Technology-Enhanced Peer Review: Benefits and Implications of Providing Multiple Reviews

Pantelis M. Papadopoulos[1*], Thomas D. Lagkas[2] and Stavros N. Demetriadis[3]
[1]Aarhus University, Aarhus, Denmark // [2]The University of Sheffield, International Faculty, CITY College, Thessaloniki, Greece // [3]Aristotle University of Thessaloniki, Thessaloniki, Greece // pmpapad@tdm.au.dk // T.Lagkas@sheffield.ac.uk // sdemetri@csd.auth.gr
[*]Corresponding author

(Submitted January 12, 2016; Revised May 23, 2016; Accepted June 26, 2016)

ABSTRACT

This study analyses the impact of self and peer feedback in technology-enhanced peer review settings. The impact of receiving peer comments ("receiver" perspective) is compared to that of reaching own insights by reviewing others' work ("giver" perspective). In this study, 38 sophomore students were randomly assigned in two conditions and engaged in peer review activity facilitated by a web-based learning environment asking them to provide multiple reviews. In the Peer Reviewed (PR) condition students both reviewed peer work and received peer comments for their own work. By contrast, in the Self Reviewed (SR) condition students provided peer reviews, but did not receive any. Instead, they were asked to perform self reviewing, before proceeding to any revisions of their work. Result showed that the two groups were comparable in all aspects, suggesting that the lack of getting peer reviews can be efficiently alleviated by other type of scaffolds such as a scripted self review process. Overall, the study provides evidence that the review "giver" perspective (as opposed to the typical "receiver" perspective) is a vital option and has noteworthy implications for the design of technological systems that aim to flexibly support more efficient peer review schemes.

Keywords

Peer review, Free selection, Technology-enhanced learning, Self review

Introduction

Peer review is a widely used instructional approach that has been proven valuable for assisting students in both the acquisition of domain-specific knowledge and the development of domain-independent skills. McConnell (2001) argued that peer reviewing offers to students the opportunity for a constructive and collaborative learning experience, by engaging them in an active learning exercise. In its most common form, peer review entails four steps orchestrated by a teacher: (a) production of the initial student work, (b) assigning of reviewers, (c) feedback production, and (d) revisions. Each step poses additional research questions, while the literature abounds with variations of the method, each focusing on different learning goals.

Peer review can be implemented with or without technology, although the former enables the teacher in applying more complex instructional designs, by addressing larger audiences, supporting multiple reviews, and managing the process more efficiently. As the use of technology gets established in the typical classroom, the number of studies in the literature that use technology to support peer review activities increases steadily.

Our work is situated within the technology-supported peer review domain, focusing on settings where students play both the roles of reviewers and reviewees and provide multiple reviews, being also free to choose the peer work they want to review. We refer to this peer review method as "free selection." Previous work showed that such freedom resulted into a greater amount of feedback produced, while increasing students' level of engagement (Papadopoulos, Lagkas, & Demetriadis, 2012). Several learning environments support multiple reviews, with varied levels of freedom allowed during the selection process (e.g., Denny, Hamer, Luxton-Reilly, & Purchase, 2008; Tsai & Liang, 2009; Tseng & Tsai, 2007; Cho & Schunn, 2007). Although researchers may control students' free selection, an interesting research question that rises is what happens to the students that due to free selection do not receive any peer feedback. In other words, what are the benefits of providing peer reviews, and are they adequate to compensate for the loss of peer review comments?

This study provides new evidence on how student performance is affected by providing and receiving peer comments, thus informing both instructors and designers of technology-enhanced learning environments that support peer review.

In order to assist students that do not receive reviews, we asked them to perform self review, by explicitly reflecting on their answers and comparing them with others'. This approach was chosen because it is easy to

implement and it is based on students' capacity and not on peers or the instructor. Regarding the effectiveness of the approach, articulating one's own thinking and producing explicit external representations to reflect upon has been proven beneficial for learning (Jang, 2007; Nückles, Hübner, & Renkl, 2009). Although, self-reviewing could be performed by all students internally, making the process explicit is expected to increase critical reflection and support deeper understandings.

Theoretical background

Peer review

Peer review is a widely implemented didactic model, often supported by technological systems that lift the overhead imposed to the instructor. As the affordances of such systems are based on research evidence relevant to the value of various learning interactions that occur during the peer review activity, the current study focuses on providing additional evidence on the issue of giving vs. receiving feedback that is fundamental for the design of such systems.

Peer review is placed within the broader socio-cultural theoretical context (the "Vygotskian" approach) emphasizing the peer dialog and interaction as a key learning mechanism. Within the peer review model the "peer dialog/interaction" is typically implemented in an indirect asynchronous mode as peers are required to interact by reviewing each other's work. The method is associated with higher-level learning skills, such as synthesis, analysis, and evaluation (Anderson & Krathwohl, 2001) as the students have the opportunity to analyse and evaluate peer work. Scardamalia and Bereiter (1994) have provided evidence that higher cognitive processes of learning are stimulated and guided by the peer review procedure, by implementing the method into school classes. The feedback provided through peer reviewing could be of greater quantity than the one provided by a busy instructor (Silva & Moreira, 2003; Wolfe, 2004), while the process of analysing peer work can support the development of students' self-evaluation skills (Davies & Berrow, 1998), and improve their attitudes and self-efficacy (Anewalt, 2005).

The method has been used extensively in various fields such as second language writing (Hansen & Liu, 2005; Lundstrom & Baker, 2009; Rouhi & Azizian, 2013), writing instruction and relevant courses at the college level (Haswell, 2005), statistics (Goldin & Ashley, 2011), psychology (Cho & MacArthur, 2010), and computer science (Liou & Peng, 2009; Luxton-Reilly, 2009).

