activities of interaction design?", "Which of these is NOT a basic activity of interaction design?"* etc. Particularly striking is that despite 21 of the 27 groups lamenting the poor quality and repetitive nature of many questions, it is almost inevitable that students from many of these groups would have contributed to the problem themselves. This occurrence is not well discussed in the existing literature on PeerWise (Bates *et al.* 2012, Biggins *et al.* 2015, Denny *et al.* 2008, Denny 2013, Denny *et al.* 2012, Humpage 2014, Renzo 2014, Rhodes 2013).

*Software-related weaknesses***:** Criticisms of the PeerWise software itself were generally minor, e.g. the lack of search functionality, and a "relatively plain interface" (G24). However five groups noted they would have liked PeerWise to allow for more than one correct answer when authoring questions. Two of these groups specifically mentioned the discursive/subjective nature of Interaction Design made the need for a single-correct-answer problematic.

> *"A more realistic alternative is to allow [multiple-answer] questions, where the user selects several correct answers. This would be a simple extension to PeerWise's current format and would allow less objective topics to be assessed in greater depth and allow several questions on the same topic to be condensed into a single one"* (G21)

> *"Providing a [single] 'correct' answer is difficult… Interaction Design is subjective: sub-topics like prototyping incite debate. In essence, [PeerWise requiring] one correct answer is limited; it's possible many answers apply to questions given Interaction Design's variability and subjectivity"* (G6)

These findings raise the question of whether adapting PeerWise to allow Multiple-Answer Questions (MAQs) would potentially increase its applicability and relevance to disciplines characterised by more subjectivity.

Assumptions

The thematic analysis of the reflective essays made visible a number of assumptions that students had made about the PeerWise system that we would not otherwise have realised. The most prevalent of these assumptions are discussed below:

Lack of supervisor involvement: Because of the default anonymity of the system, PeerWise provides lecturers with an option to identify themselves through an "Administrator Comment" function. We elected not to activate this function when commenting due to our wish not to make the students feel they were under surveillance. Perhaps as a result of this decision, the "lack" of moderation emerged as a recurring concern among several of the reflective essays. Many of the 21 groups who flagged question quality as a weakness attributed this problem in varying degrees to the (apparent) lack of moderation of the questions:

> *"Since there was no supervisor involvement, this led students being able to post questions of a poor standard"* (G27)

> *"The system could do with some moderating to ensure all the questions are relevant, so people aren't wasting their time answering questions written by someone who isn't aware what we are examined on"* (G23)

In reality, some moderation did occur in the form of comments upon certain questions from the teaching team, as well as four model questions posted early in the semester by one of the instructors. However, as the lecturers chose not to make their authority conspicuous, this moderation was not acknowledged by many students and students lamented the lack of an authoritative presence. Given our findings, we suggest that lecturers' online presence should be made more visible for PeerWise deployments.

PeerWise is just a revision aid: The PeerWise component of the course was originally explained to the students as an opportunity to improve knowledge of a topic through "learning by teaching", i.e. authoring an exam question themselves. We dedicated lecture time to MCQ generation and "what makes a good question". Yet despite this, it became apparent through the reflective essays that several groups still perceived PeerWise *primarily* as a revision tool:

> *"To help with our revision for exams, we were given the task of creating at least four multiple-choice questions each, and answering just as many"* (G2).

> *"The most beneficial part of all, however, was answering other people's questions…PeerWise is definitely not a complete replacement for a traditional revision, but a great addition to already established revision methods none the less"* (G7)

Perhaps worryingly, two groups voiced their disappointment that PeerWise was *"unreliable as the main way of learning for an important exam"* (G13), and that *"revisions straight out of the source material would be more effective for the amount of time spent"* (G25). These comments suggest these groups envisaged PeerWise as a platform built specifically for exam revision and had (unsurprisingly) found it lacking as a complete replacement for traditional revision.

