The PeerWise System of Student Contributed Assessment Questions

Paul Denny **Andrew Luxton-Reilly** **John Hamer**

Department of Computer Science
University of Auckland

Abstract

Large test banks of multiple choice questions (MCQs) are popular resources for students wishing to quickly learn course material. However, they are time consuming to create and offer a somewhat limited learning experience.

PeerWise offers an innovative approach that enhances standard teaching and learning practice by requiring students to participate in the construction and evaluation of MCQs. The system encourages the development of higher order cognitive skills and enhances student learning with virtually no additional cost to teaching staff.

We have now used PeerWise in several large undergraduate programming courses, and report here on the design of the system and its user interface, identify several related systems, discuss our motivation and underlying teaching philosophy, and present some usage and performance results.

1 Introduction

This paper describes an innovative system used to augment standard teaching and learning practice by means of student generated multiple choice questions. Traditionally, multiple choice questions (MCQs) have been constructed by teaching staff and used for summative assessment. Multiple choice question test banks have also been used in a drill and practice fashion by students. Asking students to generate a MCQ and provide an appropriate explanation of the answer yields a much richer and deeper learning experience. Furthermore, asking students to critically evaluate existing MCQ items and provide formative feedback about the quality of the question requires the application of higher order cognitive skills such as making critical judgements.

The system described in this paper incorporates these aspects resulting in a productive and fertile learning environment. We note a number of benefits:

- a large test bank of MCQs are created at low cost to teaching staff;
- students can use the test bank to perform drill and practice;
- the assessment process itself is publicly discussed, focusing attention on the learning outcomes of a course and the means of assessing such outcomes. Students benefit from a better understanding of the teaching and learning goals;
- the activity of writing a question requires a student to think carefully about a topic in the course and how it relates to the desired learning outcomes. Creating distractors requires a student to consider possible misconceptions. The act of writing an explanation requires students to express their understanding of a concept in their own words, deepening their own knowledge and enhancing their written communication skills;
- evaluating existing questions incorporates higher-order cognitive skills, requiring a student to consider not only the content, but what makes a particular question more effective than other questions. The element of peer assessment and provision for feedback moves a student from being merely the recipient of an educational process to being an active partner, creating, assessing and making critical judgements.

We are in the early stages of formally evaluating the benefits noted above, and report here some preliminary results on the usage of this system by students in a large undergraduate course.

2 Description of the system

PeerWise is a web-based repository of multiple-choice questions with alternatives and explanations written by students as part of their required coursework.

All activity, such as developing new questions, answering existing questions, and rating and providing feedback on questions is confidential.

After logging in, the main menu (fig. 1) is divided into three sections entitled: “Your questions”, “Answered Questions” and “Unanswered questions”. The role of each of these sections is described next.

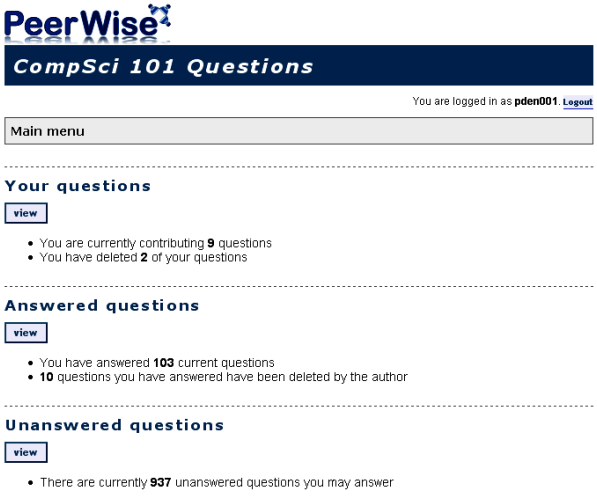

Figure 1: The main menu for the PeerWise system

2.1 Your questions

This section (fig. 2) allows a student to review all of the questions they have contributed to the system. The questions are displayed in a table with columns listing the date the question was developed, the number of responses, and the rating. The table can be sorted on any of these keys. A specific item can be selected from the table, to display details such as how often each alternative was selected and any feedback provided by students who have answered it. There is also a column in the table which displays the perceived difficulty of the question, as rated by students who have answered it. Another column displays whether or not the question is "suitable", which occurs when it has a rating greater than 2, and the most popular alternative selected is the correct alternative. If either of these conditions is not met, it may indicate that the question is overly tricky, or contains errors.

