Article

PeerWise Provides Significant Academic Benefits to Biological Science Students Across Diverse Learning Tasks, But with Minimal Instructor Intervention

H. A. McQueen†,*
C. Shields†
D. J. Finnegan†
J. Higham†
M. W. Simmen‡

From the †Institute of cell Biology, School of Biological Sciences, University of Edinburgh, King’s Buildings, Mayfield Road, Edinburgh EH9 3JR, United Kingdom, ‡School of Biomedical Sciences, University of Edinburgh, George Square, Edinburgh EH8 9XD, United Kingdom

Abstract

We demonstrate that student engagement with PeerWise, an online tool that allows students to author and answer multiple-choice questions (MCQs), is associated with enhanced academic performance across diverse assessment types on a second year Genetics course. Benefits were consistent over three course deliveries, with differential benefits bestowed on groups of different prior ability. A rating scheme, to assess the educational quality of students’ questions, is presented and demonstrates that our students are able intuitively to make such quality assessments, and that the process of authoring high quality questions alone does not explain the academic benefits. We further test the benefits of providing additional PeerWise support and conclude that PeerWise works efficiently with minimal intervention, and can be reliably assessed using automatically generated PeerWise scores. © 2014 by The International Union of Biochemistry and Molecular Biology, 00(0):000–000, 2014.

Keywords: PeerWise; collaborative learning; participatory; peer learning

Introduction

PeerWise is a free online resource that allows students to ask, answer and discuss one another’s MCQs [1–3]. Through using the tool, students become active creators of content for their own learning, and take part in multiway dialogue between class peers via discussion threads following on from questions [4]. As well as earning an overall PeerWise score (based on the number and peer-rated quality of questions, answers and comments submitted), students can also earn “badges” when they attain certain goals (e.g., when they answer a certain number of questions correctly or when they contribute to discussions). This gaming-style feature of PeerWise is intended to encourage healthy competition, and was demonstrated to boost engagement in a study where only one half of a class of >1,000 students were exposed to the badge system [5]. A growing body of literature demonstrates that academic benefits are associated with PeerWise engagement [1, 4, 6–8]. To date it has been unclear whether these benefits relate to single academic components within courses, such as answering of MCQs or problem questions (e.g., by allowing “rehearsal” of such tasks), or whether benefits might be reaped over diverse tasks, which would evidence wider and deeper learning benefits, leading to all-round improvement.

Although the academic benefits of PeerWise are being increasingly recognised, it is also not clear from the literature which PeerWise activities (question setting, answering, or discussion) are associated with the benefits. Effective question setting in PeerWise is expected to require reflective engagement with course material, unlike the information recall that can be associated with answering of MCQs. If PeerWise contributes towards deep learning, it might be expected that the depth of learning would be reflected in the quality of student-generated questions. Evidence for this is conflicting, however. A previous study [4] judged the quality of second year Biochemistry students’ PeerWise questions using a rating system based on the revised Bloom’s Taxonomy [9], categorising questions into five
cognitive domains of remembering, understanding, applying, analyzing, and evaluating or creating. In a class of 107 students, 56% of questions were in the lowest category and none fell into the highest two categories. A similar, albeit more discipline-specific, rating system based on five Bloom-type categories was employed in a study of ~150 first year Physics students [10] and found, in contrast, that only a small percentage (<5%) of questions fell into the lowest category. Most questions (~80%) straddled the higher categories of application and analysis. This effect was preserved over two course deliveries. This marked difference in question quality between studies could be due to the subject matter, or study design, or might stem from the extra support sessions provided to assist the Physics students to create questions of high quality. A separate study [11] found that offering introductory sessions to students prior to using PeerWise had little or no impact on the numbers of questions the students went on to author or answer, but did significantly increase the level of engagement as measured by the extent of commenting.