Discussion

Currently, literature which reports on deployments of the PeerWise software is disproportionately biased towards experiences from STEMM subjects (Biggins *et al.* 2015, Hancock *et al.* 2018, Harris, Walsh, Denny and Smith 2018, Hudson, Jarstfer and Persky 2018, Kay *et al.* 2018, Mac Raighne *et al.* 2015, Rhodes 2013) compared to discursive subjects (Humpage 2014, Renzo 2014). Accordingly, a primary aim of this paper was to assess whether the desirable outcomes associated with implementing student-authored MCQ software such as PeerWise (i.e. engagement with course content and student satisfaction) reported in accounts from STEMM deployments (e.g. Biggins *et al.* 2015, Hancock *et al.* 2018, Mac Raighne *et al.* 2015, Tatachar and Kominski 2017) may be replicable in discursive subjects, where course content is characterised by more nuance and interpretation. Our use of reflective essays to gather qualitative insights from students on their experience of PeerWise is thus far novel in the literature.

Of the two experiences published to date of PeerWise deployed in discursive subjects (Renzo 2015, Humpage 2014), our findings overall reflect those of Renzo (2014), in that students were generally positive of PeerWise as a tool for learning in class in their written reflections. Analysis of quantitative engagement metrics suggests students took their participation seriously and used PeerWise extensively as a revision tool. Yet similar to Humpage (2014), our deployment of PeerWise was characterised by poor question quality which decreased the overall value of the question pool as a learning aid.

Our finding that PeerWise was adopted enthusiastically as a revision tool is in agreement with other literature (Rhodes 2013, McClean 2015, Denny 2013, Renzo 2014), as is our finding that PeerWise participation is positively correlated to exam performance (Denny *et al*. 2017). However, further work comparing PeerWise participation to exam performance is required, as our findings may also be explained by the greater tendency for higher achieving students to engage with PeerWise than lower achieving students in general (Howe *et al.* 2018). PeerWise proved a popular – and seemingly effective – revision aid. Yet insights from the reflective essays found many students assumed that PeerWise was *primarily* concerned with exam revision. This is in contrast with the stated aims of PeerWise to encourage deeper learning by creating (rather than only answering) MCQs (Denny *et al.* 2008) and is not reported by other studies which measure participation through quantitative measures only (Rhodes 2013, McClean 2015, Denny 2013, Renzo 2014), suggestive that quantitative participation data alone is not necessarily a good indicator of meaningful engagement with the system. We encourage future practitioners trialling peer-learning software to additionally gather qualitative data on students' experiences rather than relying on quantitative data alone.

Despite the poor overall question quality and some misplaced assumptions from some students regarding the level of moderation and intended use of the system, on balance, we suggest that PeerWise does have the potential to be an enjoyable learning experience for students and to foster engagement in course content in discursive subjects. Yet this potential appears contingent on a number of factors related to how the software is deployed by course instructors. In the following paragraphs we outline recommendations for how instructors of discursive subjects may best realise

the positive learning outcomes of student-authored MCQ software and two considerations for the design of the software itself to better meet the needs of discursive subjects.

Supporting instructors using student-authored MCQ software in discursive subjects

Care is required to assist students in creating engaging MCQs from discursive material where there may (potentially) be fewer definitions or obvious right/wrong answers for MCQ formulation compared to STEMM subjects (Humpage 2014). Accordingly, we advocate class time spent on question creation and use of the software early in the semester, as we did. Subsequently, moderation of questions is necessary throughout the semester to pick up on inappropriate questions/comments and encourage quality questions. Our own approach to moderating was relatively laisse faire, however on reflection, this could be improved, given many students were not aware of our presence in the system. We argue that a highly-visible moderation of students' including constructive comments on students' questions may valuable in discursive subjects and that this guidance is likely to be welcomed rather than being considered unwanted "surveillance" or a deterrent from creativity as we had initially anticipated. Additional to giving persuasive comments, such as *"No more variations on the 'four key activities' theme – be original!"* we advocate a hands-on approach to moderation in the form of lecturers flagging poor quality or repeated questions with a formal *"Revise and resubmit"* notice (in the form of an "Administrator Comment"). . This more hands-on approach could be reinforced by instructors emphasising early in the semester that the benefit of PeerWise is contingent upon students spending time researching and authoring quality MCQ questions and that the exercise is more than just a revision aid.