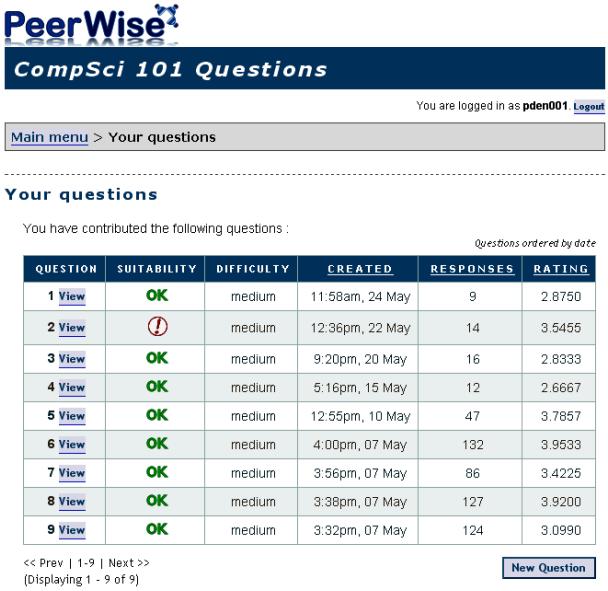

Figure 2: Page showing the questions written by the student

When creating a new question, the contributor needs to provide a question stem, at least two and up to five alternatives, an indication of which alternative is correct, and an explanation of why that is the correct alternative. The explanation is shown to all students who answer the question, and serves to assist students who select an incorrect alternative to identify their misunderstanding.

Each new question can be tagged with the name of any relevant course topics, which allows students using the system for revision to easily find questions of interest. As soon as a question is contributed, it will appear in the "Unanswered questions" section (fig. 3) for other students in the course.

2.2 Unanswered questions

Each question in the system is available to every student in the course. The unanswered questions are organised into a table that can be sorted by the order they were developed, or by the number of responses they have received, or by the rating they have been given.

Once a student selects a specific item to answer, the question stem and the alternatives for that question are displayed. The student then selects the alternative they believe to be correct, at which point they will be shown the correct alternative, as suggested by

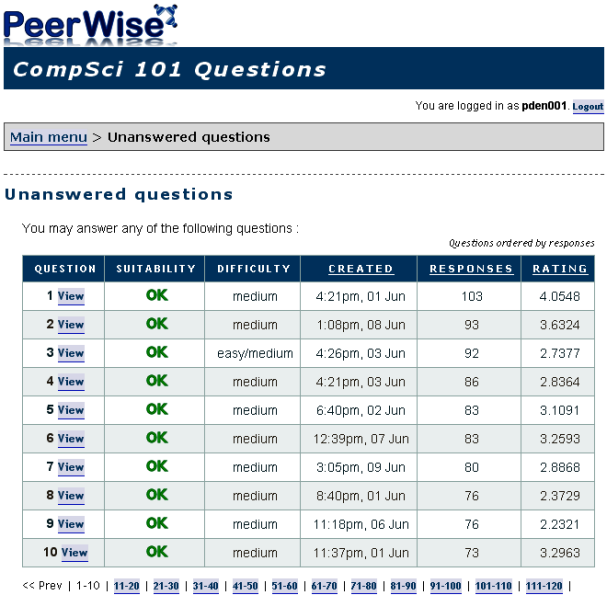

Figure 3: Page showing unanswered questions

the author of the question, as well as a histogram of all students' responses to the question. The explanation provided by the author is also displayed, along with any comments previously written by other students. In addition, a simple metric is used to approximate whether the selected answer is actually correct. The selected alternative is deemed to be correct if it agrees with the answer suggested by the author, and if this alternative is also the most popular amongst all previous responses.

After receiving this feedback, the student who answered the question has an opportunity to rate it and provide open-ended feedback. The rating scale is an integer between 0 and 5, and is expected to take into account the quality of the question, the distractors and the explanation. The student is also able to rate the difficulty of the question as either "easy", "medium" or "hard". Once a question has been answered and rated, it will always be available for review by the student in the "Answered questions" section (fig. 4).