This study reports on the use of PeerWise by three cohorts, representing three successive iterations of a second year Genetics course at The University of Edinburgh. Associations between PeerWise engagement and overall exam performance were previously demonstrated for one delivery of this course [8]. Here we asked whether using PeerWise improved our students’ abilities only to answer MCQs, or whether benefits extended to other measurable improvements in academic ability on disparate course tasks, and whether they were consistent over all three years. We also gauged the quality of student-generated questions, and asked whether students writing questions of high quality performed better academically, and whether extra introductory sessions on PeerWise impacted on question quality, on overall PeerWise engagement, and on academic performance. Finally, we present a more qualitative description of our students’ PeerWise experience, that led us to suggest the optimised PeerWise implementation strategy presented here.

### Methodology

#### Statistical Analyses

All statistical tests were carried out using IBM SPSS Statistics for Windows, Version 19.0.

#### Course Context of PeerWise Deployment

Genes and Gene Action (hereafter referred to as GGA) is an SQA Level 8 second year, Genetics course at The University of Edinburgh. The course is taught over 11 weeks in the second semester. Assessment is divided between coursework (30%), including a problem question, a data handling test and the PeerWise component, and an exam (70%), comprising an essay (25%), MCQs (25%), and a problem (20%). Data presented in this study are based on three consecutive deliveries of the GGA course (2011, 2012, and 2013). Cohort sizes averaged ~250, many of whom had previously taken a first year Genetics course, Molecules, Genes, and Cells, which is used in this study as a measure of relevant prior ability.

#### Guidance to Students on Using PeerWise

Increasing levels of PeerWise-specific support were offered to the students in each of the three successive iterations of the course. Guidance provided and key indicators of levels of activity displayed by each cohort are summarized in Table 1.

#### Course and Feedback Data

PeerWise scores were generated automatically and collected directly from the PeerWise system, as measures of individual engagement. The PeerWise score is based on the numbers of questions, answers and comments a student contributes, how his/her questions are rated by peers, and how often their contributions are agreed with by their peers. Only data for those students who were PeerWise-active and for whom prior ability scores were available are included in the study.

Course evaluation included focus group discussions and questionnaire responses collected at the end of semester, online and/or during a practical class, and contained targeted questions about students’ experiences of PeerWise. Students also had the opportunity to express their opinions about PeerWise within the questionnaire’s open response section.

#### Question Quality Rating Scheme

An 11-point rating scheme was devised that aimed to measure question quality in an objective and repeatable way. The scheme included recognition of questions achieving defined cognitive learning domains, based on the revised Bloom’s Taxonomy [9]. A subset of 20 randomly selected questions was independently rated by two of the authors, and concordance between raters was measured (by Intraclass Correlation, Model 2, single measures). Any questions where there had been disagreement between markers were revisited and discussed, in order to redefine quality category boundaries. The mark scheme was refined accordingly, and a further set of 20 questions was independently marked by both parties, whereupon inter-rater reliability was found to be very good (Intraclass Correlation Coefficient 90%). A summary of the rubric for categorising questions is shown in Table 2.

A total of 221 questions, spread across the three cohort repositories, were rated. Questions from 26 students attending the optional support sessions, and from 26 students selected for matched ability were included, while all other questions were selected at random.

#### Results

PeerWise Engagement Was Positively and Significantly Associated with Performance on All Components of the Course

Associations between engagement with PeerWise, as measured by PeerWise score, and marks in several components
of the GGA course in three successive years of delivery are shown in Table 3. Our hypothesis was that students would benefit from engaging with PeerWise, and that this would be reflected in enhanced academic performance. First order partial correlations (one-tailed) between PeerWise score and marks for exam MCQ (controlling for prior ability) were performed. As expected, positive significant associations were found for all cohorts. More surprisingly,
Similar effects were also found for exam essay, exam problem question, overall exam and coursework marks, with the exceptions of exam essay marks in 2011 and exam problem question marks in 2012 (Table 3). Significant correlations between PeerWise score and overall course work and exam marks were uncovered in all years of delivery. This confirms and extends the findings of a previous study [8] which noted a positive association between PeerWise engagement and exam marks for GGA in 2012, and shows that such effects are consistent year on year, and importantly across diverse course components.

For the purposes of comparison, marks awarded for another significant course component, the data-handling test, were also compared with overall course performance. These marks were significantly correlated, even after controlling for prior ability, having comparable correlation coefficients to those found for PeerWise (Table 4).