Researchers stress the fact that peer review offers to students the chance of developing a range of skills important in the development of language and writing ability, such as meaningful interaction with peers, a greater exposure to ideas, and new perspectives on the writing process (Hansen & Liu, 2005; Lundstrom & Baker, 2009). Certain studies support the use and adoption of peer review of writing (e.g., Cho & Schunn, 2007; Cho, Schunn, & Charney, 2006), emphasising that when students get peer feedback and revise their written work, they improve their writing skills (Cho & MacArthur, 2010).

Technology-supported peer review

Peer review requires a significant volume of information exchange and may pose a difficult administrative overhead for the instructor. Technology can lift this overhead, by distributing material, collecting student work, granting access to peer work, guiding students in the review process, and providing a comprehensive picture to the instructor that orchestrates the process. Arguably, one of the greatest benefits of using technology in peer review is to have students perform multiple reviews. Addressing this issue, many studies used technology-enhanced learning environments and explored the benefits emerging from increasing the number of peer assessors (e.g., Tsai & Liang, 2009; Tseng & Tsai, 2007), by comparing single versus multiple peer reviews (Cho & Schunn, 2007). For example, Cho and Schunn (2007) reported that students that received feedback from multiple peers in the SWoRD web-based peer review system improved their writing quality more than students that received feedback from a single expert.

Systems such as Wolfe's (Wolfe, 2004), PeerWise (Denny, Hamer, Luxton-Reilly, & Purchase, 2008), and curriculearn.dk (http://www.curriculearn.dk) support peer review settings without limiting the number of reviews a student can perform. In such systems, results showed that students with higher grades tend to contribute more than weaker students, resulting in a greater amount of higher quality feedback being produced (Luxton-Reilly, 2009).

The literature abounds with studies that use technology-enhanced learning environments around peer feedback. For example, Silva and Moreira (2003) supported peer interaction through the WebCoM system, while Liu and Tsai (2005) used web-based peer assessment to support conceptual awareness. Obviously, technology can enhance an already flexible and powerful instructional tool such as the peer review. However, the affordances of technological systems do have a significant impact on the efficiency of implementing any specific version of peer review technique. Thus, informed design approaches for improving learning systems to better support peer review-based learning are still in need.

Giving vs. receiving feedback

Although peer review is a well-known and in-depth explored learning technique, investigating the benefits emerging for peers from providing reviews – as opposed to simply receiving – is currently a critical issue in peer review literature. For example, Dunlap and Grabinger (2003) argued that reviewing someone else's work can be beneficial for the student in reflecting on and articulating own views and ideas, thus eventually improving own work. In addition, Ko and Rossen (2004) stated that while reviewing others, students receive perspectives other than the instructor's and this process could provide further insights.

Researchers refer to "assessors vs. assessees" (Li, Liu, & Steckelberg, 2010) or "givers vs. receivers" (Lundstorm & Baker, 2009; Rouhi & Azizian, 2013) and report different learning outcomes identified for these two different peer roles. In contrast to what is typically considered as the main peer review merit (that is, getting reviews), available research reveals that students who provide reviews to their peers ("assessors" or "givers") reach higher levels of learning gains in comparison to students who typically only receive peer reviews ("assessees" or "receivers"). For example, Li et al. (2010) reported a significant relationship between the quality of students' final projects and the quality of peer feedback the students provided, while no such relationship was identified for the feedback students received. Similarly, Lundstrom and Baker (2009) concluded that students who reviewed their peers' writings were significantly benefited in their own writing, outperforming those students who only received peer feedback. However, studies exploring this issue are considered limited and authors suggested that further research is needed to explore the various scenarios and roles of students in assessor/assessee situations (e.g., Li et al., 2010). Finally, Rouhi and Azizian (2013) presented similar findings, with "givers" outperforming "receivers" in second language writing, emphasizing also that most studies have touched this issue rather superficially without providing strong empirical evidence.

Research motivation, hypotheses, and research questions

Based on the above background, our motivation is to deeper understand the impact of engaging students in different modes of peer-review activity. This, in turn, is expected to generate well-informed suggestions for the design of technological systems that support the activity. The current study focuses on the role that the two distinct processes play in peer review activity: (a) the externally originated review comments that guide students in revising their work (the "receiver" perspective), and (b) the insights developed by students themselves, based on the experience of offering reviews to their peers (the "giver" perspective).

In the study, we compared the performance of two groups. Although both groups provided reviews to peers, students in the Peer Reviewed (PR) group received reviews from fellow students, while students in the Self Reviewed (SR) group did not receive peer reviews and instead they were guided through a self review process. We tested the two following directional research hypotheses:

- H1(revision): "Students in the PR group will perform better in revising their own work than students in the SR group."
- H2(conceptual): "Students in the PR group will perform better in acquiring domain conceptual knowledge than students in the SR group."

In addition, the study also explored research questions regarding the attitudes and review strategies applied by the students and how these affected the learning outcome. More specifically:

- RQ1: How will the students select which peer work to review?
- RQ2: How will the volume and quality of peer work reviewed affect reviewer's performance?
- RQ3: How will the volume and quality of peer comments received affect author's performance?
- RQ4: How will the free selection peer process affect students' attitudes towards the activity?

Method

Participants

The study employed 38 sophomore students (20-21 years old) who volunteered to participate. All students were majoring in Informatics and Telecommunications Engineering in a 5-year study program and were enrolled in the "Network Planning and Design" (NP&D) course. The activity was an optional part of the course, and a bonus grade was awarded to participants who successfully completed all phases. We randomly distributed students into the groups:

- Peer Reviewed: 18 students, 12 males and 6 females
- Self Reviewed: 20 students, 12 males and 8 females

Students were not aware of the distribution and we informed them beforehand of the research nature of the activity and the possibility that some of them would not receive peer reviews.