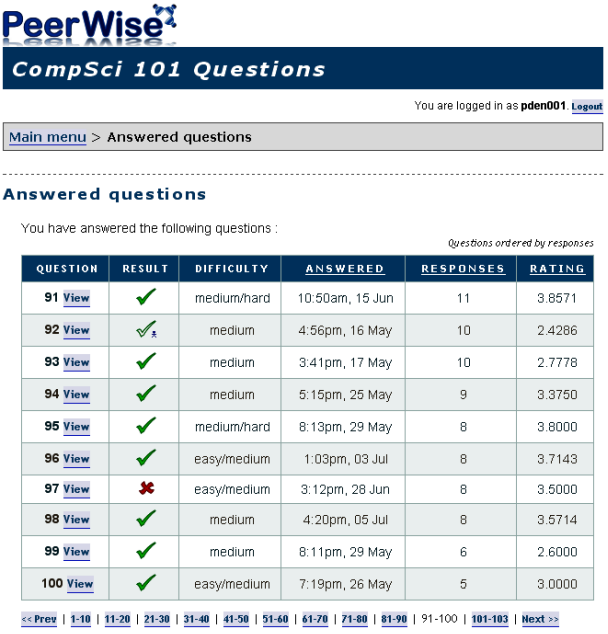

Figure 4: Page showing answered questions

2.3 Answered questions

All previously answered questions are available and can be reviewed at any time. As other students provide responses, the accuracy of the correctness metric improves. The table that displays the answered questions can be sorted by the order in which the questions were answered, by the total number of responses to the question, or by the question rating.

A basic leaderboard is also available (fig. 5), which ranks students contributions. It was included to provide some motivation for participation well beyond the minimum requirements for assessment. Tables on the leaderboard display the top rated questions, and rank students on the number of questions they have answered, the popularity of the questions they have contributed, and the popularity of any open-ended comments they have written during the rating process.

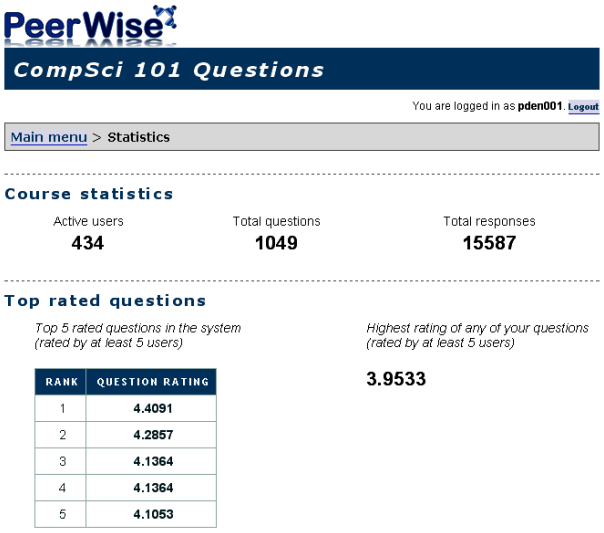

Figure 5: Page showing leaderboard

3 Related work

A number of systems have been developed with similar goals to our own.

Arthur (2006) reports on a class with streams of about 50 students. Each week a group of three or four from each stream write six to eight questions (sufficient for a 10-minute quiz), which are then submitted electronically. The quiz is taken by students in the next stream. The exercise helps students identify the key concepts and methods for a topic. Arthur notes that students get to consider plausible alternatives, and thus explore variations in the meaning of a concept. Unlike published test banks, which become obsolete quickly, the approach also ensures the quiz questions are up-to-date.

Yu et al. (2005) has students construct a question stem, with one correct answer and three alternatives. Hints, cues and reference/citations, etc. can be added as learning aids or feedback during drill-and-practice exercises. Questions are entered online, where they are peer-assessed. The peer-assessment interface includes pull-down menus listing common mistakes. In addition, students can type in detailed suggestions for further refinement of the examined item. After the peer assessment phase, questions are transferred from the temporary item bank and into an item bank database to be used for drill-and-practice exercises. The class decides on the number of questions for a given learning unit for the drill-and-practice exercises. On completion, feedback as to the percentage of correct answers and a review button for the missed ones is provided for review purposes.

