PeerWise: Replication Study of a Student-Collaborative Self-Testing Web Service in a U.S. Setting

Paul Denny
Dept. of Computer Science
The University of Auckland
Auckland, NZ
paul@cs.auckland.ac.nz

Brian Hanks
Computer Science Info Sys.
Fort Lewis College
Durango, CO USA
hanks_b@fortlewis.edu

Beth Simon
Computer Science and Engr. Dept.
University of California, San Diego
La Jolla, CA USA
bsimon@cs.ucsd.edu

ABSTRACT
PeerWise is a collaborative web-based system that engages students in the creation and evaluation of a test bank of multiple-choice questions. Previous studies involving two large CS1 courses in New Zealand have provided preliminary evidence that PeerWise usage is positively correlated with exam performance. In addition, it was reported that the student generated assessment was mostly free from errors and were clearly written, and the students appeared to positively value the system. Here we report on the first use of PeerWise in a CS1.5 course (second programming course) in the United States. Although the usage model was modified slightly to accommodate pedagogical factors, we highlight similar positive outcomes to those observed in New Zealand. Of particular note, students who were most active using PeerWise improved their rank in class relative to their peers who were less active.

Categories and Subject Descriptors
K.3.2 [Computer Science Education]: Introductory Programming – abstract programming concepts

General Terms
Algorithms, Human Factors

Keywords
CS1, MCQ, multiple choice question, peer assessment, educational technology, collaborative systems.

1. INTRODUCTION
Traditionally, multiple-choice questions (MCQs) have been created by instructors and used for summative assessment. PeerWise is a web-based tool that provides a framework for students to work collaboratively with their classmates to create a large repository of MCQs [3]. By creating the question bank, rather than just accessing an existing one, students are transformed from passive recipients to active learners. The process of creating a question engages a student in reflecting on the learning outcomes of the course, and requires they express their question clearly and consider misconceptions when proposing distracters (plausible, yet incorrect, answer options). Students are also encouraged to justify their understanding by explaining the answer to their question in their own words. The process of self-explanation promotes deeper conceptual understanding, as shown by Aleven and Koedinger [1] in a study of mathematics students.

Once questions have been submitted, students can answer them as a self-assessment activity and several opportunities for peer-comparison arise. Upon selecting an answer, a student is shown the question author’s answer and the most popular answer of their peers. The explanation for the question is also displayed, and students are able to read and contribute to the discussion threads associated with each question. Students are encouraged to evaluate the quality of the questions and explanations, a process that requires behaviors at the highest level in Bloom's taxonomy of the cognitive domain [2].

PeerWise provides regular opportunities for students to test their knowledge with questions aligned with course learning outcomes. Roediger et al. [8] showed that repeated testing, where students are regularly required to recall and apply their knowledge, is better than repeated studying at improving students' long term retention of information. He asserts "incorporating more frequent testing into a course may improve students' learning and promote retention of material long after a course has ended".

Murphy and Tenenberg [7] empirically investigated the extent of Computer Science students' knowledge of their own learning, and concluded that, "students should be provided with regular opportunities for empirical validation of their knowledge as well as being taught the metacognitive skills of regular self-testing". As an instructional strategy, PeerWise may indeed go some way towards addressing this call, by providing students with regular opportunities for self-testing and reflection.

The use of PeerWise has been evaluated previously in two CS1 courses taught at the University of Auckland. The driving force behind this prior work was a desire to address three common concerns expressed by instructors interested in using the tool in their own courses.

1. Repository quality. Instructors were unsure of the ability of students to write good questions that would form a useful learning resource.

2. Learning benefits. As an always-available online resource, another instructor concern was whether students who became actively engaged using the tool were spending their time wisely. Does PeerWise appear to offer any measurable learning benefits to those students who become most involved with it?
3. Student perceptions of activity value. Given that PeerWise is entirely student driven, it places virtually no burden on the course instructor. Instructors were interested in whether students would perceive it simply as an increase in their workload or whether they would enjoy the activity. In particular, voluntary student contributions to PeerWise would be a good sign that they value the peer-generated resource.

In this paper we report on the first use of PeerWise in the United States. This study examines data from two independent terms (Winter 2009 and Spring 2009) of a CS1.5 course taught at the University of California, San Diego (UCSD). We performed the study to investigate whether cultural and logistical differences have an impact on the effectiveness of this tool. We highlight the differences from its use in New Zealand, and compare our results to several of those from earlier work.

