This is the accepted version of a paper presented at *2017 ACM Conference on Innovation and Technology in Computer Science Education, Bologna, Italy, 3-6 July, 2017*.

Citation for the original published paper:

https://doi.org/10.1145/3059009.3059059

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-68419
How Tool Support and Peer Scoring Improved Our Students’ Attitudes Toward Peer Reviews

Daniel Toll  
Dept. of Computer Science  
Linnaeus University  
Kalmar, Sweden  
daniel.toll@lnu.se

Anna Wingkvist  
Dept. of Computer Science  
Linnaeus University  
Växjö, Sweden  
anna.wingkvist@lnu.se

ABSTRACT
We wanted to introduce peer reviews for the final report in a course on Software Testing. The students however had experienced issues with peer reviews in a previous course which made this a challenge. To get a better understanding of the situation, we distributed a pre-questionnaire to the students and 48 of the 83 students provided their expectations on peer reviews. To deal with some of the perceived issues, we developed a peer review tool where we introduce anonymity, grading of reviews, teacher interventions, as well as let students score and comment on the reviews they receive. In total, 67 reports were submitted by 83 students and 325 reviews were completed. We also distributed a post-questionnaire and this was answered by 48 students. Results from our study show that 27 students expected incorrect feedback but only 13 students agreed to have got incorrect feedback. The students also reported that they found the feedback from their peers more valuable (+15%) than expected, and 88% of the students think they learned from doing peer reviews. There are still some issues with peer reviews but overall we find that the students’ attitudes toward peer reviews have improved.

KEYWORDS
Peer Review, Peer Grading, Software Testing, Courseware

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

For more information, visit ACM’s Copyright and Permissions page at http://www.acm.org/publications/policies/copyright.jsp.

1 INTRODUCTION
The original idea was that we were interested in using peer review in a Software Testing course (7.5 ECTS). The course was run on 50% study pace over ten weeks during late autumn 2016. Our Computer Science students write test plan reports as a final assignment and these reports have in previous iterations of the course been read and graded by teachers. However, with more than 80 students, we felt that reviewing these reports will be too costly with respect to time. Since we knew that the students had experienced peer reviews in a previous course and expressed some issues related to this, we needed to investigate further before implementation. They mentioned, for example, issues related to lack of anonymity, feedback, and motivation, but also fears of getting misleading feedback.

The aim to use peer review for this particular course is that students will receive feedback from each other while we as teachers can focus on the introduction of a tool, support the review process, resolve potential disputes, and deal with final grading. We also think the students would benefit from reviewing each other’s work. They will learn to be independent and apply critical thinking since they need to judge the quality of other students test plan reports and motivate their standpoint.

Initially, we see two issues with letting students provide feedback. First, it can be argued that students do not have enough knowledge to review each other’s work. Second, students may not want to put in the required effort to review, i.e., what Hamer et al. [3] call “rogue” reviewers. We handle the first issue by providing an assignment that is specific and well defined, so anyone who had the knowledge to complete the assignment should be able to review reports. We address the second issue by providing an incentive; reviewing other students’ reports can result in a higher grade. This, in combination with multiple reviewers per report, should increase the chance that a report gets properly reviewed and that the overall feedback provided is relevant.

Additionally, students can respond to, and score, reviews they receive. They can rate a review as “Not even an attempt”, “Failed”, “Sufficient”, “Good”, and “Excellent”. If a review is scored as “Failed” by a student, a teacher should intervene and resolve the dispute between the student and the reviewer. Note that the final grading is done by teachers, which is why we refer to the activity done by students as scoring.¹

We present how we introduced a tool for peer review that supports anonymous scoring of reports, the possibility for students to score received reviews, teacher intervention when disputes occur, and teacher grading of both reviews and reports. Knowing that the students had experience with peer review from a previous course and expressed concerns made us cautious and curious. So, to understand the situation better, we wanted to know the students’ expectations before the introduction of our tool. We also want to compare these expectations to their experiences after they tried the tool-supported peer review to see if the issues remained, and if so, to what extent.

We pose two research questions (RQ).

¹In other papers this can be called marking, peer grading, etc
RQ1. What are the students’ expectations of receiving feedback from their peers before and experiences with receiving feedback from their peers after tool-supported peer reviews?