**Benefits of PeerWise Engagement Were Consistently Pronounced for Two Distinct Ability Quartiles**

Students were divided into quartiles, based on their prior relevant learning ability (overall marks on a first year Genetics course). Quartile 1 represented “low” ability, quartile 2 represented “low/intermediate,” quartile 3 represented “intermediate/high” and quartile 4 represented “high” ability students. Within each ability quartile, students were ranked based on their PeerWise scores and then divided into two equally sized groups of highly PeerWise-active (HPA) and low PeerWise-active (LPA) individuals. Final course marks of HPA/LPA students within quartiles were compared. We hypothesised that those students who were more PeerWise-active would perform better on GGA compared with less PeerWise-active students. For every HPA/LPA pair, within each quartile, HPA students attained higher average overall marks for GGA compared with their LPA counterparts (Fig. 1). To test for significance of these differences, independent t-tests (one-tailed) were carried out on each HPA/LPA pair within each quartile (analysis based on that of [1]). Significant differences between HPA and LPA students were discovered in quartile 2 for the 2011 and 2013 cohorts, and for quartile four students in all three cohorts, but not for quartiles 1 or 3.

**Good Agreement Was Seen Between Instructor and Student Ratings of Question Quality**

Strong significant correlations (Pearson’s r, one-tailed) were found between student ratings of questions (submitted to the PeerWise system each time a question is answered) and instructor ratings using the question quality rating scheme (described in the methodology section). Student ratings for all three cohorts (2011, 2012, and 2013) are plotted against mark scheme ratings in Fig. 2 ($r = 0.73; p < 0.001$).

**Students Created Questions of Good Quality, But Question Quality Alone Did Not Predict Performance**

Figure 3 shows the proportion of questions, selected at random from each cohort, that were rated as representative of each of the five Bloom’s Taxonomy categories, as defined

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Aspect of question</th>
<th>Points available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar and clarity of wording</td>
<td>1</td>
</tr>
<tr>
<td>Question should be clearly expressed and easy to follow</td>
<td></td>
</tr>
<tr>
<td>Correctness of author’s selected answer</td>
<td>1</td>
</tr>
<tr>
<td>Unless obviously wrong, the author’s answer was assumed to be correct if agreed with by the majority of 15 answering peers. Below this threshold, correctness was cross-checked</td>
<td></td>
</tr>
<tr>
<td>Feasibility of distractors</td>
<td>1</td>
</tr>
<tr>
<td>Should include only one possible correct answer. Author’s incorrect answers should be feasible enough to test knowledge</td>
<td></td>
</tr>
<tr>
<td>Quality of author’s explanation</td>
<td>3</td>
</tr>
<tr>
<td>0=Explanation missing;</td>
<td></td>
</tr>
<tr>
<td>1=Minimal explanation;</td>
<td></td>
</tr>
<tr>
<td>2=Good explanation that informs why correct answer is right and provides some background;</td>
<td></td>
</tr>
<tr>
<td>3=Thorough explanation that would improve the understanding of someone who answered incorrectly</td>
<td></td>
</tr>
<tr>
<td>Bloom’s Taxonomy rating</td>
<td>5</td>
</tr>
<tr>
<td>1=Factual recall only;</td>
<td></td>
</tr>
<tr>
<td>2=Comprehension e.g. compare/contrast/summarize-type questions;</td>
<td></td>
</tr>
<tr>
<td>3=Application of knowledge;</td>
<td></td>
</tr>
<tr>
<td>4=Analysis of information from a variety of sources e.g. questions requiring calculations/deductive processes;</td>
<td></td>
</tr>
<tr>
<td>5=Evaluation and synthesis of material from various sources</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
</tr>
</tbody>
</table>

This confirms and extends the findings of a previous study [8] which noted a positive association between PeerWise engagement and exam marks for GGA in 2012, and shows that such effects are consistent year on year, and importantly across diverse course components.

For the purposes of comparison, marks awarded for another significant course component, the data-handling test, were also compared with overall course performance. These marks were significantly correlated, even after controlling for prior ability, having comparable correlation coefficients to those found for PeerWise (Table 4).
in Table 2. In each year, one quarter of questions were consistently found in Category 3 or above, while the majority of questions fell into Category 2 (“comprehension”). As can be seen from Fig. 3, the proportion of questions found in each quality category is strikingly similar between cohorts.