Learning environment

We have developed our own web-based learning environment that has been used in various domains in the past. The benefit of using a custom-made tool is that we could easily alter the study conditions according to our research interests. Studying in our learning environment typically entails answering open-ended questions of realistic scenarios, grounding the answers on related past cases that present similar problems and the way they were addressed. The learning environment implemented the free selection peer review protocol (described next), granting access to multiple peer work in the review phase.

Design

The study followed a pre-test post-test research design to compare the performance of the two groups. The study had six distinct phases: Pre-test, Study, Review, Revise, Post-test, and Interview. The study conditions were identical for all students throughout the activity, except from the Revise phase, in the beginning of which the PR group received the peer comments and the SR group performed self review. Two instructors of the NP&D course served as raters.

Procedure and study conditions

Pre-test phase

In the Pre-test phase, we recorded students' prior domain knowledge in class using a written test with 6 open-ended questions (e.g., "How can the security requirements of a network affect its architecture?").

Study phase

The Study phase started right after the pre-test and lasted one week. The students logged in the environment (from wherever and whenever they wanted) and worked on the available material, providing answers (in the following: solutions) to open-ended questions of 3 plausible scenarios (3 answers in total). Students had to take into account the specific conditions and the presented context and propose their own computer networks as solutions to the scenarios.

Review phase

Next, in the Review phase that lasted 4 days, the students had to review, in a double-blind process, the solutions their peers gave to the scenarios. All 38 participants had to provide reviews to the solutions of the 18 students of the PR group. In order not to overwhelm the students, for each of the 3 scenarios, participants received peer solutions from a randomly selected subgroup of 9 PR students. The solutions were presented in random order in a "solution grid" (Figure 1) and each grid was unique. However, each solution appeared the same number of times in total, thus having the same chances of getting a peer review.

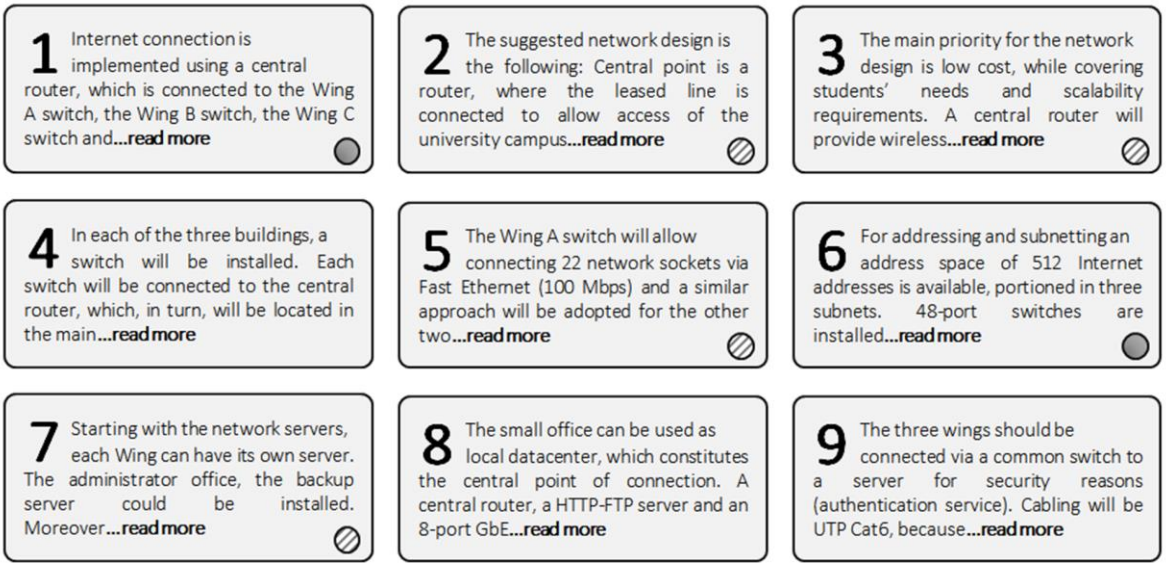

Figure 1. The solution grid of a scenario. According to this figure, the student has read solutions 1, 2, 3, 5, 6, and 7, and has reviewed solutions 1 and 6 of this particular scenario

