Research paper

Making the most of using PeerWise in education

Lasse Hakulinen and *Ari Korhonen*
Department of Computer Science and Engineering,
Aalto University School of Science and Technology, Finland
lhakulin@cs.hut.fi, archie@cs.hut.fi

Abstract

PeerWise is a system that enables collaborative learning by allowing students to create, answer and discuss multiple-choice questions. In PeerWise, students create a repository of multiple-choice questions on the topics covered in a course. The system does not require course staff to be actively involved, but rather encourages students to rate and comment questions made by their peers.

In this work, we report on the results of using the system in several courses at the Aalto University School of Science and Technology. Especially, we have studied the quality of the student-generated questions on the Data Structures and Algorithms course. We found that there are some flawed questions, but in overall, the students are capable of creating questions of good quality. We also discuss improvements to recognize flawed questions more efficiently.

PeerWise is a system that depends heavily on the collaboration and interaction of students. In our experiences, systems such as PeerWise are quite sensitive to the way they are integrated into the course. In order to fully benefit from the collaborative nature of PeerWise, students should be motivated to use the system in a way that supports the learning goals of the course. Thus, we also suggest some guidelines that should be borne in mind when using PeerWise on a course.

Keywords: PeerWise, multiple-choice questions, CSCW, collaborative learning, peer assessment

1. Introduction

Collaboration in education can appear in many different forms. For example, students can interact with teachers and each other in lectures or small study groups. Collaboration can be encouraged by creating an environment where collaboration is easy and serves the purpose specific to the learning goals of the course. It can be carried out by creating facilities for face-to-face interactions or introducing a computer system that enables collaboration among users.

In this work, we study the usage of the PeerWise [3] system that was used on several courses at the Aalto University School of Science and Technology in 2009 and 2010. PeerWise is a system where students can create multiple-choice questions and answer

the questions created by other students. They can also comment and rate the quality of such questions while they answer them.

In order to be useful, the questions in the system should be good enough quality. As the multiple-choice questions are made by novices—what comes to the topic—there is no guarantee that the questions nor the choices are well formed. Although, not all questions were good quality, we are able to conclude that students can create questions of good quality even if the question repository is not moderated by the course staff.

The level of expertness is not the only factor what comes to the quality of the questions. We found that it is important to make sure that PeerWise is properly integrated into the course and the students are motivated to use it. Based on the experiences from the courses that used PeerWise, we have given some guidelines how students and teachers could get the most out of PeerWise.

2. Related work

The success of educational technology used in Computer Supported Cooperative Work (CSCW) depends not only on the systems used, but also the way the systems are utilized and whether the students are motivated to use them. It is very important to find the best practices for each course and system to get the best possible benefit from the CSCW systems. Krejns et al. [9] focused on identifying the pitfalls for social interaction in computer-supported collaborative learning environments. They stated that it can not be taken for granted that the participants will socially interact simply because the environment makes it possible. Orlikowski [10] got similar results when exploring the introduction of groupware into an organization.

The material used for studying in CSCW system might be produced by students. In that case, the reliability of the study material is of concern. If the material is not moderated by course staff, there is a possibility that there are some incorrect or misleading material. On the other hand, if the material is reviewed before publishing, the level of interaction is reduced significantly and benefits of collaboration might be lost. Kittur et al. [8] researched the trustworthiness of articles in wiki environments. They stated that the level of trust to the content of wikis and other collaborative systems may be increased by showing users the history of contributors and the level of stability of the content. Similar results were reported also by Suh et al. [11]. They introduced a social dynamic analysis tool called WikiDashboard that improves the transparency on wiki articles. They found out that the increased transparency can improve the interpretation, communication, and trustworthiness of wiki articles.

The effects of PeerWise usage on learning were studied by Denny et al. [1] on a standard first-year programming course. They divided the class into four quartiles based on a midterm examination, which was held before any use of PeerWise. They divided each of these quartiles into two groups based on the students' activity level in PeerWise and compared the exam results of the most and the least active students of each quartile. They found that an active use of PeerWise during the semester was strongly related to students' better performance in the final exam. In the final exam, they had multiple-choice questions as well as written sections. The results showed that an active use of PeerWise

was related to a better performance in both, the multiple-choice, and the written questions of the final exam. Similar results were found when replicating the study in a CS1.5 course at USCD [2].

