This is a repository copy of *How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course*.

How to Improve the Peer Review Method: Free-Selection vs. Assigned-Pair Protocol Evaluated in a Computer Networking Course

Pantelis M. Papadopoulos [a,*], Thomas D. Lagkas [b], Stavros N. Demetriadis [a]

[a] Aristotle University of Thessaloniki, Thessaloniki, Greece
[b] University of Western Macedonia, Kozani, Greece

Abstract

This study provides field research evidence on the efficiency of a "free-selection" peer review assignment protocol as compared to the typically implemented "assigned-pair" protocol. The study employed 54 sophomore students who were randomly assigned into three groups: Assigned-Pair (AP) (the teacher assigns student works for review to student pairs), Free-Selection (FS) (students are allowed to freely explore and select peer work for review), and No Review (NR) (control group). AP and FS student groups studied and reviewed peer work in the domain of Computer Networking, supported by a web-based environment designed to facilitate the two peer review protocols. Our results indicate that students following the Free Selection protocol demonstrate (a) better domain learning outcomes, and (b) better reviewer skills, compared to the AP condition. Overall, the study analyzes the benefits and shortcomings of the FS vs. AP review assignment protocol, providing evidence that the FS condition can be multiply beneficial to students who engage in peer review activities.

Keywords: Teaching/learning strategies; Pedagogical issues; Interactive learning environments; Multimedia/hypermedia systems.

1. Introduction

Peer review is an instructional method aiming to help students elaborate on domain-specific knowledge, while simultaneously developing methodological review skills. McConnell (2001) argues that peer reviewing offers to students the opportunity for a constructive and collaborative learning experience, by engaging them in an active learning exercise. Typically, peer review is a teacher-led activity, where the instructor assigns to each student (or group of students) the review of a piece of work (a written or verbal deliverable produced by another peer/group), according to specific quality criteria. The review then becomes available to the student authors and is used as a means for reflection and revision of the deliverable.

We use the term "assigned-pair protocol" here to refer to the class of peer review methods that involve static author-reviewer dyads. Students in a dyad can play both roles and review each other's work. The overhead for the instructor is well contained and the activity is straightforward for the students. While there are studies

---

[*] Corresponding author: P.M. Papadopoulos, pmpapad@csd.auth.gr, (tel.) (+30) 6938789030, fax (+30) 2310998419, Aristotle University of Thessaloniki, PO Box 114, 54124 Thessaloniki, Greece.

that assign multiple reviewers to a single peer (e.g., Tseng & Tsai, 2007; Tsai & Liang, 2009), in the context of our study we use the term "assigned-pair" to refer to a more common research design where students in a dyad are assigned exclusively to each other. This design, of course, is not without drawbacks. One review for a single peer might not be enough for the author to get valuable comments and suggestions for improvement. Additionally, the benefit for the reviewer may be limited, since she gets just one more point of view (the author's). Finally, the method requires a level of stability in the dyads formation throughout the activity, while a bad pairing may have negative results.

Our focus was to (a) enhance the learning benefits of peer review for the students, without increasing the minimum amount of work that they had to do, and (b) keep instructor's overhead low. Towards this direction, we decided to investigate the efficiency of a "free-selection protocol", where there are no dyads and students are free to browse all peer work and select what to review.

In the following, we present the theoretical background of our approach (section 2) our research method details (section 3), the study results (section 4), outcomes discussion (section 5), and also implications for the instructors on how to benefit from applying the free selection technique (section 6).

2. Theoretical Background

2.1 Peer Review

Peer review is similar to - but should not be confused with - "peer assessment" which refers to the activity of assessing student/group performance in relation to a group task (Loddington et al., 2008). Peer review is different mainly because it includes a "peer revision" phase, that is, a phase where students revise their drafts based on their peer review suggestions (Cho & MacArthur, 2010).

Peer review is primarily expected to support higher-level learning skills such as synthesis, analysis, and evaluation (Anderson & Krathwohl, 2001) as the students have the opportunity to analyze and evaluate peer work. Scardamalia and Bereiter (1997) 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. Nevertheless, learning at lower level (basic knowledge and understanding) should not be excluded since students may rehearse and elaborate domain-specific knowledge schemata when engaged in peer critiquing, by integrating new information that they had not seen before (Turner & Pérez-Quiñones, 2009).