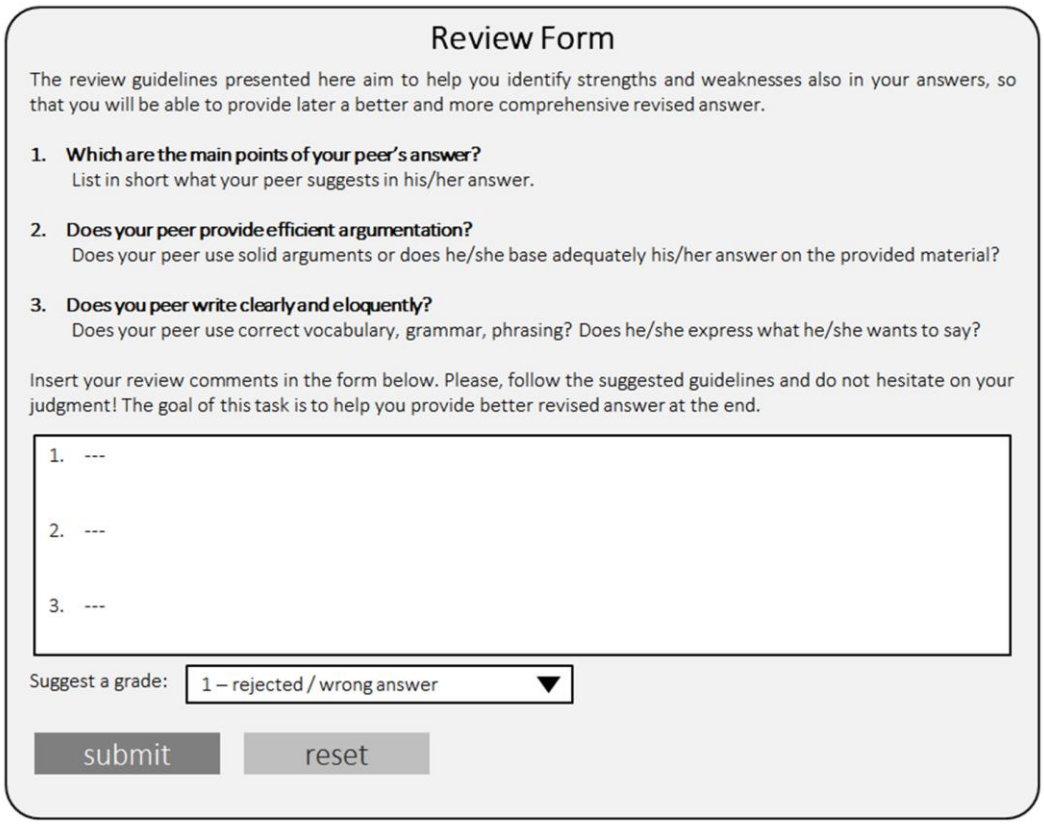

Figure 2. Review form

The students were able to read as many peer solutions they wanted and they had to perform at least one review for each of the 3 scenarios. Of course, it was possible for some solution to receive more than one review, while for others to receive none. The latter, however, did not occur, since the number of reviewers (i.e., PR+SR = 38) was much higher than the number of solution authors (i.e., PR = 18), resulting to all solutions to receive at least one review.

Students had to follow a review microscript, guiding them through the process, focusing on three dimensions: (a) content, (b) argumentation, and (c) clarity (Figure 2). In the content section of the review, students were expected to analyse the main points of their peer's solution, identifying for example the components of the proposed computer network. As students often fail to present adequately the reasoning behind their choices in proposing a

solution, the second guideline focused reviewers' attention on argumentation. In other words, the reviewers had to evaluate whether the author justified each aspect of the network appropriately and if he/she grounded his/her solution on the provided material. The third guideline was focused on the clarity and eloquence of the solution, guiding reviewers into giving advice on the form of the solution. Finally, along with the comments, the reviewer had to also suggest a grade according to the following scale: (1: Rejected/Wrong answer; 2: Major revisions needed; 3: Minor revisions needed; 4: Acceptable answer; 5: Very good answer).

Revise phase

The Revise phase lasted 3 days. Right after the Review phase was completed, the produced feedback was made available to the students by the system. Students in the PR group received the comments their peers had made on their solutions, while students in the SR group did not receive any peer feedback. Instead, the SR group was reminded of the review form, and was prompted to perform self review (minor differences in wording existed between peer review and self review forms, to target one's own solutions). As mentioned earlier, this was the only phase where the two groups worked in different ways. Once the system granted access to the peer comments, the PR group was able to proceed immediately to revise theirs solutions. On the contrary, students in the SR group had to submit self reviews, before they were allowed to revise their solutions.

Post-test phase

At the end of the online activity, the students took a written post-test in class focused on students' acquired domain knowledge, using 3 open-ended questions (e.g., "Which network characteristics are affected by the end-users' profile?"). The answers to these questions were not to be found as such in the study material, but rather to be constructed, by taking into account information presented in various cases.

Interview phase

Shortly after, students from each group were interviewed to record their approaches and comments on the activity. Interviews were semi-structured and focused on students' views on the activity.

Figure 3 presents the phase sequence for each group (the study variables, marked in bold, are presented in the next section).