A second possible strategy to improve question quality in PeerWise deployments in discursive subjects, might involve increasing students' literacy on what constitutes a high-quality MCQ both in theory and in practice, by providing examples of how MCQ questions can be derived from debate, not only facts or definitions. This could be complemented by instructors "seeding" the question pool. In our study, the instructors added four exemplar questions to the repository before students began using PeerWise. However, this strategy needs further exploration. Renzo (2014) provides an example of instructor-seeded question from popular music studies, highlighting how MCQs may be derived from more discursive material: *"What would Adorno most likely say about Beyoncé's music?"* The timing of this input is important and we argue that "seeding" should take place once students have already made questions out of the simple facts and definitions and some repetition is beginning to occur, additional to at the very beginning of semester (as we had). Further work may explore the effectiveness of this strategy on improving question quality.

Design considerations for MCQ software in discursive subjects

Finally, we provide two potential design adaptations to PeerWise itself, with the aim of increasing the software's relevance to non-STEMM subjects.

Self-rating: We were struck with the irony that many students in the 21 groups who complained of the low question quality in their reflective essays must have contributed to this problem themselves. A potential adaptation to PeerWise towards the aim of increasing overall question quality, is to require students to rate the quality of their own questions before they can be submitted. Taking more ownership over question quality may serve to provoke students to stop and reconsider before submitting a poorly considered or throw-away question. This modification could provide lecturers with useful data regarding individual effort per student and provide grounds to re-iterate in class how the quality of the question pool is only as good as the sum of individual efforts.

Opportunities for multiple correct answers: Finally, some reflective essay responses highlighted that PeerWise's current limitation to questions with one correct answer was problematic, given the discursive nature of Interaction Design. Ambiguity in questions or their distractors was seen as a weakness rather than an opportunity for discussion. A further design consideration PeerWise, to

increase relevance and value to non-STEMM subjects, may be to provide an option for questions with multiple correct answers. This adaptation may potentially allow for MCQ's to be leveraged better as points of discussion rather than right/wrong facts in non-STEMM subjects.

Due to the lack of research on student-authored MCQ exercises in discursive subjects, both these potential adaptations are presented only as suggestions at this point and both would be useful targets for future research.

Conclusion

Previous research shows that involving students as co-creators of revision material through self-authored MCQ generation in STEMM subjects has a great potential for learning beyond memorisation, including engagement in course content and increased student satisfaction. Based on mixed-method research, our findings suggest such benefits may be generalisable to more discursive subjects. Yet the realisation of these benefits depends on careful instructor support and could be assisted by the addition of certain features to the software itself. Based on a mixed-method research, we have formulated some recommendations to both instructors and software designers. This paper contributes to a growing debate around peer-learning software in discursive subjects, which at present is still incipient.

Acknowledgements

The authors thank the COMP2213 class of 2016 for their participation in PeerWise, for their insights into their experience with our deployment of the software and for their consent in allowing us to use the generated data for action research. This research did not receive any funding from agencies in the public, commercial, or not-for-profit sectors.

Statements on open data, ethics and conflict of interest

Students were informed of our intention to publish de-identified excerpts from their reflective essays and were offered opportunities to request their withdrawal. Ethical approval was granted from the University of Southampton use de-identified quotes from the written reflections (approval: ERGO-FPSE-20318). Quantitative data from the platform is stored on the PeerWise server and accessible by request to peerwise@cs.auckland.ac.nz. Other data presented is stored on the authors' password-protected computers and can be made available with a specific request to the authors. For full disclosure, Paul Denny is the creator of the PeerWise software, however this association does not cause a conflict of interest in the research here reported.

Biographies

Stephen Snow is a Teaching Fellow in Computer Science at the University of Southampton with over 20 peer reviewed publications. His research interests include human factors research related to the effect of indoor air quality on human performance.

Adriana Wilde is an Associate Lecturer in Computer Science at the University of St Andrews and a Fellow of the Higher Education Academy (2013). She holds an MSc in Computer Science (with a minor in Education) and is completing a PhD in Computer Science at the University of Southampton. Her research interests are educational technologies, data mining and pervasive computing.