The literature abounds with relevant studies indicating that the method is popular among educators inspired mainly by the constructivist and socio-constructivist paradigms for learning (e.g., Topping, 1998; Falchikov, 2001; Liu & Tsai, 2005) who want to challenge their students to think critically, synthesize information, and communicate science in nontechnical language (Loddington et al., 2008). The method has been used extensively in various fields (Turner & Pérez-Quiñones, 2009; Falchikov & Goldfinch 2000; Liu & Hansen 2002; Dossin 2003; Carlson & Berry 2005; Anewalt 2005; Hundhausen et al. 2009; Turner, Pérez-Quiñones, Edwards, & Chase, 2010; Liou & Peng, 2009; Goldin & Ashley, 2011; Gehringer, Ehresman, Conger, & Wagle, 2007), having a long history in writing instruction and relevant courses at the college level (DiPardo & Freedman, 1988; Haswell, 2005; Cho & MacArthur, 2010).

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; Mangelsdorf, 1992; Lundstrom & Baker, 2009). Certain studies support the use and adoption of peer review of writing (e.g., Cho & Schunn, 2007; Cho, Schunn, & Charney, 2006) emphasizing that when students get peer feedback and revise their written work they improve their writing skills (Fitzgerald, 1987; Hayes et al., 1987; MacArthur et al., 1991; McCutchen et al., 1987; Sommers, 1980; Cho & MacArthur, 2010).

2.2 Peer Review-Based Learning in the Computer Science Domain

Peer review has been used as a learning process to improve the quality of computer programs for at least 30 years (Anderson & Shneiderman, 1977; Luxton-Reilly, 2009), however, Turner & Pérez-Quiñones (2009) argue that in the Computer Science (CS) curriculum peer review is not widely used. Nevertheless, available research so far, in the CS discipline, has documented promising results (e.g. Crespo et al., 2004; Ziu et al., 2001; Turner, Pérez-Quiñones, Edwards, & Chase, 2010; Demetriadis et al., 2011). Whittington (2004) suggests that peer reviewing makes use of two important collaborative processes within the CS domain. First, students become familiar with peer reviewing processes in the context of software development for the diagnosis of programming errors and assurance of software quality (see also O'Neill, 2001). Second, their educators are familiar with peer reviewing as part of the publishing process. Typically, in a peer review cycle, an author drafts a piece of work which is then evaluated by a peer. The evaluation or critique is carried out anonymously on the basis of explicitly defined criteria and is subsequently returned to the author. The author is free to review his or her final draft based on the given critique. Yet, when practicing peer review in the classroom the instructor has a number of alternative design selections to choose from (for a detailed analysis of the peer review design space see Topping, 1998, and Turner & Pérez-Quiñones, 2009).

Certain researchers emphasize that by implementing peer review students get feedback of greater quantity than a busy teacher could reasonably provide (Wolfe, 2004; Silva & Moreira, 2003). This gives valuable feedback and the opportunity for development of critical reviewing skills. Others report benefits such as students' improved self-evaluation skills (Davies & Berrow, 1998) and improved students' attitudes and self-efficacy (Anewalt, 2005). However, Turner and Pérez-Quiñones (2009) emphasize: "While we have identified a number of potential benefits from reviewing, we have not shown that it is better than or as good as what we currently do. We require some sort of baseline to compare our efforts to. We need a control group in our experiments in order to judge effectiveness." (p. 44).

2.3 Peer review: Key research issues

Peer review takes many forms: it may be face-to-face or written, may involve numerical ratings as well as comments, and may supplement or replace instructor evaluations. The many forms of peer review may be confusing for the educator interested to implement the method. However, to increase the potential impact of peer assessment on learning, it is crucial to understand which mechanisms affect learning, and how these mechanisms can be supported (Gielen et al., 2010).

The Table 1 below emphasizes key aspects of the various peer review phases, what key questions are being investigated and what current research outcomes are available so far about these key issues.

Table 1. Peer review phases.