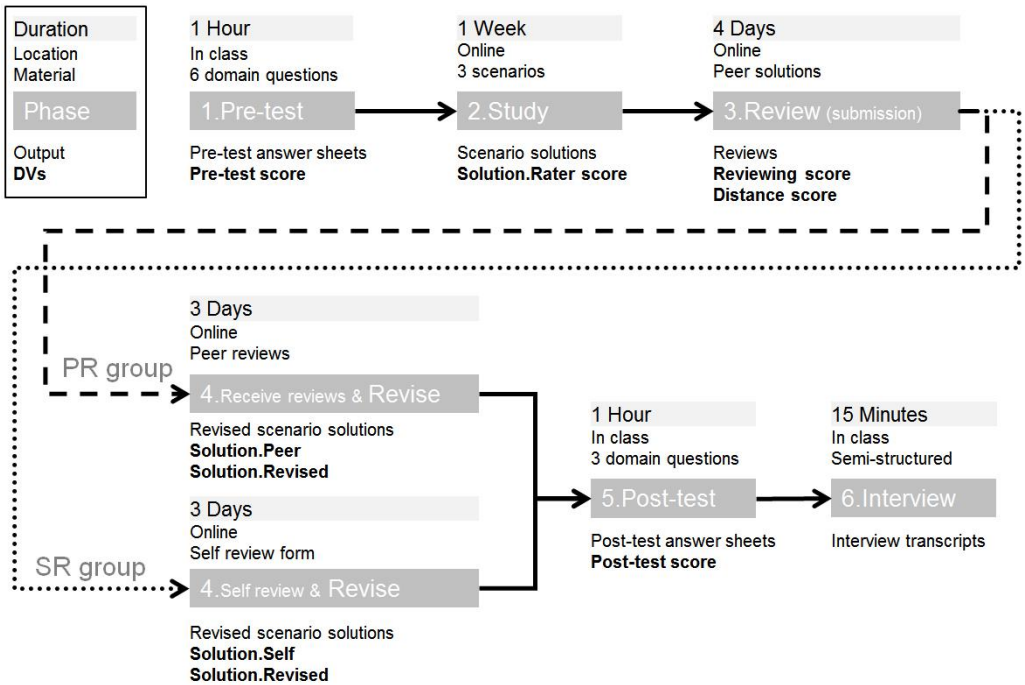

Figure 3. Phase information and sequence

Measures

The type of study conditions the students had before revising their initial solutions was the independent variable (i.e., peer reviewed or self reviewed), while students' performance (in the written tests and in the learning environment) and activity (according to system log files) were the dependent variables. Table 1 presents each dependent variable related to students' performance, along with the scale used and the way the total final score was calculated for each metric.

A 1-10 scale was used for the *Pre-* and the *Post-test* scores, while a 1-5 scale was used for the *Solution.Rater, Solution.Peer, Reviewing,* and *Solution.Revised* scores, to be in line with the scale used by the students in their review process.

Table 1. Dependent variables for students' performance

Score name (scale)	Description
Pre-test (1-10)	The mean score the student received from the two raters for the 6 questions of the pre-test instrument.
Solution.Rater (1-5)	The mean score the student received from the two raters for the initial 3 solutions in the respective scenarios of the learning environment.
Solution.Peer (1-5)	The mean score the student received from peer reviewers for the initial 3 solutions of the learning environment.
Solution.Self (1-5)	The mean score the student assigned to his/her scenario solution in the self-review form.
Reviewing (1-5)	The mean score the student received from the two raters for the quality of each review he/she submitted.
Distance (0-4)	The mean absolute difference between the review scores submitted by the student and the respective scores submitted by the raters.
Solution.Revised (1-5)	The mean score the student received from the two raters for the revised 3 solutions in the respective scenarios.
Post-test score (1-10)	The mean score the students received by the two raters for the 3 questions of the post-test instrument.

In analysing students' review strategies and the impact the review process had on their performance, the analysis included four additional dependent variables based on system log files that were related to students' review strategy (Table 2).

Table 2. Dependent variables for students' review strategy

Variable name (scale)	Description
NrReviewsSubmitted (3-27)[1]	The number of reviews submitted by a student.
NrReviewsReceived (0-42)[2]	The number of reviews received by a PR student.
QltReviewsReceived (1-5)	The mean score of the quality of the reviews (as determined by raters) received by a PR student.
QltSolutionsReviewed (1-5)	The mean score of the quality of peer solutions (as determined by raters) reviewed by a student.

Note. [1]Each student sees a 3x3 solution grid in each of the 3 available scenarios and has to review at least one peer solution per scenario. [2]Each PR student's solution appeared approximately to 14 other students (PR and SR) and each student had to submit one question in the 3 available scenarios.

Data analysis

To avoid any biases, students' paper sheets (pre- and post-test) and system print-outs (scenario solutions and reviews) were mixed and assessed blindly by the two raters, following predefined grading instructions. As a measure of inter-rater reliability, the two-way random average measures (absolute agreement) intraclass correlation coefficient (ICC) was calculated for the raters' scores.

For all statistical analyses, a level of significance at .05 was chosen. Cronbach's Alpha test was used to evaluate the internal consistency of the pre and post-test, while bivariate Pearson's correlation test was used to calculate the respective correlation between the two test instruments. Bivariate Pearson's correlation test was also used to analyse any interaction between the variables describing students' performance (i.e., Table 1) and review strategies (i.e., Table 2).

Finally, each one-on-one interview lasted approximately 15 minutes and was audio recorded. The interview transcripts were used for content analysis.

Results

Student performance

Inter-rater reliability was high for the *Pre-test* (*ICC* = .882), the *Solution.Rater* (*ICC* = .849), the *Reviewing* (*ICC* = .892), the *Solution.Revised* (*ICC* = .880), and the *Post-test* (*ICC* = .843) scores. In addition, internal consistency was high for pre (α = .821) and post-test (α = .866), while the scores of the two instruments were not correlated ($p > .05$). Table 3 presents the results regarding the performance of the two groups throughout the activity.

Table 3. Student performance in the activity

	Peer reviewed			Self reviewed			Total		
Written (scale: 1-10)	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Pre-test	2.04	(1.06)	18	2.20	(1.09)	20	2.12	(1.08)	38
Post-test	8.13	(1.40)	18	8.19	(1.37)	20	8.16	(1.36)	38
Online (scale: 1-5)	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Solution.Rater	2.73	(0.93)	18	2.43	(1.10)	20	2.60	(1.01)	38
Solution.Peer	3.32	(0.77)	18	--	--	--	3.32	(0.77)	18
Solution.Self	--	--	--	3.39	(0.90)	20	3.39	(0.90)	20
Reviewing	3.21	(0.78)	18	3.33	(0.59)	20	3.27	(0.68)	38
Distance (0-4)	1.06	(0.51)	18	0.98	(0.52)	20	1.02	(0.50)	38
Solution.Revised	3.28	(0.77)	18	3.29	(1.05)	20	3.29	(0.91)	38

T-test results did not indicate a significant difference between the two groups ($p > .05$) regarding their prior knowledge (*Pre-test* score), the quality of their initial solutions (*Solution.Rater* score), the quality of their submitted reviews (*Reviewing* score), the distance from raters' score (*Distance* score), and the quality of their revised final solutions (*Solution.Revised* score), therefore suggesting that the two groups were comparable. One-way analysis of covariance also showed that there was no significant difference ($p > .05$) in the post-test performance (*Post-test* score) of the two groups.