Paul Denny is a Senior Lecturer in Computer Science at the University of Auckland, New Zealand. He is a recipient of New Zealand's National Tertiary Teaching Excellence Award (2009) and the Computing Research and Education Association of Australasia Teaching Award (2010). His interests include developing and evaluating technologies for supporting collaborative learning, and motivating students to engage within online learning environments.

m.c. schraefel is a Professor of Computer Science and Human Performance, Fellow of the British Computer Society, and Research Chair for the Royal Academy of Engineering. Her research focuses on the design information systems to support the brain-body connection for quality of life, including fitness to learn, play and perform, and to understand through these paths how to enhance innovation, creativity and discovery.

References

PeerWise: Ask | Share | Learn (n.d.) Retrieved from http://peerwise.cs.auckland.ac.nz/ (last accessed 05/11/2017).

Bates, S.P., Galloway, R.K. & McBride, K.L. (2012). Student-generated content: Using PeerWise to enhance engagement and outcomes in introductory physics courses. In N.S. Rebello, P.V. Engelhardt & C. Singh (Eds.), AIP Conference Proceedings. 1413(1), 123-126. doi:10.1063/1.3680009

Biggins, D., Crowley, E.J., Bolat, E., Dupac, M. & Dogan, H. (2015). Using PeerWise to improve engagement and learning. In Proceedings of *The European Conference on Education*, 1-5 July 2015, Brighton, UK, 585-602.

Bottomley, S. & Denny, P. (2011). A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochemistry and Molecular Biology Education*, 39(5), 352-361. Doi:10.1002/bmb.20526

Brady, A. M. (2005). Assessment of learning with multiple-choice questions. Nurse Education in Practice, 5(4), 238-242. doi:10.1016/j.nepr.2004.12.005

Denny, P., Luxton-Reilly, A. & Hamer, J. (2008). The PeerWise system of student contributed assessment questions. In *Proceedings of the tenth conference on Australasian* computing education, Volume 78, 69-74.

Denny, P. (2013). The effect of virtual achievements on student engagement. *In Proceedings of the SIGCHI conference on human factors in computing systems*, CHI'13, Paris, France (pp.763-772). ACM. doi:10.1145/2470654.2470763

Denny, P., Tempero, E., Garbett G. & Petersen, A. (2017). Examining a Student-Generated Question Activity Using Random Topic Assignment. In Proc: Innovation and Technology in Computer Science Education (ITiCSE '17). ACM, New York, NY, USA, 146-151. doi:10.1145/3059009.3059033

Denny, P., McDonald, F., Empson, R., Kelly, P. & Petersen, A. (2018). Empirical Support for a Causal Relationship Between Gamification and Learning Outcomes, *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI 2018), Paper No. 311, Montreal, Quebec, Canada, 2018. doi: 10.1145/3173574.3173885

Devon, J., Paterson, J. H., Moffat, D.C. & McCrae, J. (2012). Evaluation of student engagement with peer feedback based on student-generated MCQs. *Innovation in Teaching and Learning in Information and Computer Sciences*, 11(1) 27-37. doi:10.11120/ital.2012.11010027

Draper, S.W. (2009) Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *BJET*, 40(2) 285-293. doi:10.1111/j.1467-8535.2008.00920.x

Duret, D., Christley, R., Denny, P. & Senior, A. (2018) Collaborative learning with PeerWise. *Research in Learning Technology*, Vol. 26, 2018. doi: 10.25304/rlt.v26.1979

Hancock, D., Hare, N., Denny, P. & Denyer, G. (2018) Improving large class performance and engagement through student-generated question banks, *Biochemistry and Molecular Biology Education*. March 2018. doi: 10.1002/bmb.21119

Harris, B., Walsh, J., Denny, P. & Smith, P. (2018) Formative student-authored question bank: perceptions, question quality and association with summative performance, *Postgraduate Medical Journal*, pages 97-103, 94(1108), February 2018. doi: 10.1136/postgradmedj-2017-135018

Heron, G. & Lerpiniere, J. (2013). Re-engineering the multiple choice question exam for social work. *European Journal of Social Work*, 16(4), 521-535. doi:10.1080/13691457.2012.691873

Howe, P., McKague, M., Lodge, J., Blunden, A. & Saw, G. (2018) PeerWise: Evaluating the Effectiveness of a Web-Based Learning Aid in a Second-Year Psychology Subject. *Psychology Learning & Teaching*. March 20, 2018. doi: 10.1177/1475725718764181