PHASE 1: "Producing Initial Student Work"

Description	Each student / group is assigned the development of a specific written/oral work
Expected Benefits	Students elaborate on domain knowledge
Key Research Questions	(no specific research questions – student work may be of various forms depending on the domain, and the learning objectives of the activity)
Research Evidence	--

PHASE 2: "Assigning Reviewers"

Description	Student work is assigned to reviewers
Expected Benefits	The review assignment protocol should maximize cognitive and metacognitive benefits expected from subsequent peer review phases
Key Research Questions	(1) Is there a preferred review assignment protocol? (i.e. assign reviews randomly in pairs, freely, matched)? If yes, on what grounds?
	(2) Explore the benefits emerging from the number of peer assessors by comparing a single assessor versus multiple peer assessors (Cho & Schunn, 2007)
	(3) Explore the benefits when matching principles for peers are applied (Van den Berg et al., 2006; Gielen et al., 2010)
Research Evidence	(1) Some studies suggest matching peers (author – reviewer) depending on the level of their skills (Crespo et al., 2004)
	(2) Some systems, such as PeerWise and that of Wolfe, do not limit the number of reviews that a student can perform. In such systems, 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)

PHASE 3: "Review/feedback production"

Description	Student reviewers are guided to provide reviews/feedback
Expected Benefits	(1) Student reviewers are provided with review guidelines, therefore they elaborate on domain-general knowledge/skills of peer review method
	(2) Student reviewers are guided to elaborate on the domain-specific knowledge
Key Research Questions	(1) Explore the role of reviewers' preparation including: (a) the training of peer assessors in assessment skills (Sluijsmans et al., 2002; Gielen et al, 2010), (b) the methods of teaching students how to provide peer feedback (Van Steendam et al., 2010; Gielen et al., 2010)
	(2) Explore the role of the feedback quality criteria (Van den Berg et al., 2006; Gielen et al., 2010)
Research Evidence	(1) The literature suggests that in order for students to successfully carry out an assessment of their peers they need to be prepared for the assessment (Loddington et al., 2008)
	(2) The quality of peer feedback can affect its impact (Gielen et al., 2010)

(3) The "givers", who focused solely on reviewing peers' writing, made more significant gains in their own writing over the course of the semester than did the "receivers", who focused solely on how to use peer feedback (Lundstrom & Baker, 2009; Li, Liu & Steckelberg, 2010; Reily, Finnerty & Terveen, 2009)

PHASE 4: "Revisions"

Description	Author students/groups are asked to revise their work based on peer reviews/feedback
Expected Benefits	Cognitive: Student authors elaborate on the domain by engaging in revision activity Metacognitive: Student authors reflect on the quality of their initial work and their peer reviews/feedback
Key Research Questions	Are revisions improved – and how – based on the peer provided review/feedback?
Research Evidence	Students receiving feedback from multiple peers improve their writing quality more than students receiving feedback from a single expert (Cho & Schunn, 2007)

Against the above background, this study focuses on Phase 2 (assigning reviewers) and explores the potential of the Free Selection assignment protocol to improve learning outcomes (that is, allowing students to browse and select for themselves peer work for review). There is already in the literature indication that the "randomly assigned pair" protocol, implemented typically by the instructors, might not be the optimal selection regarding student learning. For example, it is suggested that matching author-reviewer student pairs (depending on student author and reviewer ability) can lead to improved learning outcomes (Crespo et al., 2004). However, "matching student pairs" results also in additional instructors' workload, since it is necessary for the teacher to somehow model students' author-reviewer ability and apply an optimization algorithm for student matching. Such overhead would make the matching protocol hardly an appealing technique to employ, unless supported by appropriate technology tool. Additionally, although certain benefits have been emphasized related to enabling students provide multiple peer reviews (Luxton-Reilly, 2009), there is not any field research evidence available so far regarding the impact of a free selection technique on student learning.