2. RELATED WORK

Previous studies have reported on the use of PeerWise in two CS1 courses taught by the system creator (Denny) at the University of Auckland. In this section, we reflect on these earlier results to set the scene for our study.

Repository quality. In 2008, a repository of 617 questions was developed by a class of 407 CS1 students [6]. A sample of 10% of the questions was examined and categorized, and it was found that 93% of the questions were clearly worded and 87% had a full set of feasible distracters. The majority of the questions were free from errors, with 80% having no errors in either the question stem or any included source code, and the correct answer was indicated by the question author in 89% of the cases. In every case examined where the answer was not correctly specified by the question author, this was detected and corrected by other students in the question discussion area.

After answering a question, students were given the opportunity to rate the quality of the question on a 6-point scale (0-5). The questions in the repository could be sorted by the average rating assigned. While the repository contained a number of low quality questions, students used the ratings of their peers to answer the high quality questions more frequently than the low quality questions [6].

Learning benefits. A quantitative analysis of a 460 student CS1 course taught in 2007 suggested that students who used PeerWise actively performed better in final examinations than their peers of similar initial abilities, but who were not as active [5]. The initial abilities of students were measured by their performance on a mid-term test, administered approximately half-way into the course, before PeerWise was introduced. While significant differences in final exam performance were found between the most and least active students within all achievement quartiles based on mid-term test performance, the greatest differences were found amongst students in the lowest quartile.

Student perceptions of activity value. Surveys showed that students enjoyed using the tool [4], quoting its value as a useful exam review resource as its biggest benefit. Although the amount of course credit offered to students for their participation was small (only 2%), students used PeerWise more often than was required. In 2007, students answered more than double the number of questions necessary to receive course credit, and 44% of the class used PeerWise voluntarily for exam review after the participation deadline [3].

3. METHODOLOGY

PeerWise was used in two second-quarter programming courses (CS1.5) at UCSD in 2009. We will refer to these two courses as Winter and Spring, corresponding to the terms in which they were taught. We examined repository quality, learning benefits, and student perceptions of activity value by replicating several of the analyses of previous studies.

In the case of repository quality, we have not replicated the expert analysis of question content, but instead we acknowledge that in a student generated repository there will be a range of questions of varying qualities. We focus on whether students can effectively find high quality questions at appropriate levels of difficulty based on ratings assigned to the questions by their peers.

For our analysis of learning benefits, the small number of students in our UCSD classes precludes direct replication of the analysis by quartile in [5]. In this study, PeerWise was used over the entire term (not just the second half of the term), so we use the final course grade in the prerequisite CS1 class as a baseline performance indicator. This allows us to examine the change in class rank of each student based on their level of activity.

Student perceptions of the value of the PeerWise activity are examined by considering the levels of voluntary usage of PeerWise in the Winter and Spring classes and by student survey.

3.1 CS1 in New Zealand

CS1 at the University of Auckland in New Zealand is a 12-week introductory programming course in Java. It is taken by both majors and non-majors and enrollment typically ranges between 400-500 students. Approximately 30% of students who originally enroll in the course either fail or withdraw.

In 2007 and 2008, when the previous studies were conducted, PeerWise was introduced to students half-way through the course, immediately after a mid-term test was conducted. Students were required to contribute two questions and to answer ten questions in the 6 weeks before the end of the term, for 2% of their course grade. The system remained available for voluntary use after the end of lectures and before the final exam (a period of two weeks).

3.2 CS1.5 in the U.S.

CS1.5 at the University of California, San Diego is a 10-week second programming course using Java. In this study, both the Winter and Spring CS1.5 courses, as well as both prerequisite CS1 courses, were taught by the same instructor (Simon). Failure rates (WDF) for the Winter and Spring CS1.5 courses were 6% and 12% respectively. The course is taken primarily by majors, with 84% majors in Winter and 54% majors in Spring.

The classes were considerably smaller than in Auckland, with 79 students in Winter and 63 in Spring. However, in order to have a consistent baseline performance measure (final grade in CS1 the previous term) we report on only 73 (Winter) and 53 (Spring) students who met that criterion. This allowed us to have students use PeerWise throughout each term, unlike in previous studies where the introduction of PeerWise was delayed in order to conduct a test to measure student abilities.