In another research, the coverage of topics on student-made questions was studied on an introductory programming course by Denny et al. [5]. Although, students were able to choose the topics of the questions freely, the topics covered all the major topics in the curriculum. In addition, the quality of the questions was studied as well. Denny et al. [6] reported that 80% of the questions in an introductory programming course did not have any errors in them. Moreover, the student attitude towards PeerWise was also studied by Denny et al. [4]. They reported on that the students considered answering to other students' questions to be more educative than creating new questions. However, this opinion was not supported by the exam results. Nevertheless, they reported on that participating in discussions in PeerWise seems to have a positive effect on the exam results.

3. PeerWise

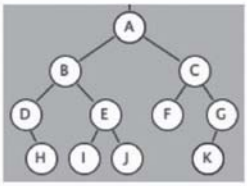
PeerWise is an online tool that enables students' interaction and collaboration while studying. Because the idea of PeerWise is to create a repository of multiple-choice questions by students, it can be used in many different courses and it is not tied to a specific field of education. PeerWise does not require the course staff to be active administrating the questions and answers, but rather encourages students to correct each other's misunderstandings.

Creating a new question is straightforward. The student author provides the question and the correct answer as well as the distractors. The author should also provide an explanation text, which is shown after the question has been answered. A preview of an example question is shown in Figure 1.

Question

This is how the text of your question will appear to others:

Below is a binary tree. If the tree is traversed in inorder, what is the order the nodes are visited?



Alternatives

This is how the alternatives for your question will appear to others:

Select your answer

OPTION	ALTERNATIVE
A	A, B, D, H, E, I, J, C, F, G, K
B	A, B, C, D, E, F, G, H, I, J, K
C	D, H, B, I, E, J, A, F, C, K, G
D	H, D, I, J, E, B, F, K, G, C, A

Answer

You have indicated that the correct alternative is **C**

Explanation

This is how your explanation will appear to others:

When traversing a binary tree in inorder, the following operations are performed recursively at each node:

1. Traverse the left subtree.
2. Visit the root.
3. Traverse the right subtree.

Choice A is the sequence when traversing the tree in preorder
Choice B is the sequence when traversing the tree in level-order
Choice D is the sequence when traversing the tree in postorder

Figure 1. Preview of a new question

After one has answered a question, the explanation as well as the choice that the author suggested to be the correct one, and the distribution of the answers of the other answerers are shown. The answerer can then rate the difficulty and the quality of the question, and possibly provide some feedback for it. The author of the question and other answerers can see all the comments written for the question. The comments provide students a convenient forum to discuss matters related to the question.

4. Research design

In this paper, we describe the usage of PeerWise at the Aalto University School of Science and Technology. We studied the usage of PeerWise and the quality of the questions in order to find good practices for making a good use of PeerWise as an educational tool. The following research questions were addressed based on the data collected from PeerWise:

RQ1: What is the quality of the student-generated questions?
RQ2: How to detect faulty questions in PeerWise?
RQ3: How do the students find the usage of PeerWise?

4.1. Research methods

PeerWise was used for the first time at Aalto on the Data Structures and Algorithms course in 2009. Based on the collected data, the quality and correctness of the questions in the repository was studied. The data was also used to study the reliability of the ratings given by the students and to develop a method for recognizing flawed questions automatically. Hakulinen has reported on these results in his Master's thesis [7]. PeerWise was utilized also on two other courses in 2009 and we continued to use it in 2010.

In order to examine the quality of the questions, a set of questions was answered and rated by a board of three experts. The board of experts consisted of three researchers who have been, or are currently, members of staff on the Data Structures and Algorithms course. All the questions in the repository—that were rated at least by five students—were sorted according to the average ratings. Then, every other question was selected to be evaluated by the experts. This way we made sure that the total of 48 evaluated questions had different average quality ratings. The instructions to answer and rate the questions were the same with students and experts. First, the assumed correct answer to the question was selected. Second, the difficulty and quality of the question was rated. However, experts did not see the comments to the question given by students in order to prevent this to affect their judgement.