Baraka & Rafaelib (2004) reports on a postgraduate MBA course in which students contribute questions to an on-line repository. The students also rank their peers' contributions.

Chang et al. (2005) operate in a "one-on-one" educational computing classroom in which every student uses a digital learning device, such as personal digital assistant, notebook, tablet PC, etc., to participate in learning activities. They have elaborated a theme of "asking a good question" in which each student generates a question and answer, then applies a self-assessment rubric before sending the question to two anonymous peer reviewers for peer-assessment. The mutual reviewers then form triads, and each triad selects two items for a class-wide discussion led by the teacher. During the discussion, the teacher points out if there are misconceptions and misunderstandings, and each student has to elaborate, justify, or clarify possible confusions on the ideas of her own item, and correct the item as necessary.

4 Discussion

Providing students with an MCQ test bank to perform drill and practice is an effective way to develop skills at the lower end of Bloom's taxonomy of educational objectives (as described by Anderson et al. (2001)), such as remembering, understanding and applying known procedures to new data, especially when the MCQ test items have an explanation of the correct response. However, we can exploit the MCQ domain to yield a much richer range of learning experiences by asking students to generate the MCQ item, write an explanation, evaluate the effectiveness of the MCQ item and provide feedback.

As teachers involved in higher education, our ultimate goal is to facilitate the transformation of students into independent life-long learners. A critical part of this transformation is engaging students in the process of reflection, peer and self assessment.

Ramsden (2003, p90) makes the claim

> "Good teaching helps students to become aware that educationally valid assessment is an opportunity to learn and to reveal the depth of one's knowledge"

He further describes a number of principles of good teaching, including striving for student independence, control and active engagement in the learning process. He claims "Good teaching fosters a sense of student control over their learning" [p.97]

We feel that students benefit greatly by participating in the development and application of assessment items. Students should be permitted to create items for any area of the course, and encouraged to align their test questions with the desired learning outcomes. Involving students in the development of assessment items in this fashion puts the educational process in focus and empowers students by providing a greater degree of control.

4.1 Reflection

Informing students about the nature and purpose of assessment activities provides a deeper awareness of learning processes. Discussing the construction of MCQ test items in class helps to demystify the assessment of learning outcomes and provides insight into how course objectives are being measured.

Before students are able to construct a MCQ test item, they must first consider what has been taught in

the course, reflect on the learning outcomes and utilize the knowledge they have constructed about the course content. Once they have identified the important concepts related to a given topic, students are in a position where they can create a question. In order to create feasible distractors, students must consider the possible misconceptions that could occur. In order to be able to explain why the answer is the correct choice and why the distractors are incorrect, students must first understand the topic at hand. The entire process of question creation entails serious reflection on the nature of the course content and the learning objectives. This process, then, promotes deep learning.

4.2 Peer Assessment

Providing good feedback is a critical aspect to effective learning. Nicol & Macfarlane-Dick (2006) have identified seven principles of feedback that help students become independent learners:

1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of reflection and self-assessment in learning;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape the teaching.

Presenting students with the opportunity to provide feedback on assessment items is an important aspect of the learning process. Students are encouraged to critically evaluate existing MCQ items. Providing feedback about questions requires students to think critically and analytically about the MCQ test items. Evaluating the questions written by their peers may encourage students to reflect on their own questions. It can also motivate dialogue about the desired learning outcomes, the best way to evaluate topics, or identify the most important concepts.

Futhermore, the nature of the comments and evaluations performed by students provides valuable information for the teaching staff, giving a sense of which topics students are engaging with, and which they are not. We feel that the feedback obtained through the system satisfies the principles outlined by Nicol & Macfarlane-Dick outlined above.

4.3 Deep Learning

If students are presented with an existing MCQ test bank, then they interact with the test bank on a superficial level, merely using it as a drill and practice opportunity. Although there are benefits to be obtained from utilising the PeerWise system to perform drill and practice, there are many greater learning opportunities provided by fully engaging in the process of constructing, using, and evaluating MCQ test items.