RQ2. Are the students less worried about getting scored by their peers after they have used the peer review tool in a new course?

2 TOOL-SUPPORTED PEER REVIEW

We introduced a tool-supported peer review in a Software Testing course given during the second year of several computer science study programs at our university. The students were presented with a requirement specification and source code for a web server software, and their assignment was to test to what extent the provided requirements were implemented. The test effort is planned, executed, and documented in a report.

When the assignment started, the students formed groups of up to four students. The students then had four weeks to complete the testing and writing up the report. At the end of the four weeks, they submitted the report describing their work including a test strategy, a test plan, manual test cases that should cover the requirements, and the results of their test efforts. 55 of the 83 students who submitted a report decided to work alone and the rest of the students worked in groups.

A total of 67 documents were uploaded by the students, which were all peer reviewed using our tool. The peer review process is described below.

2.1 Review Instruction

Before the reviews could start, the teacher described how these were to be done, wrote an instruction, and decided scoring criteria. The review form was divided into three different scoring criteria. Each of the criteria had a description, a free text input field and a scoring rubric. The scoring rubrics had five different levels: “Not even an attempt”, “Failed”, “Sufficient”, “Good”, and “Excellent”, and each level was described with a set of characteristics.

The first scoring criterion focused on the documents clarity, language, disposition, use of tables, headings, and images. The second scoring criterion focused on the completeness of the report, and the third and final scoring criterion focused on the thoroughness of the test effort. Further, to better explain the review, scoring, and grading process to the students, we recorded a short video on “How to be a helpful reviewer”. The video was released a few days before the review phase, and described the overall process as well as how to review. During the review phase, the teachers supported the students by answering questions and participating in discussions on how to score through the messaging system Slack.

2.2 Uploading and Revising

The students were given individual access to the peer review tool from the course web page. Once they were logged in, they could upload their report, preview how it would appear to the reviewers, and revise it until the upload deadline had passed. We decided to use Markdown for the reports, since students had previously used it to document their source code repositories on GitHub. We wanted an option for the students to create reports with headlines, anchor links, pictures, and tables that could easily be viewed within the tool.

2.3 Peer Reviews

The students began to review each other’s work after the deadline had passed. When a student started the task of doing a new review, the peer review tool selected a random report from the set of reports that currently had the least amount of reviews, excluding the reviewers own report.

To get a passing grade, a student had to review at least two reports, but could review more to increase their chance of getting a higher grade. To get the highest grade, they needed to write at least four reviews. If they wrote more than four reviews, only the four highest graded reviews determined their grade. To remind students at risk of not getting a grade, we emailed all students that had not completed the two reviews one day before the review deadline.

2.4 Review Response

After the review deadline, students were given access to the reviews of their report. They should respond to each review with free text input and score how useful, correct, and thorough each review was. If a report has multiple authors, each of them should respond individually. The students were told to motivate their scoring of the review and that this information was helpful for the teacher that would grade the review.

2.5 Teacher Interventions and Grading

The teachers are responsible for grading and the general quality of the students’ learning experience, so this should be supported by the peer review tool. The tool allows the teachers to browse through the submitted reports, reviews, and review responses. Since several students review each report, the teacher gets an overview of how each report was scored as well as how each of the report’s reviews was scored.

We wanted to make it easy for the teachers to detect and handle potential disputes, so we introduced a number indicator that a teacher can use to decide if an intervention is needed or not. First, if a reviewer decides on a failing grade for a report or if an author marks a review as failed, this is clearly indicated (marked with red in the tool). The teacher can view the review as well as the author’s response and determine if any of the students failed in their effort or if both are deemed to be adequate. A free text input is also provided.

Second, since several students review each report, the teachers’ interface shows the minimum, median, and maximum score for each of the three grading criteria. This allows the teacher to investigate if any review significantly diverges from the median score, and if so, the reasons for the spread. Likewise, when several students collaborated on a report, their views of a review might also diverge.

The teacher sets two grades for each student; one for the report and one for the reviews. These are weighted into a final grade for the complete assignment.