The requirements for students to use PeerWise differed between the Winter and Spring terms of CS1.5. In both, students used PeerWise on alternate weeks (when in-class quizzes were not given). In Winter this was 6 weeks – each week, half of the class contributed 1 question by Monday and the other half of the class answered 3 questions by Tuesday. This meant that the minimum
required contribution per student involved writing 3 questions and answering 9 questions over the term for 4% of the course grade. The instructor did not spend class time promoting the system, other than to say that it served as a self-assessment mechanism for the weeks where a formal quiz was not given. The instructor did not view student questions or responses during the term. Students in the Winter course created a repository of 202 questions.

In the Spring term, students used PeerWise a total of 5 weeks – each week writing 1 question and answering 3 questions earning 4% of the course grade. This increased the minimum required contribution per student to writing 5 questions and answering 15 over the term (and made for an assignment that was simpler to remember). The instructor was considerably more active in the Spring term promoting the use of PeerWise to students, and every week identified interesting questions to present in class. She discussed her rationale for why certain questions were useful, and gave suggestions for those that could be improved. She once awarded small prizes for good questions. At the end of the term, she presented research on the benefits of repeated testing and effortful recall [8], and emphasized the opportunities for such practice that PeerWise provides. Students in the Spring course created a repository of 297 questions.

4. RESULTS

In this section, we present the results of our examination of repository quality, learning benefits, and student perceptions of activity value for both the Winter and Spring classes, and make comparisons to earlier studies where appropriate.

4.1 Repository Quality

In [6], it was reported that questions rated highly for quality received a greater number of responses than poorly rated questions. It was also shown that student assigned quality ratings were well correlated with instructor ratings of the same questions. Figures 1 and 2 show the number of responses each question received in the Winter and Spring classes, against the average quality rating of that question. It is clear from these figures that poorly rated questions generally did not receive a large number of responses, consistent with the results in the New Zealand courses. This indicates that while the quality of the questions in the repository naturally varies, students do use the ratings assigned to the questions by their peers to help them determine which questions to spend their time answering.

Table 1 presents the correlations between question quality and question popularity for both the Winter and Spring courses, as well as for the New Zealand course reported in [6]. All correlation values are very similar, and above 0.3.

<table>
<thead>
<tr>
<th>Course</th>
<th>NZ</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation (R)</td>
<td>0.34</td>
<td>0.35</td>
<td>0.31</td>
</tr>
</tbody>
</table>

In addition to quality ratings, students are also given an opportunity to rate the difficulty of questions they answer on a 3-point scale (0-2). Like the quality ratings, these difficulty ratings are shared and can be used by students to answer questions at a difficulty level they are comfortable with. We are able to measure the accuracy of these difficulty ratings by comparing them with an objective measure of difficulty – the percentage of responses to a question that are correct. Table 2 summarizes the correlation between student assigned difficulty ratings and the percentage of correct responses. There are strong negative correlations in both the Winter and Spring courses (showing that questions rated more difficult by students have fewer correct responses). Although not previously reported in the New Zealand studies, data from the 2008 course yields a correlation of -0.56.

<table>
<thead>
<tr>
<th>Course</th>
<th>NZ (2008)</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation (R)</td>
<td>-0.56</td>
<td>-0.58</td>
<td>-0.60</td>
</tr>
</tbody>
</table>

4.2 Learning benefits

Previous studies have shown that students who are most active using PeerWise, particularly when measured by the number of questions answered, outperform their less active peers in formal examinations.

In the New Zealand study, the midterm test grades were used to establish a performance baseline, which was used to assess the impact of PeerWise use on student performance. In our study, each student’s final course grade in their CS1 course was used to give them an initial class rank in the CS1.5 course. At the end of the Winter and Spring courses, the student rank was again computed. Every student therefore had a change in rank depending on whether their position in class improved or declined.
relative to their peers. Figures 3 and 4 show the average change in class rank when the students are divided into quartiles based on the number of PeerWise questions they answered, where Q1 is comprised of the most active PeerWise users (the values in parentheses in each of these figures are the mean number of questions answered by students in the corresponding quartile). These figures clearly show that the more questions a student answered when using PeerWise, the greater the improvement in their class rank. In fact, the most active students were improving their class rank by an average of 8 places in a class of 73 (Winter) and by an average of nearly 6 places in a class of 53 (Spring). A one-way ANOVA on the Winter data showed that significant differences existed between the quartiles, and a Bonferroni corrected pairwise t-test showed that the difference between the rank changes in the most active (Q1) and least active (Q4) group was highly significant (p=0.004). While a one-way ANOVA on the Spring data showed that the differences were not significant (although very close, p=0.060), the probability of both charts exhibiting a monotonic trend by chance is less than 0.7%. The fact that in both cases the trend is monotonically decreasing is very strong evidence of a positive relationship between student activity and improvement in class rank.