Given the strong agreement between instructor and student ratings of question quality (Fig. 2), we considered average student ratings of questions (available for every question answered on PeerWise) to be a reliable measure of question quality. With the hypothesis that students writing questions of higher quality should benefit from this activity, and do better on the course than those writing questions of lower quality, we performed Pearson product-moment correlation (one-tailed) between question quality (as rated by peers) and question authors’ academic performance. Overall marks for the course showed a significant positive correlation with question quality ratings (Table 5). However, once prior ability was controlled for (by partial correlation to remove the effect of prior learning), a significant correlation between question quality and marks was only evident for the 2012 cohort. All significance was lost for the 2011 and 2013 cohorts. This lack of robust correlation between a student’s quality of questions and their overall performance on the course, on consideration of prior learning, was also seen when the same analysis was performed using the smaller subset of instructor rated questions (data not shown).

**Table 3**

<table>
<thead>
<tr>
<th>Academic year</th>
<th>Exam MCQ marks (25%)</th>
<th>Exam problem question marks (20%)</th>
<th>Exam essay marks (25%)</th>
<th>Overall exam marks (70%)</th>
<th>Coursework marks (25%)</th>
<th>Overall course marks (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>0.186**</td>
<td>0.187**</td>
<td>0.052(n.s.)</td>
<td>0.204**</td>
<td>0.156**</td>
<td>0.192**</td>
</tr>
<tr>
<td>2012</td>
<td>0.213**</td>
<td>0.101(n.s.)</td>
<td>0.181**</td>
<td>0.241***</td>
<td>0.231**</td>
<td>0.241***</td>
</tr>
<tr>
<td>2013</td>
<td>0.166**</td>
<td>0.117*</td>
<td>0.143*</td>
<td>0.197**</td>
<td>0.323***</td>
<td>0.195**</td>
</tr>
</tbody>
</table>

The proportion that each item of assessment contributed to the 2013 course is shown in brackets. The 5% attributed to PeerWise activity has been removed from the coursework and overall course values, since PeerWise marks were subtracted from coursework marks and final course marks for the purposes of these analyses. Statistically significant correlations are shown in bold. $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$, not significant (n.s.).

**PeerWise Used as a Simple MCQ Revision Tool Did Not Affect Exam Performance**

In 2012 and 2013 the deadline for assessed PeerWise activity was several weeks prior to the final exam. 5,294 answers were submitted in the period after the deadline in 2012 (cohort size 253), and 6,586 answers were submitted in the equivalent period in 2013 (cohort size 275). Students who had not previously completed the first year Genetics course were included in this dataset, as prior ability had no bearing on this analysis. Although this suggested an average of 22 answers per student during the voluntary revision period each year, on closer examination it was evident that only around 40% of each class had answered any questions after the PeerWise deadline (~60 questions answered per active student). This disparity in behavior between roughly two halves of these cohorts offered a serendipitous opportunity to examine the effects of voluntary use of PeerWise for straightforward drill-and-practice MCQ revision. No questions were authored during this period and very few peer discussions took place.

Since it was conceivable that students could either make academic gains from this activity, or that it may instead impede their performance (e.g., by displacing time spent on other revision activities), independent sample

**Table 4**

<table>
<thead>
<tr>
<th>Academic year</th>
<th>Correlation between data-handling test marks and overall marks for GGA</th>
<th>Partial correlation between data-handling test marks and overall marks for GGA, controlling for prior ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>0.522***</td>
<td>0.168***</td>
</tr>
<tr>
<td>2012</td>
<td>0.546***</td>
<td>0.351***</td>
</tr>
<tr>
<td>2013</td>
<td>0.629***</td>
<td>0.368***</td>
</tr>
</tbody>
</table>

Results are shown before and after controlling for prior ability, in each of 3 years of delivery. $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$, not significant (n.s.).
-tests (two-tailed) were performed for each cohort, to compare the exam MCQ, exam essay, and overall exam marks of those students who engaged voluntarily with PeerWise after the deadline with those who did not. No significant differences were found between the two groups in either 2012 or 2013 (average overall 2012 exam scores of 61.2 and 59.6 for 2013, difference between groups not significant), suggesting that voluntary answering of questions in the revision period did not have any measurable positive or negative impact on exam performance.