---

1https://slack.com/
2https://daringfireball.net/projects/markdown/
3https://github.com
3 CAPTURE STUDENTS’ EXPECTATIONS AND EXPERIENCE

We distributed a pre-questionnaire (to be answered anonymously) to capture the students’ expectations of peer reviews. Furthermore, to investigate the students’ experiences from using our tool-supported review process, we distributed a post-questionnaire (also to be answered anonymously). The students had completed their reviews when they answered the post-questionnaire, but no teacher interventions or grading had taken place.

We posed six statements in the pre-questionnaire and eight in the post-questionnaire. Answers were given on a five-grade Likert-scale: Strongly agree, Agree, Neutral, Disagree, and Strongly disagree. Statements that investigate before and after opinions are identified with .a in the pre-questionnaire and .b in the post-questionnaire.

S1.a I think my peers can give me valuable feedback on my report.
S1.b I think my peers can give me ‘incorrect’ feedback on my report.
S2.a I’m afraid my peers will give me “incorrect” feedback on my report.
S2.b I’m afraid my peers will give me “incorrect” scores on my report.
S3.a I’m afraid my peers will give me “incorrect” scores on my report.
S3.b I have received “incorrect” feedback on my report.
S4.a I’m afraid I will give “incorrect” scores on another student’s report.
S4.b I’m afraid I will give “incorrect” scores on another student’s report.
S5.a I think the teacher’s feedback is more useful than my peer’s feedback.
S5.b I think the teacher’s feedback is more useful than my peer’s feedback.
S6 I look forward to reviewing other students’ reports.

We also provided a free text option “Please share your view on peer reviews” to let students share their opinions on the use of peer reviews.

For the post-questionnaire we adjusted statements 1 to 5 from the pre-questionnaire to capture the students’ experience after the peer review ended. We also added the same free text option for this questionnaire.

S1.b I have received valuable feedback on my report by my peers.
S2.b I have received “incorrect” feedback on my report by my peers.
S3.b I have received “incorrect” scores on my report by my peers.
S4.b I’m afraid I have scored other student’s report incorrectly.
S5.b I think the teacher’s feedback is more useful than my peer’s feedback.
S6 It was fun to score other student’s reports.
S8 I learned from reading other students reports.

Statements S6 in the pre-questionnaire, and S7 and S8 in the post-questionnaire were adopted from Turner and Pérez-Quinones [8], but reformulated to fit the assignment context.

4 RESULTS

The pre-questionnaire attracted 48 responses. In total, 325 reviews were written for 67 reports by 83 students. On average, a student wrote 4 reviews and received 5 reviews. One student completed 22 reviews! The post-questionnaire was distributed to the students after they responded to their reviews but before the teachers had graded their work. It was distributed to all students that had submitted a report. The number of students that responded was also 48. Note that this is not necessarily the same 48 students, since the questionnaires were answered anonymously. All responses to the statements can be seen in Table 1.

4.1 Received Valuable Feedback from Peers (S1)

In the pre-questionnaire, 28 students (58%) agreed or strongly agreed with that their peers can give valuable feedback (S1.a). 17 students were neutral and only three students disagreed.

In the post-questionnaire responses to S1.b, the number of students that agree to have received useful feedback rose to 35 (73%). Three students mentioned they got better reviews than they expected, and one wrote: “After receiving mostly good and honest feedback my attitude changed to be more positive”.

4.2 Receive Incorrect Feedback from Peers (S2)

In the pre-questionnaire, seven students commented in the free text section that they had experienced getting poor feedback from peers and getting feedback that was contradictory. This is visible in S2.a, where 27 students (56%) agree that they are afraid of getting incorrect feedback. Only five students (10%) disagree. One student remarked that the motivation to do good reviews had dropped since they received poor quality reviews in a previous course.

In the post-questionnaire 13 (27%) students agree that they have received incorrect feedback while 24 (50%) students disagreed or strongly disagreed with this statement (S2.b).

4.3 Receive Incorrect Scores from Peers (S3)

More than half of the students (52%) in the pre-questionnaire feared receiving incorrect scores. One student even expressed that only teachers should score and grade students. In total, 15 (31%) students in the post-questionnaire agreed that they had experienced scoring that did not match their expectations. However, the number of students that disagreed improved from 5 (10%) in the pre-questionnaire to 24 (50%) after experiencing getting scores by their peers.

When it comes to students scoring students, the post-questionnaire revealed in the free text option that students experienced variation in the scores that they got from the reviewers and in one case the student wrote “got correct scores but for the wrong reasons”.