We’d like to highlight that the relationship between PeerWise use and change in class rank was similar in both courses, even though the pattern of PeerWise use in the two courses differed markedly. In Winter, those students who were the highest performing in the prerequisite CS1 course tended to be the most active PeerWise users, while in the Spring course it was the lower performing CS1 students who were more active. This suggests that using PeerWise helps all students regardless of their academic standing.

4.3 Student perceptions of activity value
Voluntary usage of PeerWise is a good indicator that students value the peer-generated resource. The total number of questions that students were required to answer in order to be awarded their participation marks was 9 and 15 in the Winter and Spring courses respectively. In Spring, 72% of students answered questions beyond the requirement – a marked increase from 45% who answered more than required in Winter (Table 3). This occurred despite the higher requirement in Spring. This may be attributed to the fact that the instructor spent more time promoting the benefits of the system in this course.

Table 3. Number of students per activity level

<table>
<thead>
<tr>
<th>Activity level (no. questions answered)</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>10-15</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>16-19</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>20-29</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>30-49</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>50-99</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>&gt;99</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
These results are similar to those seen in earlier studies, in which there is clear evidence that many students use the question repository voluntarily, answering questions beyond the requirement for their grade.

Finally, on a survey conducted in the Spring course, student perceptions of the value of PeerWise were generally positive (Table 4), consistent with the earlier results in [4].

Table 4. Survey results (Spring, N=60)

<table>
<thead>
<tr>
<th>Question</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing new questions helped me learn</td>
<td>47%</td>
<td>37%</td>
<td>17%</td>
</tr>
<tr>
<td>Answering other student's questions helped me learn</td>
<td>65%</td>
<td>28%</td>
<td>7%</td>
</tr>
<tr>
<td>I would recommend other professors to offer PeerWise in their class</td>
<td>50%</td>
<td>29%</td>
<td>21%</td>
</tr>
</tbody>
</table>

5. DISCUSSION AND FUTURE WORK

Based on our experience using PeerWise in the Winter and Spring courses discussed in this paper, we make the following observations and suggestions for future work.

Repository quality. Accurate student ratings may improve the usefulness of the repository as a review resource. In both Winter and Spring, students made use of the quality ratings to decide which questions to answer, and the difficulty ratings appeared to be very accurate, yielding strong negative correlations with the percentage of correct responses. There may also be options we can explore to better help students find useful questions. Possible options include providing training to students in how to rate questions, rewarding students for accurately rating questions, requiring students to rate more questions (rating questions was voluntary in our studies), and giving greater weighting to instructor, TA, or “top” student ratings.

Learning benefits. Interestingly, the students answering the greatest number of questions in the Winter course tended to be those with a stronger background (based on their CS1 prerequisite mark), whereas in the Spring course it was the students with a weaker background who were more active. We believe this is due to the fact that the instructor promoted the use of PeerWise much more actively in the Spring course. Without repeated instructor encouragement, most use of the tool was made by the already high performing students. However, what is striking is that the relationship between PeerWise use and change in class rank was similar in the Winter and Spring courses (see Figures 3 and 4), indicating that PeerWise may potentially help all students regardless of academic rank. This corroborates previous results [5] that the most active students perform better than the least active students of similar initial abilities.

We also acknowledge that our measure of activity (number of questions answered) is, in some sense, a measure of time on task. However we are pleased to note that this was an entirely student-driven activity, and many students participated to a greater extent than they were required.

Student perceptions of activity value. Students answered significantly more questions than they were required in both courses we studied. This voluntary usage, combined with generally positive survey results, indicates that many students do value both the activity and the peer-created resource. The numbers alone tell only part of the story. In future work we plan a qualitative exploration of several issues including: is the quality of a student’s questions influenced by the number of questions they answer, how do students choose which questions to answer (and does that vary based on whether they are answering voluntarily or to receive course credit); and how do students make their studying choices including investigating how PeerWise use impacts on other forms of studying (reading the textbook, reviewing quizzes, etc.).

6. CONCLUSIONS

In this paper, we reported on the first use of PeerWise in Computer Science courses in the United States. Data collected from two consecutive terms of a CS1.5 course taught at UCSD support results of earlier studies examining CS1 courses taught in New Zealand.

By establishing a learning community that incorporates active learning, peer review, and peer tutoring, we believe PeerWise has the potential to be a powerful teaching tool.

7. REFERENCES