2.4 Research Motivation and Hypotheses

Considering the above, we argue that applying the free selection protocol may result in minimized instructors' workload and improved student learning outcomes when engaged in peer review. Our objective is to improve the benefits of the method without however increasing the instructors' overhead. Following this perspective, this study applies an experimental research protocol to provide field research evidence on the possible benefits emerging from implementing the "free selection" protocol as compared to the "assigned pair" one. We tested three null hypotheses:

- H_{01} (review): "Students in both Assigned-Pair and Free-Selection groups perform the same as reviewers in a double-blind peer review activity".

- H_{02} (revision): "Students in both AP and FS groups perform the same in revising their answers, after they receive comments in a double-blind peer review activity".
- H_{03} (conceptual): "Students in both AP and FS groups perform the same in a test on acquisition of ill-structured domain conceptual knowledge".

3. Method

3.1 Participants

The study employed 54 sophomore students (32 males and 22 females) majoring in Informatics and Telecommunications Engineering in a 5-year study program. Students volunteered to participate in the study and we awarded a bonus grade for the laboratory course to students who successfully completed all the phases of the study. We randomly assigned students into three groups:
- Assigned-Pair (AP): 20 students, 12 males and 8 females randomly assigned into 10 same-gender dyads.
- Free-Selection (FS): 17 students, 9 males and 8 females.
- No review (NR): 17 students, 11 males and 6 females.

Students were domain novices and they had never before been typically engaged in case-based learning as undergraduates.

3.2 Domain of Instruction

Although the peer review method has been used to both well- and ill-structured domains, in this study we focus mainly on the latter. According to Spiro et al. (1992), in an ill-structured domain (a) knowledge application entails the simultaneous interactive involvement of multiple schemas, perspectives, organizational principles, each of which is individually complex, (b) the pattern of conceptual incidence and interaction varies substantially across cases nominally of the same type (i.e., the domain involves across-case irregularity). As such, in ill-structured domains alternative solutions and solving paths are not only acceptable, but expected, while the existence of an ideal solution is not always certain. We believe that this allows for more space for discussion and argumentation among students during the review process and provides an ideal test bed for the review protocol we propose.

The domain of instruction was "Network Planning and Design" (NP&D), which is a typical ill-structured domain characterized by complexity and irregularity. The outcome of a NP&D technical process results from analyzing user requirements and demands compromise in balancing technology against financial limitations (Norris & Pretty, 2000). The network designer has to solve an ill-defined problem set by the client. The role of the designer is to analyze the requirements, which are usually not fully specified, and follow a loosely described procedure to develop a practical solution. Hence, the success of a NP&D project depends greatly on past experience. Computer network design involves topological design and network synthesis, which are best conceived through studying realistic situations. Students in Computer Engineering learn to face realistic complex problems and they can be greatly benefited by project-based learning methods (Martinez-Mones et al., 2005). This is the reason why several researchers employ cases and plausible scenarios in their studies concerning network-related instruction. For example, Linge and Parsons (2006) employed realistic cases in a "Data Telecommunications and Networks"

module and they concluded that this is an ideal method for teaching computer network design. Gendron and Jarmoszko (2003) successfully utilized relevant real world problems to teach "Data Communications and Networking", while Noor (2003) applied the same learning principles to teach "Network Design and Management".

The suitability of the NP&D instruction domain for case study based teaching is supported by the fact that a specific problem may lead to multiple acceptable solutions according to the given requirements. Considering, for example, the case briefly described in Appendix A, the options of 24-port and 32-port switches both could appear as acceptable approaches when designing the departments' networks, on the grounds of tradeoff between network scalability and cost. Similarly, Cat5e and Cat6 cables could be accepted for usage in future gigabit Ethernet connections, depending on the subjective prioritization of the network performance, scalability, and cost.

3.3 The Learning Environment

For the purpose of our research, we developed a web-based environment that supported students in studying in the domain and performing the review procedure according to the respective protocol.

The students had to read supporting material, presented to them as past experiences in the NP&D domain, and provide answers to open-ended questions of related plausible scenarios. The scenarios presented to the students referred to various installations of computer network systems in new or restructured facilities, while the supporting material referred to similar projects highlighting important domain factors, such as the cost of the project, efficiency requirements, expansion requirements, and the traffic type and end-users' profile (see Appendix A for an excerpt).