4.4 Give Incorrect Scores to Peers (S4)

22 (46%) students worried about scoring other students incorrectly in the pre-questionnaire. This was reduced to 9 (19%) in the post-questionnaire. Two students mention concerns that their grading depends on how lucky they are to receive good reviews themselves and/or easy reports to review.

4.5 Teachers Feedback is More Valuable (S5)

The number of students that think the teacher’s feedback is more valuable was comparable between the pre-questionnaire (30) and post-questionnaire (27) and thus remained around 60%. Two students mentioned the need for teachers to take a strong role in grading in the post-questionnaire; one of them wrote that he/she has confidence that it will happen in this case.
4.6 Attitudes to Peer Review (S6) & Scoring (S7)
According to the pre-questionnaire, students generally liked peer reviews and considered reading each other’s reports to be good for learning. Three students indicated that peer reviews increased their motivation to do a good job writing the report and that it is fun to do reviews since they get to share their expertise. 23 respondents (48%) looked forward to writing reviews while 13 respondents (27%) disagreed. Nearly identical numbers of students, i.e., 24 of the respondents (50%) thought scoring other students was fun in the post-questionnaire, while 9 students (19%) disagreed or strongly disagreed.

A student in the pre-questionnaire explained that their reason for poor motivation came from the need to invest a lot of time reading and reviewing other’s reports. Since the reviews are taking place around Christmas, one student reported that they felt even less motivated to invest effort during this time of the year.

In the post-questionnaire three students mention that the process did require a lot of work, and one student suggested that they should not be required to do more than two reviews to get a good grade. One student wrote he/she felt motivated by the possibility of getting a better grade by writing more reviews.

4.7 Learned from Reviewing (S8)
In the post-questionnaire, 17 of the students strongly agreed and 25 students agreed to have learned from the process. Thus, in total 88% of the students state they learned from reviewing other students’ reports. Three of the students also emphasized this in their free text responses.

5 VALIDITY
There are three main validity issues related to the way we collect information from our students.

First, in order to track actual changes in attitude for individual students, we need to pair responses from the pre- and post-questionnaires. Our anonymous questionnaire only allows us to study changes for the entire group, and not for individual students. Consequently, since the same students occur in both groups, the samples cannot be considered independent. Statistical tests for ordinal numbers with independent samples can however be used, but with less power to detect significant results. We use Mann-Whitney U test\(^5\) to check for statistical significance.

Second, the populations sampled are not exactly the same, since we only distributed the post-questionnaire to students that submitted a report. So, any attitude changes might result from students who were most against peer reviews dropped out before trying the tool. However, we have seen no such indication and students that dropped out cited other reasons.

Third, in the pre-questionnaire, we ask about the students’ fear of something and in the post-questionnaire we ask about their experience of it occurring. Thus, it can be argued that if a student fears something and it did not occur during the peer review, they might still fear it. The free text responses show several attitude changes for the better, but one student wrote that this experience had not changed his or her (negative) view.

6 DISCUSSION
Bauer et al. [1], Hamer et al. [3], Orsmond et al. [5], Søndergaard [7] have all investigated the drawbacks and benefits of peer reviews in the context of computer science education. These all investigate students’ attitude after the review process. We also investigate whether the students’ attitudes toward the peer review process were changed.

Søndergaard [7] used online peer reviews in a course on compiler construction, where students had to review source code written by other students. He reports that students found the peer review task very valuable and that it particularly helped them to better understand their own progress in the course. 79% felt that other students’ feedback was useful. His students had the non-mandatory option of commenting on the reviewer’s reviews, similar to what we do in our study. The student’s view of this option was however not investigated in detail. In our study, only two students disagreed to have received valuable feedback (S1) while 72% of the students agreed or strongly agreed. This is an increase of 15% compared to the pre-questionnaire. Therefore, we think our students found value in each other’s feedback. This is close to the findings of Bauer et al., where students expressed that they found their peers competent, thorough, and helpful. Bauer et al. [1] investigated how students perceived online peer reviews in a computer science course on scientific writing. In a questionnaire, the students reported that online peer reviews were a very positive experience, that reading other students’ work changed their perception of their own work, and that they became more confident in the quality of their own work. Similarly, our students reported that they felt it useful to see how other solved the same problem as they had been working on.