In addition, paired-samples t-test results showed that in both groups the revised solutions were significantly improved, when compared to the initial ones (PR: $t[17] = 4.618$, $p = .001$, $d = .66$; SR: $t[19] = 3.669$, $p = .003$, $d = .82$), suggesting that the second week of the activity (Review and Revise phases) had an important impact on students' learning.

Reviewing strategies and bivariate correlations

Usage data analysis showed that most of the students' solutions in the scenarios were approximately half a page long, with the maximum length being one full page. Students read almost all the available solutions ($M = 8.12$, $SD = 1.45$) in the grid of each respective scenario and t-test results showed that the two groups were comparable ($p > .05$) regarding the number of reviews submitted and the quality of the solutions students chose to review (Table 4).

Table 4. System activity metrics

	Peer reviewed			Self reviewed			Total		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
NrReviewsSubmitted (3-27)	5.50	(0.92)	18	5.90	(0.80)	20	5.71	(0.88)	38
NrReviewsReceived (0-42)	10.77	(2.75)	18	--	--	--	10.77	(2.75)	18
QltReviewsReceived (1-5)	3.25	(0.51)	18	--	--	--	3.25	(0.51)	18
QltSolutionsReviewed (1-5)	2.33	(0.82)	18	2.43	(0.65)	20	2.38	(0.72)	38

Regarding the research questions of the student on students' review strategy, Tables 5 and 6 present the bivariate Pearson's correlation test results for the two groups. To increase readability, only the statistically significant correlations are presented. The two groups were comparable presenting the same bivariate correlations. This is in line with the previous results showing that students in the two conditions performed the same.

Table 5. Bivariate correlations in Peer Reviewed group

	Review strategy				Authoring skills					Reviewing skills	
	NrReviewsSubmitted	NrReviewsReceived	QltReviewsReceived	QltSolutionsReviewed	Pre-test	Solution.Rater	Solution.Peer	Solution.Revised	Post-test	Reviewing	Distance
NrReviewsSubmitted	1					.503*				.425+	
NrReviewsReceived		1									
QltReviewsReceived			1								
QltSolutionsReviewed				1		.577*		.644**	.558*	.813**	-.813**
Pre-test					1						
Solution.Rater						1		.806**	.703**	.606*	-.452+
Solution.Peer							1	.612*		.702**	-.675*
Solution.Revised								1	.597*	.636*	-.448+
Post_test									1		
Reviewing										1	-.756**
Distance											1

Note. $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$.

Table 6. Bivariate correlations in Self Reviewed group

	Review strategy				Authoring skills					Reviewing skills	
	NrReviewsSubmitted	NrReviewsReceived	QltReviewsReceived	QltSolutionsReviewed	Pre-test	Solution.Rater	Solution.Self	Solution.Revised	Post-test	Reviewing	Distance
NrReviewsSubmitted	1					.534*				.444+	
NrReviewsReceived		n.a.									
QltReviewsReceived			n.a.								
QltSolutionsReviewed				1		.604*		.671**	.601*	.790**	-.790**
Pre-test					1						
Solution.Rater						1		.800**	.742**	.580*	-.478+
Solution.Self							1	.612*		.702**	-.621*
Solution.Revised								1	.559*	.657*	-.501+
Post_test									1		
Reviewing										1	-.820**
Distance											1

Note. $^{+}p < .10$; $^{*}p < .05$; $^{**}p < .01$.

Table 7 combines and summarizes the most important findings (F#) of the bivariate correlation analysis.

Table 7. Findings of the bivariate correlation analysis

#	Correlated variables	Finding
F1	(Solution.Rater – QltSolutionsReviewed) and (Solution.Rater – NrReviewsSubmitted)	Students that have provided better initial solutions also submitted more reviews, selecting mainly high quality peer solutions to review.
F2	(Solution.Rater/Peer/Self – Reviewing – Solution.Revised)	Students that submitted good initial solutions (as determined by raters, peers, or themselves) also submitted high quality reviews and revised solutions.

F3	(QltSolutionsReviewed – Reviewing – Solution.Revised)	Students that reviewed high quality peer solutions had also better performance in reviewing and revising.
F4	(Reviewing – Distance)	Students that submitted high quality reviews suggested also grades closer to the ones suggested by the raters.
F5	(NrReviewsSubmitted – Reviewing)	The higher the number of reviews submitted by a student, the higher the quality of the reviews. (Tentative)
F6	(Post-test – Solution.Rater – Solution.Revised – QltSolutionsReviewed)	Post-test performance is correlated to the quality of initial and revised solutions, and the quality of solutions reviewed.
F7	(Post-test -/- Reviewing)	No correlation between post-test performance and reviewing.
F8	(NrReviewsReceived -/- any) and (QltReviewsReceived -/- any)	The number and quality of received reviews were not correlated to any other variable in the study.

Interviews

The lack of peer comments in SR group did not raise any major concerns, with students in both groups evaluating positively their experience in the activity and the free selection peer review process. A small number of SR students ($N = 6$) said that they would like to receive peer comments, although they also expressed the belief that their revised solutions would not have been affected much by this.

Results showed that students submitted, on average, almost double the number of minimum reviews required ($M = 5.71$, $SD = 0.88$). Students mentioned the usefulness of forming and articulating a written review in clarifying their own understandings, while the format and length of the solutions in the activity made the review process easier.