4.1.1. Classification of the questions

Based on the expert judgement, the questions were classified into two classes. The first class is labelled as *"bad"* and it contains questions that are incorrect or somehow unclear. In other words, the class contains questions that should have been corrected by the corresponding student. The second class is labelled as *"good"*, and it contains questions that were good quality, and did not need modifications.

Classification of the questions was done based on 1) the average quality rating given by the experts, and 2) the correct answers selected by the experts. All the questions whose average quality rating was below 2 were classified to the class *bad*. In addition, the questions whose average quality rating was exactly 2, and the experts did not agree with the correct solution were classified to the class *bad*. All the other questions were classified to the class *good*.

4.1.2. Feedback from the students

The student feedback was collected at the end of the Data Structures and Algorithms course in 2010. Students were asked to give a numeric grade from 1 to 4 for the usefulness for their learning. They rated the answering to questions and the creation of new questions separately.

Students were also able to give open-ended feedback on PeerWise and other tools used on the course. In order to study the student attitude, the open-ended feedback on PeerWise was categorized and examined. First, the first author of this paper read the feedback and created the categorization based on the comments on PeerWise. Then the second author used the same categorization to classify the same comments again.

5. Results

From the 48 questions in the question set, ten questions were classified to class *bad* and 38 questions were classified to class *good* by the experts. PeerWise spotted five out of the ten bad questions. In the list of unanswered questions, PeerWise shows an icon representing the suitability of the question. If the average rating of the question given by the students is below 2 or the answer provided by the authoring student is not the most popular answer, the question is tagged with a red exclamation mark meaning that there might be a problem with the quality of the question. Thus, five out of the ten bad questions were tagged with an exclamation mark in PeerWise. However, another five bad questions were missed by PeerWise.

The results of the classification is shown in Figure 2. In the figure, the suitability is calculated as:

$$\text{suitabity: } \frac{A_{count}}{A_{count} + D_{count}} \qquad (1)$$

where A_{count} is the number of students who agreed with the authoring student and D_{count} is the number of answers that the most popular distractor received. When *suitability* is low (below 0.5), PeerWise tags the question respectively. In Figure 2 the questions that PeerWise considers to be good quality are in the top right corner.

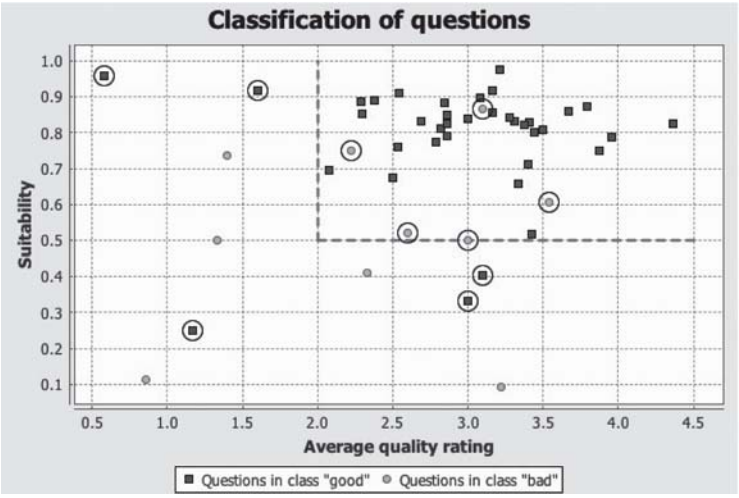


Figure 2. The questions classified to classes "good" and "bad" based on the data from the expert board. PeerWise tags a question to be possibly low quality if the average rating is below 2 or the answer selected by the author is not the most popular one (suitability < 0.5). All the questions whose classifications made by the experts and PeerWise do not match are circled.

Five of the questions that were classified to class *bad* based on the data from the experts were not noticed by PeerWise (circled dots). Similarly, five questions that the experts considered to be high enough quality were unnecessarily tagged with an exclamation mark by PeerWise (circled squares). In addition, there are 38 questions that both the experts and PeerWise considered to be good quality (uncircled squares). Moreover, the uncircled dots are those questions that both considered to be low quality.

5.1. Feedback

On the Data Structures and Algorithms course in 2010, we looked into the course feedback in order to find out whether the students liked better the creation of questions or answering questions made by their peers. In the following, we report on both the numerical feedback and answers from open-ended questions.