Smith et al. [6] used peer reviews as part of a testing course. Their students, like ours, submitted double-blind reviews and wrote a report on what they learned from the review process and from evaluating the received peer reviews. However, in their case, the teachers were grading the reviews without peer scoring. Smith et al. found that 90% of their students reported to have learned from

---

\(^5\)This can also be called a Wilcoxon rank-sum test

---

Table 1: Responses to the pre- and post-questionnaire. N=48

<table>
<thead>
<tr>
<th></th>
<th>S1.a</th>
<th>S1.b</th>
<th>S2.a</th>
<th>S2.b</th>
<th>S3.a</th>
<th>S3.b</th>
<th>S4.a</th>
<th>S4.b</th>
<th>S5.a</th>
<th>S5.b</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly agree</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>12</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>Agree</td>
<td>21</td>
<td>30</td>
<td>23</td>
<td>12</td>
<td>21</td>
<td>13</td>
<td>20</td>
<td>7</td>
<td>18</td>
<td>17</td>
<td>20</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Neutral</td>
<td>17</td>
<td>11</td>
<td>16</td>
<td>11</td>
<td>14</td>
<td>17</td>
<td>19</td>
<td>18</td>
<td>12</td>
<td>15</td>
<td>13</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Disagree</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>22</td>
<td>7</td>
<td>15</td>
<td>5</td>
<td>18</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Strongly disagree</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

---
Our results and the other studies cited above clearly suggests that there are benefits of using online peer reviews. There are however a number of problems that needs to be addressed, especially when the students are also scoring each other’s as pointed out by Hamer et al. [3]:

1. Assignments and reviews have to be distributed and collected in an efficient way.
2. Anonymity must be ensured.
3. Validity and reliability of the scoring must be ensured.
4. Scoring disputes must be dealt with.
5. The students must be motivated to take the review task seriously.
6. The effect of rogue reviewers, i.e., reviewers who do not take the task seriously and submit bad and misleading reviews, must be minimized.
7. Plagiarism must be detected and handled.

Problem 1 is addressed by our online tool for distributing and submitting reviews. Regarding anonymity (problem 2), 13% of the respondents of the questionnaire used by Bauer et al. [1] would have judged differently if they were absolutely anonymous. We ensure double-blind anonymity in our tool and think it allows the students to write what they think.

According to Moskal et al. [4], scoring validity (problem 3) can be achieved using scoring rubrics. Scoring rubrics means that instead of giving one total score on what is reviewed, the reviewers rate their level of agreement on a number of aspects using, e.g., a Likert-scale. Moskal et al. [4] point out several benefits of using scoring rubrics:

- Students know how they will be scored before submitting their assignment, typically resulting in higher quality submissions.
- It can help teachers track the knowledge development of students over time.
- The total score on a submission differed only slightly between different reviewers, increasing the reliability of the peer review task.

Falchikov [2] argues that peer scoring with marks has little meaning if the peer who scores has no knowledge about the standards. Using scoring rubrics helps the students know what standard is expected of their submissions.

We used scoring rubrics to addresses problem 3. We also used statistical tests to analyze if the use of our rubrics was adequate. In the pre-questionnaire, 25 students (52%) expected to receive incorrect scores (S3), however only 15 students (31%) experienced it. In Figure 1, we see that more students disagree to have received incorrect scores than expected it in the pre-questionnaire. A Mann-Whitney U Test suggests a statistically significant difference in that fewer students experienced incorrect scoring than expected it \((U = 473, p < 0.01)\). We investigated the students free text answers, and they state that receiving variation in scores lowers their confidence for using peer reviews, which we interpret as the scores given were either not properly motivated or students simply did not agree with the motivations. Orsmond et al. [5] studied peer scoring and found that clear scoring criteria is very important to get consistent results. Therefore, we think that our scoring rubrics could be made more specific and with better instructions to increase the reliability of the peer scoring. However, it was noted that our students gained confidence in giving scores and fewer students experienced problems scoring other students than expected it (S4). A Mann-Whitney U Test suggests a statistically significant change \((U = 489.5, p < 0.01)\).

Note that teacher interventions, i.e., where incorrect scores were removed, happened after the post-questionnaire was completed. We expect that the students’ experience concerning scoring will improve when scoring disputes are resolved by the teachers.