Regarding the review procedure, the environment was generic enough, not to interfere with the main characteristics of the review protocols. The system was responsible for collecting all deliverables, granting access to students to peer work, and monitoring student activity throughout the phases. Especially for the Free-Selection setting, the role of a technological tool was important. As we will describe in detail in next sections, the basic concept behind the FS protocol is rather simple; however, the managerial task of distributing each deliverable to every student in a class and keeping track of who is reviewing what could pose a considerable overhead to an instructor implementing a paper-based FS protocol. We should underline here that the learning environment itself is not part of the analysis. An instructor could implement the two review protocols employing less sophisticated tools, such as forums and spreadsheets. We decided to develop our own system, so that we can better tailor the data collection during the activity according to the study needs.

3.4 Design

We used a pre-test post-test experimental research design to compare the performance of the different groups. The type of peer review performed by the students was the independent variable and students' performance in the written tests and in the learning environment were the dependent variables. All students worked individually throughout the activity, since the students in the FS and the AP groups were engaged in a double-blinded peer review process. The study had five distinct phases: Pre-test, Study, Review & Revise, Post-test, and Interview.

3.5 Pre- and Post-Testing

The pre-test was a prior domain knowledge instrument that included a set of 6 open-ended question items relevant to domain conceptual knowledge (e.g., "How can the security requirements of a network affect its architecture?"). The post-test also focused on acquired domain-specific conceptual knowledge including three domain conceptual knowledge 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.

3.6 Procedure

In the Pre-test phase, students completed the prior domain knowledge instrument in class. During the Study phase, all students logged in the environment (from wherever and whenever they wanted) and worked on 3 scenarios. Students had to read the resource material and based on that to provide answers to the open-ended scenario questions. They were allowed one week to complete the activity and study conditions were common for all the students.

Next, in the Review & Revise phase the students had to review, in a double-blinded process, the answers their peers gave to the scenarios in the previous phase (Study). Furthermore, the students were able, in case they wanted to, to revise their own answers according to the comments received from their peers. The Review & Revise phase also lasted one week. More specifically, we allowed a 4-day period for all the reviews, while the parallel revision of the previous answers lasted an additional 3-day period. Students in the No Review group skipped this phase and continued directly to the Post-test phase. To compensate the effort differences between the groups, we assigned an additional placebo task to the NR group after the Post-test phase. In the placebo task, NR students had to select and design one of the networks they proposed in the scenarios in a design software tool.

After the Review & Revise phase, the students took a written post-test in class and shortly after that, we interviewed the students from each group to record their approaches and comments on the activity. Interviews were semi-structured and focused on students' views on the activity and more particularly on the Review & Revise phase. Naturally, we skipped the topic of reviewing for the NR students. (Please see Figure 3 at the end of this section for a detailed representation of phase sequencing).

3.7 Treatment

The study conditions were the same for all students during the first week (study phase). In general, the students had to study the resource material and propose suitable solutions to the problems depicted in the respected scenarios.

In the second week, the NR group worked out of the environment, while the AP and FS groups continued with the review and revise phase, which was different for these two groups. Students in the AP group were randomly paired and had to blindly review each other's answers in the scenario questions of the previous phase. Hence, each student in the AP group had to submit 3 reviews (one for each answer to

the respective 3 scenarios). The steps that a student in the AP group had to follow while studying a scenario could be summarized as follow:

1. Submit a review to your peer's answer
2. Wait until your peer submits a review to your own answer
3. Submit a revised answer to the scenario, along with detailed justification of any changes made.