Students that have been involved in the creative process of constructing MCQ test items that form part of the MCQ test bank will tend to bring critical analysis skills to bear. They are not only interested in getting the right answer to an MCQ question, but will also be involved in the evaluation of the question. Students engage with the MCQ test item by asking questions such as:

- How effective was this question at evaluating one of the learning outcomes?
- What makes the item a good or bad question?
- Were the distractors effective?
- How could the item be improved?

These kind of questions engage the student in a deeper and richer learning experience.

4.4 Staff benefits

There are significant benefits for staff. The development of MCQ test banks is a very time consuming activity for staff. By placing the creation of the MCQ items in the hands of the students, a significant degree of time is saved. This is a fast, low cost way for staff to have access to a large body of MCQ test items designed specifically to test the course content. By evaluating the topic areas that students have created questions for, staff can get a sense of which topics students are more confident with and which topics students are not engaged with.

Looking at the results provides further information about how students are performing, and what kinds of misconceptions students have. Analysing the feedback can often reveal further insight into the student perception of topics within the course.

Utilising an automated system for MCQ is scalable to large classes. The larger the class, the more effective the system will perform. The number of high-quality questions will be greater. Students will therefore have access to a higher number of effective questions.

5 Preliminary results

The PeerWise system was used for the first time with a large class (500 students) in a 12 week semester course running from March to June 2007. It was introduced to the students during a lecture approximately halfway into the 12 week course.

Contributing to PeerWise was an assessed activity, but worth only a small fraction of the students' final marks. The students were required to develop at least two questions, and answer and rate at least ten questions in the system for 2% of their final grade. The deadline for this contribution was June 1st, the last day of the semester, approximately five weeks after the system was introduced. The final exam for the course, which is the major assessed activity contributing 60% towards the final grade, was held on June 15th. We will refer to the two-week period between the coursework deadline (1st June) and the final exam (15th June) as the "study period".

We analysed the usage of PeerWise from when it was introduced until the day of the final exam, to investigate how the level of participation varied across students of different abilities, and what the level of participation was like during the study period.

For this discussion, we have ranked the students according to their final result in the course. Figures 6 and 7 show, for each student in ranked order, the number of questions they contributed, and the number of questions they answered. Although, as expected, there is a slight trend of higher participation at the top end of the class, we noticed that the overall participation was fairly uniform across students of

different abilities. It was pleasing to see that even students near the lower end of the class were willing to contribute more than the minimum that was required for assessment.

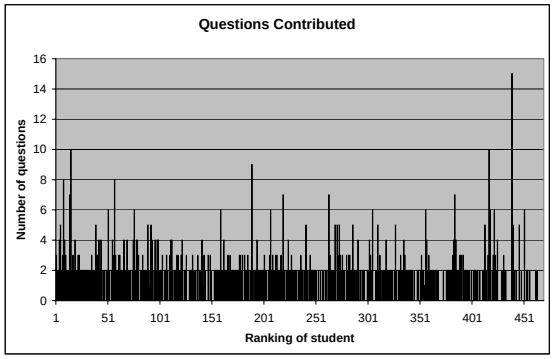

Figure 6: Number of questions contributed, by student rank

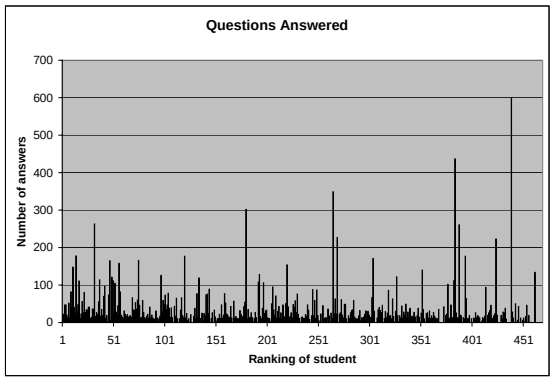

Figure 7: Number of questions answered, by student rank

We were interested to see whether the students would find any real value in voluntarily using the system during the study period, after their contributions had been assessed.

Figures 8 and 9 show the number of questions that were contributed to the system, and the number of existing questions that were answered, from the day the system was introduced until the day of the final exam.