To motivate the students (problem 5), the peer review task (minimum of doing two reviews) was mandatory to pass the assignment and students could increase their chances of getting a higher grade by doing more reviews.

Our strategy to minimize the effect of rogue reviewers (problem 6) is to let the authors of a report score each review. This score guides the teacher to investigate the review. A bad review can be disqualified and this can be seen by both author and reviewer. With this transparency we hope to ensure the validity and reliability of the review process (problem 3) to the students. This strategy also addresses the problem of scoring disputes (problem 4). If scores differ much for a submission, the teachers can investigate the reasons for this and suggest a consensus.

The expectations of receiving incorrect feedback was high (S2) with five students disagreeing and 27 students agreeing to this in the pre-questionnaire. The number of students who agreed with having received incorrect feedback in the post-questionnaire was 13. A Mann-Whitney U Test suggests that fewer students experienced incorrect feedback than expected it \((U = 462, p < 0.01)\). Half of the post-questionnaire respondents disagree with having experienced incorrect feedback. This can also be seen in Figure 2, where a shift towards disagreement is evident. Therefore, we think that the
Figure 2: More students expected to receive incorrect feedback from their peers (pre) than experienced it (post).

scoring of reviews motivated our students to spend more time and to write better reviews.

The total number of students that agreed with having experienced either a bad review or an incorrect score (or both) from one of their peers was 20 (41%). This shows that these problems are still widespread and that a large number of students experience them. The teachers resolved conflicts between authors and reviewers 19 times (28%) on 17 different documents. In total, the teachers detected, inspected, and resolved 78 feedback and grading issues for 40 (60%) of the submitted reports. Therefore, we hope that the impact of those bad reviews were minimized. When we compare the students responses on “if the teachers feedback is more valuable” (S5), we can see no significant difference. However, this is to be expected since the students had not given any feedback when the post-questionnaire data was collected.

When we investigate the students’ attitude towards peer review (S6) and scoring (S7), only 50% were positive. Considering the tool and our efforts, we expected better numbers. But, as one student mentioned, the reviews took place over Christmas and involved a lot of work therefore we think this could be the reason why only 50% were positive.

Plagiarism (problem 7) is currently not dealt within our review tool. We do however have plans of considering it in the future.

7 CONCLUSIONS

From prior knowledge, answers, and analysis of statements S1, S2, S5, S6, and S8 as well as from the free text we find that the student group had experienced problems with peer reviews before taking part of the tool-supported peer review process. It was clear that many students expected problems with bad feedback (RQ1). A third of the students experienced at least one bad review. However, all students got a minimum of three reviews and on average five reviews, so it is unlikely that a student got only bad reviews.

The responses to statements S3, S4, and S7 showed that students had concerns on giving and receiving scores from their peers. The post-questionnaire results indicate that fewer students actually experienced problems (RQ2). However, students experienced a variation in their scores, and a third of the students reported that they got at least one incorrect score. We therefore conclude that we need to improve our instructions and scoring rubrics.

We find that the overall attitude toward peer reviews did improve and students experienced fewer issues than they expected.

8 FUTURE WORK

While scoring reviews seemed to motivate our students, we are concerned that students thought it was easier to get a high grade on their reviews if the report they reviewed had obvious flaws. We plan to investigate the factors that affected the students’ grades in more detail using data collected with the tool. We also want to investigate different means of improving the review process, such as adding a checklist to the review instructions. This was suggested by one student in the post-questionnaire.

We aim to improve our research method in three ways for our next study. First, we want to make sure the pre- and post-questionnaires are comparable. Second, we aim to make sure the questionnaire respondents can be tracked anonymously between the pre- and post-questionnaire, perhaps with a unidirectional code. Third, we want to separate questions into tool-, process-, and learning benefit-statements. Since many students reported that peer reviews stimulated their learning it would be interesting to investigate more in detail what learning benefits students actually perceived.

Finally, to improve the peer review system itself we plan to include handling plagiarism and support more languages and formats.

ACKNOWLEDGMENTS

We want to express our appreciation to Johan Hagelbäck who provided valuable feedback to the research. We are also grateful to Morgan Ericsson who moderated this paper and in that line improved the text significantly.

REFERENCES