The difference between the Assigned-Pair and the Free-Selection groups was that students in the latter were able to see all the answers in their group and decide which to review. The answers were presented in random order in the "answer grid" (Figure 1). At first, only the first 200 characters of each answer was shown followed by a "read more" link. Each time a student clicked on that link, the system was recording the study time and the position of the answer in the answer grid.
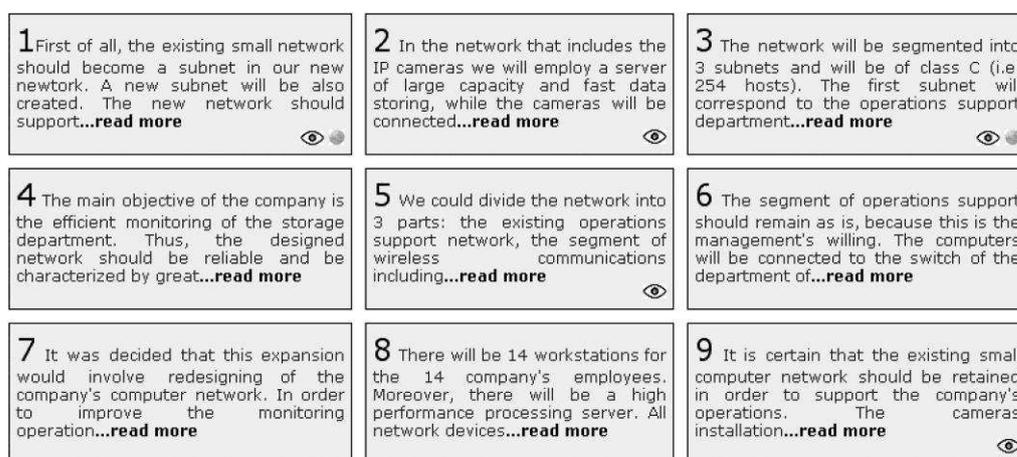

Figure 1. Part of the answer grid. According to this figure, the student has read answers 1, 2, 3, 5, and 9 (marked by an eye icon), and has reviewed answers 1 and 3 (marked by a bullet icon). The complete grid had 16 answers for the Free-Selection group.

The students were able to read as many answers they liked and they had to perform at least one review for each of the 3 scenarios (at least 3 reviews in total). At this point, we encouraged students to submit more reviews per scenario to increase the probability of all students receiving at least one review per answer. This was the only motive we gave the students, since no additional credit was given for such an act. Of course, it was possible for some answers to receive more that one review, while for others to receive none. For this reason, we were prepared to directly assign additional reviews to students during the 3-day "just revision" period of the Review and Revise phase. However, we acknowledge that there are various other strategies that span from "students should have complete freedom in the selection process" to "every student should get at least one review for each submitted answer".

Students in the AP and the FS groups had to follow a review microscript guiding them through the process, focusing on (a) content, (b) argumentation, and (c) expression (Figure 2). 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).

Figure 2. Review guidelines and form.

3.8 Data Analysis

Two subject-matter experts (SMEs), who had served as reviewers of the learning material, also served as raters. The SMEs are instructors of the "Network Planning and Design" course and have years of experience in network engineering research. To avoid any biases, students' paper sheets (pre- and post-test) and system print-outs (scenario answers and reviews) were mixed and assessed blindly by the two raters. The raters followed predefined instructions on how to assess each specific item. Eventually, each student received five scores from each rater, one from her peers, and one was constructed. Table 2 presents the meaning of each score, the scale used, and the way the total final score was calculated for each metric. The scores are presented in chronological order.

Table 2. Dependent Variables

Score name	Explanation, scale, and final calculation
Pre-test score	The mean score of the 6 conceptual knowledge questions of the pre-test instrument. Scale: 1-10. Total: mean of 2 raters.
Scenario.SME-score	The mean score of the initial 3 answers in the respective scenarios of the learning environment. Scale: 1-5. Total: mean of 2 raters.

Scenario.Peer-score	The mean score the student received from peers for the initial 3 answers of the learning environment. Students in the AP group received exactly 3 scores, while students in the FS group received at least 3. Scale: 1-5. Total: mean of all scores from peers.
Distance score	The mean (absolute) difference between the review scores submitted by the student and the respective scores submitted by the raters. Scale: 0-4. Total: mean difference between a student-rater and the SME-raters.
Reviewing score	The mean score for every review the student submitted in the Review & Revision phase. Exactly 3 reviews for AP students; at least 3 reviews for FS students. Scale: 1-5. Total: mean of 2 raters.
Scenario.Revised score	The mean score of the revised 3 answers in the respective scenarios of the learning environment. Scale: 1-5. Total: mean of 2 raters.
Post-test score	The mean score for the 3 conceptual knowledge questions of the post-test instrument. Scale: 1-10. Total: mean of 2 raters.