In figure 8, there is a dramatic increase in the number of questions being contributed to the system on the days immediately prior to the assessment deadline, after which very few new questions are added. This pattern matches our expectations in that the increased rate of contributions is explained by students simply working to the deadline, and during the study period most students preferred more traditional revision techniques over spending time developing new questions.

We found the pattern in figure 9 quite interesting and much more marked than we expected. During the study period, the system was used very heavily for drill and practice revision. This seems to suggest

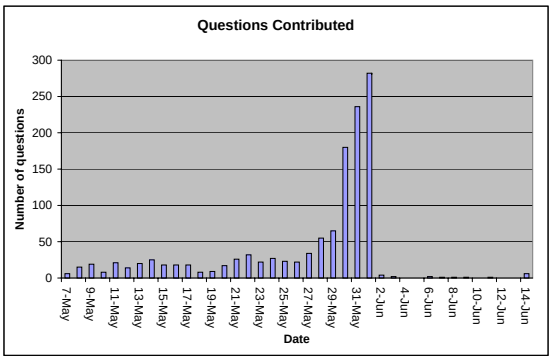

Figure 8: Number of questions contributed, by day

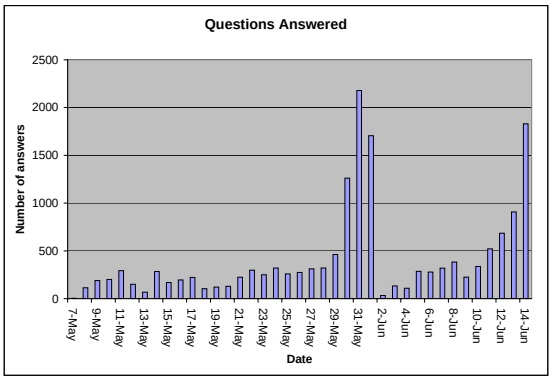

Figure 9: Number of questions answered, by day

that the students saw real value in the repository of questions developed by their peers. In fact, for most of the study period, questions were being answered at a faster rate than we saw for much of the time prior to the assessment deadline.

5.1 Question ratings

Developing a question that received a high rating was not a simple task. We noticed that any new question that was too similar in style to existing questions struggled to get a good rating.

Typically, the questions that were highly rated had a unique style, or a very detailed explanation. Well explained answers were often rewarded with positive comments such as:

> *"Very brilliant explanation, even more detail than the cousebook!!!"*
>
> *"The explanation is long but very detailed. Someone that got it wrong is sure to get it right now."*

On the other hand, many of the poorly rated questions either contained some kind of error, or did not have an explanation.

6 Future work

Over the course of the semester, we have received some feedback regarding the usability of the system

and we will look to improve certain aspects of the interface before next semester. The PeerWise system has also attracted interest from staff teaching large classes outside the Computer Science department. We plan to collect further data from these classes in order to determine how effective the system is with a broad range of students from a wide range of subject areas.

References

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J. & Wittrock, M., eds (2001), *A taxonomy for learning and teaching and assessing: A revision of Bloom's taxonomy of educational objectives*, Addison Wesley Longman.

Arthur, N. (2006), 'Using student-generated assessment items to enhance teamwork, feedback and the learning process', *Synergy* **24**, 21–23. `www.itl.usyd.edu.au/synergy`.

Baraka, M. & Rafaelib, S. (2004), 'On-line question-posing and peer-assessment as means for web-based knowledge sharing in learning', *International Journal of Human-Computer Studies* **61**, 84–103.

Chang, S.-B., Huang, H.-M., Tung, K.-J. & Chan, T.-W. (2005), Agq: a model of student question generation supported by one-on-one educational computing, *in* 'CSCL'05: Proceedings of the 2005 Conference on Computer Support for Collaborative Learning', International Society of the Learning Sciences, pp. 28–32.

Nicol, D. J. & Macfarlane-Dick, D. (2006), 'Formative assessment and self-regulated learning: A model and seven principles of good feedback practice', *Studies in Higher Education* **31**(2), 199–218.

Ramsden, P. (2003), *Learning to teach in higher education*, 2nd edn, RoutledgeFalmer, London; New York.

Yu, F.-Y., Liu, Y.-H. & Chan, T.-W. (2005), 'A web-based learning system for question posing and peer assessment', *Innovations in Education and Teaching International* **42**(4), 337–348.