Additional Support Sessions Encouraged Peer Interaction, But Did Not Lead to Creation of Higher Quality Questions, Nor Improved Academic Performance
Data for 26 students attending optional PeerWise support sessions in 2013 (described in Table 1) was compared with data for 26 nonattending counterparts matched for equal prior learning (Fig. 4). We expected that attending the support sessions would encourage students to write better quality questions on PeerWise, and that they would engage more with the tool. Moreover, if these activities were critical for enhanced academic performance, we would also expect to see better overall marks on the course for attendees.

The attendees and their counterparts were spread across all four prior ability quartiles (data not shown). Surprisingly, paired sample t-tests (one-tailed) did not show those who attended the support sessions to write questions of higher quality than their academic counterparts, and nor did they perform better on the course (Figs. 4a and 4b). Attendees did contribute more to discussions in the commenting section of PeerWise, as evidenced by significantly greater character numbers within comments (Fig. 4c), and they also spent more time on PeerWise (measured by distinct days of activity—Fig. 4d), compared to their nonattending counterparts.

A Highly Competitive PeerWise Environment Can be Unpopular with Students
Student feedback on the use of PeerWise was encouraging in the first two years but less so in 2013. For iterations of
the GGA course in 2012 and 2013, identical end-of-course feedback questions relating to PeerWise were posed. The responses (in percentages) are shown in Fig. 5. In 2012, 92% of respondents agreed that PeerWise had improved their understanding of the course “a lot” (irrespective of whether they enjoyed the experience). In 2013, the class was less positive in attitude towards PeerWise, with only 55% agreeing with the same question and 20% of respondents (34 students) answering that PeerWise had helped them “not at all”. Through free-response surveys of the “best and worst” aspects of the GGA course in both years, as well as discussions at staff-student liaison meetings, we were able to ascertain that negative attitudes towards Peerwise were almost always connected to the peer-dependent element of the scoring system or, specifically in the unusually active 2013 cohort (see Table 1), the highly competitive environment they encountered on the system. In 2013 several students felt that the effort they were expending, as a result of this competition, was not well aligned with the proportion of final marks available for PeerWise activity on the course.

Discussion

This study has uncovered positive, statistically significant, associations between engagement with PeerWise and academic performance on a second year Genetics course, which were repeated over three successive course deliveries. As expected, the $r$ values of partial correlations presented here are small, since many factors contribute to academic performance and PeerWise is a relatively minor intervention. Indeed, another assessed formative learning exercise, the course data handling component, was also shown to have a statistically significant positive effect on overall course performance (Table 4). Our results reinforce the findings of other studies that have reported positive associations between PeerWise engagement and course grades across several STEM subject areas [1, 6–8]. Moreover, our data show that engagement with PeerWise did not only boost students’ ability to correctly answer MCQs in the final course exam, but that it was also strongly associated with performance across other diverse aspects of the course. This agrees with the findings of Denny et al. [1] who also noted a correlation between PeerWise activity and written exam work but whose correlation analyses did not correct for prior ability. Our findings demonstrate PeerWise engagement is robustly associated with benefits in coursework, exam essay and exam problem questions and suggests that effective PeerWise use encourages deep engagement with the subject matter, and promotes integration and application of what has been learned, allowing transference of learning across diverse tasks.