In the numerical feedback (1–4 in Likert scale), we asked the students to indicate how beneficial they find 1) answering and 2) creating new questions in the PeerWise system. According to the data, the students considered answering questions (average 2.49) to be a little bit more beneficial than creating new questions (2.34).

All the open-ended feedback related to PeerWise was categorized into positive and negative feedback. Positive feedback was then divided into three subcategories that are: *general, creating questions* and *answering questions*. The *general* category includes all the feedback that were positive but did not mention any specific part of PeerWise to be useful. The *creating questions* category includes all the feedback that considered the creation of questions to be especially useful in PeerWise. Similarly, the *answering questions* category has all the feedback that considered answering to other students' questions to

be useful. Negative feedback was divided only into two subcategories that are: *general* and *quality*. The *general* category includes the feedback that were negative but did not specify any specific reason for the feedback. The *quality* category contains the feedback that criticize the quality of the questions in the repository.

The open-ended course feedback contained 14 comments. After some negotiation, both authors agreed on the following classification of the comments. 11 comments were positive and 3 comments were negative. The *positive general* subcategory had 9 comments, *question creation* had 2 comments, and *answering questions* did not have comments at all. The *negative general* subcategory had 2 comments and the *quality* subcategory had one comment in it.

6. Discussion

In this paper, we have studied a system called PeerWise that enables collaborative learning. The system provides a learning environment in which students can create as well as rate and comment multiple-choice questions created by their peers. The system is originally from Auckland, New Zealand. The original developers have studied the system earlier and find it to be a novel addition to the arsenal of teaching and learning tools. They have concluded that most of the questions created by students are good quality and enhance learning. However, a small number of questions can be low quality.

6.1. Feedback

Although, low quality questions exist, those are not necessarily bad for learning. The course feedback in our study revealed that some of the students perceived misconceptions, but only after their peers addressed that their own question was flawed. Thus, students can correct their way of thinking not only by answering questions, but also by writing low quality questions that they need to correct after some feedback. Following, there is an example of an open-ended feedback from the Data Structures and Algorithms course, which indicates that finding an error can also be useful for learning: *"PeerWise questions worked well – finding an error from own/others' questions even more educational than solving them."*

Based on the numerical ratings, it can be seen that the students considered answering questions to be more useful than creating new questions. Similar results were also reported by Denny et al. [4] in a previous study. However, in the open-ended feedback, the creation of questions was mentioned to be useful in two comments, when answering questions was not mentioned at all. In our opinion, it is highly likely that the usefulness of the system depends on the way students use it. For example, if a new question is created with only a little effort, it is likely that it is not very educational for the student.

6.2. Quality of the questions

Since recognizing faulty questions is important, one of the research questions was how to detect such flawed questions. PeerWise can detect some of the questions based on the low rating a question receives. Also, the comments written by peers can help a student to correct his or her question. However, a small fragment of questions remains that is never

corrected during a course. We have studied how to detect such questions and how to improve the system in order to better facilitate the learning process.

First, the role of course integration should not be neglected. We need mechanisms to quarantee that the students visit the system often enough to recognize their possible flawed questions. This can be achieved partially by paying attention to how the system is utilized in a course. Typically, multiple deadlines are needed to force students to come back and notice the quality of their questions. In addition, the system could send email if the quality of questions is detected to be too low. This, however, rises the question how to recognize such low quality questions.

In his Master's thesis, Hakulinen [7] compared several methods to rank questions according to their quality. The marking scheme used in the current version of PeerWise can detect only a limited number of flawed questions. Hakulinen has suggested another scheme that could improve the system to recognize more questions. However, we believe that the current feedback form in PeerWise could be improved to gather more information in this sense to improve the scheme even further. The current system has six steps as follows:

1. a student reads a question and selects one of the alternatives
2. correct answer (according to the authoring student and the majority of other students) is shown
3. descriptions of the authoring student are shown
4. comments by other students are shown
5. the student can give ratings for the question
6. the student can comment or agree/disagree with other students comments

We suggest to change the above so that, in addition to the comments, the student can change his or her opinion about the correct alternative. In addition, at this last phase, the student should have the option to choose several of the alternatives in case the question was originally flawed and more than one answer could be considered to be correct. Moreover, a new alternative should appear among the possible choices: none of the above. This would make it easier to perceive flawed questions as the problem could be deduced by examining the quantitative data only (and not the qualitative comments). This would help at least the automatic testing of the quality of the questions, but could also provide new information for the authoring students about how other students have accepted his or her question.