Students in PR group found the peer comments generally useful claiming that they decided to revise their solutions after a comment they received. Of course, there were also students that expressed disagreement with reviewers' opinions. Similarly, students in SR group stated that the self review helped them to better analyse their initial solutions, mentioning that the structured process of looking back to their own work after having reviewed others' made their weakness more visible to them. However, many students said that they had taken the decision to revise their solutions even before self review, usually because of something they had read in peers' solutions. More specifically, students mentioned that there were cases in which peer solutions presented arguments which "made more sense" than theirs or "presented the same solution to the scenario, but in a more clear way."

Discussion

Hypotheses testing

The comparison between the two conditions showed that students in the two conditions performed the same in every way. In the light of these results, research hypotheses H_1(revision) and H_2(conceptual) are rejected, suggesting that the process of giving and receiving peer reviews was not enough for the Peer Reviewed group to outperform the Self Reviewed group that gave peer reviews but did not receive any, performing self review instead. As such, alternative hypotheses are stated:

- Ha1(revision): "Students in the PR group will not perform better in revising their own work than students in the SR group."
- Ha2(conceptual): "Students in the PR group will not perform better in acquiring domain conceptual knowledge than students in the SR group."

Research questions analysis

Analysing RQ1, finding *F1* of the bivariate correlation analysis (Table 7) suggests that students' selection strategy was correlated to their own levels of understanding, meaning that strong students tend to review more solutions, selecting *also* higher quality peer work. On the contrary, less competent students selected peer

solutions at their own levels of understanding (weaker solutions), while analysis shows that they reviewed lower numbers of peer solutions, thus indicating minimum effort strategy.

Regarding authoring, reviewing, and revising skills, finding *F2* suggests that authoring an initial solution, reviewing peer work, and revising own work are all correlated. Although correlation analysis does not refer to causality, having in mind the sequence of the study phases, it seems that the level of knowledge students acquire during the first week of the activity is indicative of their performance in the phases to come. In other words, strong students are able to perform well in all the phases, while weak students struggle throughout the activity. This is important, because it dictates the need for an additional instructional intervention that would improve the performance of weak students. Such an intervention could be based on finding *F3* and students' interview comments in which they appreciated clear and well-grounded peer solutions. So, addressing RQ2, our assumption is that students may benefit more when they spend their time reviewing good solutions. We maintain that the deeper insights and better grounded argumentation found in higher quality peer work underlines the discrepancy between strong and weak solutions, and supports the generation of such self-constructed deeper understanding that could trigger weak students to improve their work. Having in mind also *F1*, it appears that weak students tend to select lower quality solutions, thus missing the opportunity to get deeper insights while reviewing. So, returning to the question raised earlier, one beneficial intervention could be the promotion of high quality peer work for reviewing, in an effort to guide weak students to a more useful review process. In our study, for example, this could be done by selecting only high quality solutions in the grids presented to weak students. Of course, this approach needs further validation through future research.

Further in the analysis, findings *F4* and *F5* were expected. Students that, according to the raters, provide better feedback to their peers are expected to also grade their peers similarly to raters, while the students' ability to review peer work is expected to improve as students spend more time reviewing.

Finally, as far as students' performance in acquiring domain knowledge is concerned, findings *F6* and *F7* showed that the scores achieved in the post-test were strongly correlated to the quality of the initial solutions, the quality of the revised solutions, and the quality of solutions reviewed, but there was no correlation between post-test scores and students' skills in reviewing peer work. One explanation for this could be that the post-test instrument was strongly focused on conceptual domain knowledge, and as such it might not be suitable to also assess students' reviewing skills. This does not mean that the review process itself did not have an impact on learning outcomes. Results already showed existing correlations between *Reviewing* score and other dependent variables, and a significant improvement between initial and revised solutions. On the other hand, addressing RQ3, *F8* suggests that neither the number nor the quality of peer comments received have an effect on students' performance.

Regarding RQ4, results showed that the majority of students submitted almost double the number of required reviews. In explaining this attitude, students cited (a) the usefulness of the review process for their own benefit, and (b) the short length of the peer solutions. Students' ability to link the process of formulating and writing an analysis on someone else's work with clarifying their own understandings suggests metacognition. However, a research question is raised here regarding the efficiency of the same review setting for students with lower metacognitive skills. Such question could be the focus of future research on peer review, as the level of metacognition could vary and affect students' engagement in a learning activity significantly. Students' positive attitude towards the review process applied is not new, as it has also appeared in a previous work of ours on defining and comparing the free selection protocol with assigned review settings (Papadopoulos et al., 2012). Furthermore, the students' attitudes are in line with studies that report students' tendency to contribute more when not restricted by the assignment protocol (e.g., Denny et al., 2008; Luxton-Reilly, 2009; Wolfe, 2004). Nevertheless, since all the students in this study volunteered to participate, one should also consider that students' engagement may differ in settings where such activity is obligatory. Finally, students' claim that in some cases the decision for revising their solutions was taken while reading peer solutions, demonstrates the impact the process of providing reviews could have on students' performance.

A possible explanation

In peer review activity, peers are generally expected to deeper understand the domain and improve their contributions based principally on feedback comments offered by other peers. However, in a context where peers are offering multiple reviews they may also get significant benefits from the process of reviewing others' work. The current study provided evidence that the reviews offered by other peers did not help the PR group students to outperform those in SR group who did not receive this kind of comments. Thus, reviews from peers may not

be of substantial help when compared to the insights that students develop, when reviewing others' work and are prompted explicitly to self review their work. We suggest that there are various reasons leading to failure of peer review comments to further improve the PR performance: (a) PR students may ignore peer reviews, doubting their quality, as they do not originate from an expert, (b) poor quality of peer review does not actually help PR students to improve their solutions, and (c) peer reviews are of high quality, but do not actually offer any additional insights for improvement that the PR students could not develop themselves while reviewing others.