High ability and low/intermediate ability students on the GGA course (Quartiles 4 and 2, respectively, Fig. 1) appeared consistently to have gained the most benefit from PeerWise, as evidenced by comparing the performance of HPA and LPA students within ability quartiles. This pattern confirms the findings of a previous study using one cohort’s data from this course [8], and the variable penetrance of PeerWise, particularly for mid-ability students, is also consistent with results from a variety of datasets [1, 7, 8]. It is perhaps unsurprising that high ability students benefit greatly from an extra activity such as PeerWise, given that those students are expected to engage well with all aspects of their course and have a
firm knowledge base on which to build. Benefits demonstrated for low/intermediate ability students (Quartile 2) illustrate that PeerWise also assists less able students. It has previously been suggested that, for low/intermediate ability students, PeerWise could offer an activity that is challenging enough to be useful, but not so challenging as to be beyond their grasp [8]. Perhaps the lowest ability group could not engage effectively due to a weak knowledge base or habitual nonengagement. Indeed, “high” PeerWise engagement levels in the lowest ability quartile are considerably lower than those of other quartiles within each of the three year groups (Fig. 1). It remains unclear why the high/intermediate group (Quartile 3) should derive lower benefits from PeerWise engagement compared with other groups, but one explanation could be that this group might take a “strategic” learning approach, putting in the minimum effort required to gain marks. Given the accumulating evidence on depressed PeerWise benefits for mid-ability groups [1, 7, 8], useful further investigation might look in more detail at the nature of the engagement with PeerWise of different ability groups.

### TABLE 5

*r* values of Pearson product-moment correlations of average peer question ratings per student and their overall marks for the GGA course (PeerWise component removed)

<table>
<thead>
<tr>
<th>Academic year</th>
<th>Number of students</th>
<th>Correlation between overall course marks and peer-rated question quality</th>
<th>Partial correlation between overall course marks and peer-rated question quality, controlling for prior ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>226</td>
<td>0.185**</td>
<td>−0.076 (n.s.)</td>
</tr>
<tr>
<td>2012</td>
<td>193</td>
<td>0.294***</td>
<td>0.173*</td>
</tr>
<tr>
<td>2013</td>
<td>237</td>
<td>0.185**</td>
<td>0.064 (n.s.)</td>
</tr>
</tbody>
</table>

Results are shown before and after controlling for prior ability, in each of 3 years of delivery. *p < 0.05*, *p < 0.01***, *p < 0.001***, not significant (n.s.).

Comparison of attendees of support sessions (■) and nonattendees (□). Paired samples *t*-tests (one-tailed) were performed to determine whether support session attendees performed better compared with their non-attending academic counterparts with respect to (a) quality of authored questions, (b) overall course marks, (c) character length of all comments, and (d) distinct days of activity. Error bars show standard error of the mean. *n* = 52; *p < 0.05*, *p < 0.01***.
Our findings indicate that the learning benefits associated with PeerWise engagement results from a combination of reflective PeerWise question setting and peer discussion of questions, but not the drill-and-practice MCQ answering which we found to have no effect on exam outcome. As predicted, the “strongest” students created questions of the highest quality and, although a positive correlation was found between peer-rated question quality and academic performance on the GGA course, this was not generally statistically significant once prior ability was taken into account. This perhaps shows that writing questions of high quality on PeerWise has the potential to aid learning, but is not the whole story. Since question quality was not always associated with improved academic performance, while overall PeerWise engagement was, we suggest that the act of creating questions in the higher cognitive domains of Bloom’s ladder was not the sole route of benefits, which must instead arise from engagement with the wider set of activities that PeerWise offers, including post-question multipeer discussion. Indeed, it is widely accepted that dialogue-rich learning environments (including peer–peer dialogue) aid conceptual development in learners [12].

Two previous studies of PeerWise question quality found different levels of sophistication in student-authored questions. In one, the majority of questions created by second year Biology students were of low quality [4]. Conversely, a study of first year Physics students found only a small percentage (5%) of questions to be in the lowest quality category [10]. This study found that about one third of questions mapped onto the lowest learning domain, with one quarter of questions consistently mapping onto the uppermost three domains. This pattern was consistent across all 3 years (Fig. 3) and did not respond to varying levels of support given (Table 1). Students within our study who had attended extra support sessions prior to embarking on PeerWise did spend more time on the system, and contributed more to discussions, echoing the findings of another study [11]. However, within our cohort, attending extra support sessions did not alter outcomes in terms of numbers or quality ratings of questions authored, nor final academic performance. The quality of our students’ questions was intermediate between those described in the two other existing studies of question quality ([4] and [10]), although it is important to note that each study used its own appropriately customised rating scheme. It is also conceivable that some subject matter lends itself more readily to the creation of questions in higher cognitive domains, and that the course content in this study and the Physics study [10] represented such material. For our biology students we conclude that extra support did not improve the question quality or academic performance and that our previous level of introduction to PeerWise (~15 min of lecture time along with online and written materials) was sufficient. This “hands off” approach has the added benefit of entrusting to the students a greater sense of “ownership” of the exercise. Our sample size of 52 students in the extra support analysis would be too small to derive separate conclusions about each of the four ability quartiles, but an interesting avenue for further study would be to target support towards the lowest and intermediate/high ability quartile groups to analyse whether this would increase the PeerWise benefits from these less benefiting students.