A 1-10 scale was used for the Pre- and the Post-test scores. On the contrary, a 1-5 scale was used for the Scenario.SME, Scenario.Peer, Reviewing, and Scenario.Revised scores, to be in line with the scale used by the students in their review process. The deviation between the rater scores was not to exceed the 20% level (two grades on the 1-10; one grade on the 1-5 assessment scale), else raters had to discuss the issue and reach a consensus. As a measure of inter-rater reliability, we calculated the two-way random average measures (absolute agreement) intraclass correlation coefficient (ICC) for the raters' scores. Figure 3 presents the phase sequencing, along with the characteristics and the data we collected in each phase.

For all statistical analyses a level of significance at .05 was chosen. To validate the use of the parametric tests, we investigated the respective test assumptions and results showed that none of the assumptions were violated.

Interviews were conducted to better understand how students of different groups worked and perceived the activity during the Study phase. Interviews were semi-structured and focused on students' views on the activity and more particularly on the peer review process.
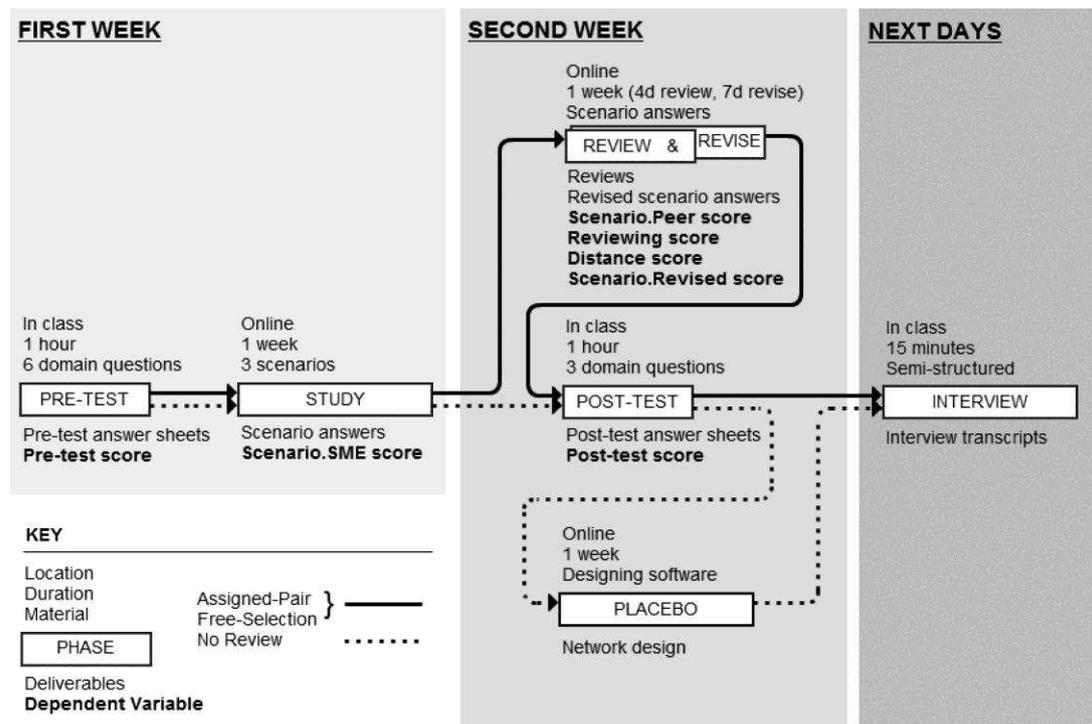
Figure 3. Time schedule, phase sequence, and metrics for the three groups.

4. Results

Inter-rater reliability was high for the Pre-test (ICC = .901), the Scenario.SME (ICC = .856), the Reviewing (ICC = .860), the Scenario.Revised (ICC = .877), and the Post-test (ICC = .828) scores. Table 3 presents the results regarding students' performance throughout the activity.