Hudson, S., Jarstfer, M. & Persky, A. (2018) Student Learning with Generated and Answered Peer-written Questions, *American Journal of Pharmaceutical Education*, Volume 82, Issue 2, Article 6315, 2018. doi: 10.5688/ajpe6315

Humpage, L. (2014). PeerWise: A useful learning tool for Sociology? *New Zealand Sociology* 29(1), 135. Retrieved from: https://search.proquest.com/docview/1556030288?accountid=13963

Hunsu, N.J., Adesope, O. & Bayly, D.J. (2016). A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Computers & Education, 94*, 102-119. doi:10.1016/j.compedu.2015.11.013

Kay, A., Hardy, J. & Galloway, R. (2018) Learning from peer feedback on student-generated multiple choice questions: Views of introductory physics students, *Physical Review Physics Education Research* 14, 010119, 2018. doi: 10.1103/PhysRevPhysEducRes.14.010119

Liamputtong, P. & Ezzy, D. (2005). Qualitative research methods, Oxford University Press, UK.

Luxton-Reilly, A. (2009). A systematic review of tools that support peer assessment. *Computer Science Education, 19*(4) 209-232. doi:10.1080/08993400903384844

McClean, S. (2015). Implementing PeerWise to engage students in collaborative learning. Perspectives on Pedagogy and Practice, 6, 89-96. Retrieved from: http://uir.ulster.ac.uk/33108/

McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. *Medical teacher* 26(8), 709-712. doi: 10.1080/01421590400013495

McKenna, C. (2001). Introducing computers into the assessment process: what is the impact upon academic practice? Paper presented at the Higher Education Close Up Conference 2, Lancaster University, 16-18 July 2001. Retrieved from: http://www.leeds.ac.uk/educol/documents/00001805.htm

Mac Raighne, A., Casey, M. M., Howard, R. & Ryan, B. (2015). Student attitudes to an online, peer-instruction, revision aid in science education. *Journal of Perspectives in Applied Academic Practice*, 3(1). doi:10.14297/jpaap.v3i1.135

New, S., Clarke, S., Lindsay, K. & McKenna, C. (2004) INQUIRE: A Case Study in Evaluating the Potential of Online NCQ Tests in a Discursive Subject. *Research in Learning Technology*, 12 (3), 249-260. doi:10.1080/0968776042000259564

Nicol, D. (2007) E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education* 31(1), 53–64. doi:10.1080/03098770601167922

Rea, P. & McClure, J. (2012). The use of PeerWise in a student selected component within the medical curriculum led to improvement in examination performance. In *Proceedings of Empowering Student Learners in Higher Education*, Glasgow, Scotland. Retrieved from: http://eprints.gla.ac.uk/83664/

Renzo, A. (2014). Multiple-choice questions in the Humanities: a case study of PeerWise in a first-year Popular Music course. In B. Hegarty, J. McDonald & S.-K. Loke (Eds.), *Rhetoric and Reality: Critical perspectives on educational technology. Proceedings ASCILITE Dunedin 2014* (pp. 560-564). Retrieved from: http://hdl.handle.net/1959.14/329207

Rhodes, J. (2013). Using PeerWise to knowledge build and consolidate knowledge in nursing education. *Southern Institute of Technology Journal of Applied Research (SITJAR)*. Retrieved from: https://www.sit.ac.nz/Portals/0/upload/documents/sitjar/SITJAR%202013-Rhodes.pdf

Sykes, A., Hamer, J. & Purchase, H. (2018) PeerWise as a Vehicle for Uncertainty-resolution to Enhance Student Learning, *Journal of Perspectives in Applied Academic Practice*, Volume 6, Issue 1, pages 55-65, 2018. doi: 10.14297/jpaap.v6i1.317

Tatachar, A. & Kominski, C. (2017) Assessing a traditional case-based application exercise and a student question creation exercise on student performance and perceptions, *Currents in Pharmacy Teaching and Learning*, Volume 9, Issue 4, July 2017. doi: 10.1016/j.cptl.2017.03.007