Irrespective of additional support, we were encouraged to observe that instructor marking and peer ratings of question quality exhibited a strong, statistically significant positive correlation (r = 0.73; p < 0.001, Fig. 2). This indicates that students were intuitively able to rank questions similarly to staff, suggesting a firm grasp of the course material, as well as a natural appreciation of the relevance of higher cognitive activities. Previous studies
have calculated and used a “combined measure of activity” (CMA), which ranks students depending on their volume of activity in the areas of question authoring, answering, commenting and days of activity [1, 8], in preference to PeerWise score. Our findings suggest that the PeerWise scoring system (which relies heavily upon peer question ratings) is fair and reliable enough to be translated into marks, negating the need for staff to spend extra time individually marking questions. We found the PeerWise score to be strongly correlated with CMA for our cohorts (data not shown), but also consider it to better represent true engagement, since it includes a (peer-rated) reflection of question and comment quality. The PeerWise score has the added benefit of being automatically generated.

Most studies that report on student uptake of PeerWise reveal voluntary use of the system beyond what is formally required. This has taken the form of extra commenting [2], answering questions after assessment deadlines [2], and creating and answering more questions than required [2, 5, 11]. We noted similar enthusiasm in behavior as well as general positivity among our students towards PeerWise. However, a dissatisfied faction was evident in the 2013 deployment, where frenetic PeerWise activity led to heightened concern amongst students about the marking of PeerWise, which was known to depend on Peerwise performance relative to the performance of peers. Several students in 2013 created large numbers of questions and attained PeerWise scores that were much higher than the majority of the class (Table 1) and, as is normal, the scores of (anonymous) peers were visible to other students on the leaderboard. The competitiveness that ensued went well beyond the “healthy” competition that the system intends to encourage, as evidenced by our students’ end-of-course feedback (Fig. 5). Our experiences and discussions with other PeerWise users (colleagues and students) have now led us to devise a transparent, predefined scoring system, suitable for a class size of around 200–350 students. We have set a maximum number of ten questions that students can contribute, and reward PeerWise scores above 3,000, 4,000, and 5,000 with 3%, 4%, and 5% respectively of course credit. We award 2% to students performing the minimal requirements but failing to reach a PeerWise score of 3,000 and 1% for activity below the minimum. We now employ this system for the GGA course and it has also been adopted across a number of our other Biological Science courses this year, providing a consistency of experience for students encountering PeerWise on multiple courses.

Conclusions

Our results suggest that effective student engagement with PeerWise has the potential to foster deep learning, improving disparate skills such as essay writing and problem solving. The benefits of PeerWise do not depend on drill-and-practice answering of questions, nor are they wholly dependent on question setting, but appear to result from a combination of reflective question setting and good engagement in peer discussion. We have found the PeerWise score, generated by the system itself, to be a valid means of deriving marks for course credit that incorporates a reliable peer-assessed measurement of question quality, and that lengthy introductory sessions to the tool are not necessary. Both of these latter findings emphasise the point that PeerWise is an easily implemented learning intervention, requiring minimal instructor time, and generating unforeseen academic benefits.

Acknowledgments

This work was funded by a Principal’s Teaching Award Scheme grant from the University of Edinburgh. We would like to acknowledge the members of the Student-Generated Content for Learning (SGC4L) project team for regular discussions about PeerWise (SGC4L website: https://www.wiki.ed.ac.uk/display/SGC4L/Home).

References