7. Conclusion and further development

The quality of the student-made questions in PeerWise seems to be good. However, we have reasons to believe that the quality of the questions and the feasibility of the system can be enhanced even further. Several issues arised in our study to be further studied. We have suggested a number of improvements for the system to tackle these issues.

The course integration is a critical task that need to be carefully planned. We suggest to set several deadlines for a single course in order to force students to visit the system often enough. This would improve the quality of questions as the student can see the other

students comments, and thus can made corrections in his or her own questions. In addition, the system itself could be developed further by implementing new feedback forms and by notifying students about possibly flawed questions. One suggestion is to send email if the rating of a questions goes too low. However, this requires that the system could recognize flawed questions easily. This could be achieved by improving the feedback form to allow the students to change their opinion about the correct choice after answering. If the system is used for summative evaluation, changing the opinion is not a problem. Both, the first choice as well as the last choice, can be saved to be used in evaluation.

In this paper, we concentrated on the usage of PeerWise on the Data Structures and Algorithms course. However, PeerWise is a system that is not tied to a specific field of education. We believe that the results represented in this paper are relevant regardless of the educational discipline. In PeerWise, learning is not solely based on answering and creating questions, but also on collaborating and discussing the context of the questions with other students. In order to make the most of it, students should be motivated to take advantage of the collaborative nature of PeerWise.

References

[1] Denny, P., Hamer, J., Luxton-Reilly, A., and Purchase, H. PeerWise: students sharing their multiple choice questions. In *Proceeding of the fourth international workshop on Computing education research*, pages 51–58. ACM, 2008.

[2] Denny, P., Hanks, B., and Simon, B. Peerwise: replication study of a student-collaborative self-testing web service in a u.s. setting. In *SIGCSE '10: Proceedings of the 41st ACM technical symposium on Computer science education*, pages 421– 425, New York, NY, USA, 2010. ACM.

[3] Denny, P., Luxton-Reilly, A., and Hamer, J. The peerwise system of student contributed assessment questions. In *ACE '08: Proceedings of the tenth conference on Australasian computing education*, pages 69 –74, Darlinghurst, Australia, 2008. Australian Computer Society, Inc.

[4] Denny, P., Luxton-Reilly, A., and Hamer, J. Student use of the peerwise system. In *ITiCSE '08: Proceedings of the 13th annual conference on Innovation and technology in computer science education*, pages 73 –77, New York, NY, USA, 2008. ACM.

[5] Denny, P., Luxton-Reilly, A., Hamer, J., and Purchase, H. Coverage of course topics in a student generated mcq repository. In *ITiCSE '09: Proceedings of the 14th annual ACM SIGCSE conference on Innovation and technology in computer science education*, pages 11–15, New York, NY, USA, 2009. ACM.

[6] Denny, P., Luxton-Reilly, A., and Simon, B. Quality of student contributed questions using peerwise. In M. Hamilton and T. Clear, editors, *Eleventh Australasian Computing Education Conference (ACE 2009)*, volume 95 of *CRPIT*, pages 45 –53, Wellington, New Zealand, 2009. ACS.

[7] Hakulinen, L. Using computer supported cooperative work systems in computer science education – case: Peerwise at tkk. Master's thesis, Faculty of Information and Natural Sciences, School of Science and Technology, Aalto University, April 2010.

[8] Kittur, A., Suh, B., and Chi, E. H. Can you ever trust a wiki? : impacting perceived trustworthiness in wikipedia. In *CSCW '08: Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 477– 480, New York, NY, USA, 2008. ACM.

[9] Kreijns, K, Kirschner, P.A., and Jochems, W. Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in human behavior*, 19(3):335 –353, 2003.

[10] Orlikowski, W. J. Learning from notes: organizational issues in groupware implementation. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 362 – 369, New York, NY, USA, 1992. ACM.

[11] Suh, B., Chi, E. H., Kittur, A., and Pendleton, B. A. Lifting the veil: improving accountability and social transparency in wikipedia with wikidashboard. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1037–1040, New York, NY, USA, 2008. ACM.