Additionally, as Lundstrom and Baker (2009) suggested, review providing is a cognitive activity that can be understood within the framework of socio-cognitive perspective (Vygotsky, 1986) relevant to student-provider's zone of proximal development (ZPD). It seems that reviewing peers' contributions acts as a scaffold that enables student-providers to successfully build understandings within the limits of their own ZPD, thus achieving to improve their performance as authors and reviewers. It might be also possible that this cognitive mechanism is further facilitated by the positive influence of affective and motivational factors, as student-providers develop ownership of their newly constructed understanding and are more willing to accept the corrective feedback implied. This, of course, is a claim open to future research and validation.

Implications

Free selection had positive effects on students' attitudes towards the activity. At the same time, it is clear that not all selection strategies work the same for the students. Based on correlation analysis and students' statements, we maintain that there is a possible connection between reviewing high quality peer work and higher student performance. A design implication would be to promote this high quality peer work for review, especially to weak students, perhaps by allowing students to rate the quality of peer work and designing the learning system in a way that encourages students to review high level work. The way this approach can be implemented without lessening the role of free selection depends on the context of the learning activity.

Conclusions

This study provides concrete evidence showing that in free selection peer review settings, where conducting multiple reviews is possible, the lack of peer reviews can be easily alleviated by supporting students through a simple self review process. We are not suggesting, of course, the elimination of peer comments. Students also expressed positive opinions for some of the peer reviews they received. The question that rises naturally is how the students are going to make use of free selection and how the spread of reviews will cover all the peer work. The instructor can opt to modify the setting in an effort to ascertain that each student will receive peer reviews. However, what the findings of this study suggest is that such a modification may not be necessary. Thus, systems for peer review can be designed to offer "receiving peer reviews" and "practicing self review" as equally important alternatives for the instructor to choose from.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A Taxonomy for learning, teaching, and assessing: A Revision of Bloom's Taxonomy of educational objectives*. New York, NY: Longman

Anewalt, K. (2005). Using peer review as a vehicle for communication skill development and active learning. *Journal of Computing in Small Colleges, 21*(2), 148-155.

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*, 328-338.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A Web-based reciprocal peer review system. *Computers and Education, 48*(3), 409-426.

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication, 23*, 260-294.

Davies, R. & Berrow, T. (1998). An Evaluation of the use of computer peer review for developing higher-level skills. *Computers and Education, 30*(1/2), 111-115.

Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. C. (2008). PeerWise: Students sharing their multiple choice questions. In *Proceeding of the Fourth international Workshop on Computing Education Research* (pp. 51-58). New York, NY: ACM.

Dunlap, J. C., & Grabinger, R. S. (2003). Preparing students for lifelong learning: A Review of instructional methodologies. *Performance Improvement Quarterly, 16*(2), 6-25.

Goldin, I. M., & Ashley, K. D. (2011). Peering inside peer-review with Bayesian models. In G. Biswas. S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial Intelligence in Education* (pp. 90–97). Berlin, Germany: Springer-Verlag.

Hansen, J., & Liu, J. (2005). Guiding principles for effective peer response. *ELT Journal, 59*, 31-38.

Haswell, R. H. (2005). NCTE/CCCC's recent war on scholarship. *Written Communication, 22*, 198-223.

Jang, S. J. (2007). A Study of students' construction of science knowledge: Talk and writing in a collaborative group. *Educational Research, 49*(1), 65-81.

Ko, S., & Rossen, S. (2004). *Teaching online: A Practical guide* (2nd ed.). Boston, MA: Houghton Mifflin.

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525–536.

Liou, H. C., & Peng, Z. Y. (2009). Training effects on computer-mediated peer review. *System, 37*, 514–525.

Liu, C. C., & Tsai, C. M. (2005). Peer assessment through web-based knowledge acquisition: tools to support conceptual awareness. *Innovations in Education and Teaching International, 42*, 43-59.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The Benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*, 30-43.

Luxton-Reilly, A. (2009). A Systematic review of tools that support peer assessment. *Computer Science Education, 19*(4), 209-232.

McConnell, J. (2001). *Active and cooperative learning. Analysis of algorithms: An active learning approach*. Burlington, MA: Jones & Bartlett Publications.

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*(3), 259-271.

Papadopoulos, P. M., Lagkas, T. D., & Demetriadis, S. N. (2012). How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course. *Computers & Education, 59(2)*, 182-195.

Rouhi, A., & Azizian, E. (2013). Peer review: Is giving corrective feedback better than receiving it in L2 writing? *Procedia - Social and Behavioral Sciences, 93*, 1349–1354.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences, 3*(3), 265-283.

Silva, E., & Moreira, D. (2003). WebCoM: A Tool to use peer review to improve student interaction. *Journal on Eduational Resources in Computing, 3*(1). doi:10.1145/958795.958798

Tsai, C.-C., & Liang, J.-C. (2009). The Development of science activities via on-line peer assessment: The Role of scientific epistemological views. *Instructional Science, 37*, 293-310.

Tseng, S., & Tsai, C.-C. (2007). On-line peer assessment and the role of peer feedback: A Study of high school computer course. *Computer & Education, 49*, 1161–1174.

Vygotsky, L. S. (1986). *Thought and language*. London, UK: Longman.

Wolfe, W. J. (2004). Online student peer reviews. In *Proceedings of the 5th conference on Information technology education* (pp. 33-37). New York, NY: ACM Press.