Table 3. Student Performance in the Activity

	Assigned-Pair			Free-Selection			No Review			Total		
	M	SD	n	M	SD	n	M	SD	n	M	SD	n
(scale: 1-10)												
Pre-test	2.69	(1.07)	20	2.59	(0.83)	17	2.73	(0.88)	17	2.67	(0.93)	54
(scale: 1-5)												
Scenario.SME	2.81	(1.00)	20	2.88	(0.86)	17	2.96	(0.72)	17	2.87	(0.86)	54
Scenario.Peer	3.98	(1.13)	20	3.56	(1.28)	17	--	--	--	3.79	(1.19)	37
Reviewing	3.09	(0.83)	20	3.64	(0.63)	17	--	--	--	3.34	(0.84)	37
Distance (0-4)	1.25	(0.68)	20	0.77	(0.37)	17	--	--	--	1.03	(0.60)	37
Scenario.Revised	3.29	(0.75)	20	3.45	(0.62)	17	--	--	--	3.36	(0.72)	37
(scale: 1-10)												

| Post-test | 7.71 (0.95) 20 | 8.43 (0.81) 17 | 6.85 (1.23) 17 | 7.66 (1.04) 54 |

In order to analyze the data and be able to address the research questions, we performed a number of different statistical tests. We present the results of these tests in Table 4 and we elaborate on our findings in the sections that follow.

Table 4. Statistical Tests Performed

Test Description	Results
T1. Difference in the Pre-test scores (AP vs FS vs NR)	One-way ANOVA $F_{(51,2)} = 0.013$, $p = 0.902$
T2. Difference in the Scenario.SME scores (AP vs FS vs NR)	One-way ANOVA $F_{(51,2)} = 0.081$, $p = 0.922$
T3. Difference in the Distance scores (AP vs FS)	t-test $t[35] = 2.567$, $p = 0.015$
T4. Difference in the Reviewing scores (AP vs FS)	t-test $t[35] = 1.887$, $p = 0.072$
T5. Correlation between Reviewing and Distance (AP $\cup$ FS)	Pearson's r $r = -0.396$, $p = 0.000$
T6. Difference in the Scenario.Revised scores (AP vs FS)	t-test $t[35] = 0.554$, $p = 0.581$
T7. Distance between Scenario.SME and Scenario.Revised scores (AP vs FS)	Paired-samples t-test AP: $t[19] = 1.682$, $p = 0.116$ FS: $t[16] = 5.162$, $p = 0.000$
T8. Difference in the Post-test scores (AP vs FS vs NR)	One-way ANCOVA $F_{(50,2)} = 9.017$, $p = 0.000$ NR-AP: $p = 0.013$ NR-FS: $p = 0.000$ AP-FS: $p = 0.048$

4.1 Pre-test Phase

One-way analysis of variance (ANOVA) results showed that the three groups were comparable regarding their prior knowledge, scoring very low in the pre-test instrument (Table 4: T1).

4.2 Study Phase

Students' performance was average in answering the three scenarios of the learning environment. One-way ANOVA results showed that there was no significant difference among the groups (Table 4: T2). This was expected, since the study conditions were the same for all students.

4.3 Review & Revise Phase

As we mentioned above, only the Assigned-Pair and the Free-Selection groups continued to the Review & Revise phase, while the No Review group skipped this phase and continued immediately with the post-test.

4.3.1 Usage Data Analysis

The big challenge in usage analysis was to examine the students' approaches, especially in the FS group. Considering the length of the answers provided by the students, we decided to set a threshold at 30 seconds for all the visits that we would accept as actual answer readings in our analysis. This amount of time should be enough for a brief reading, while shorter time periods usually suggest that the student is just browsing through the answers. Regarding the answers' position, usage data analysis showed that students were browsing the whole grid before selecting which answer(s) to review. This means that the answers that appeared first in the grid were not favored over the others.

Based on the above, usage data analysis showed that FS students read in average more than 8 answers out of the total 16 in each scenario ($M = 8.25$, $SD = 2.98$, min = 3, max = 12). Furthermore, FS students reviewed in average almost 2 answers per scenario ($M = 1.90$, $SD = 0.92$, min = 1, max = 4), and of course they received the same number of reviews for each of their answers ($M = 1.90$, $SD = 0.64$, min = 1, max = 3). We were expecting to have several answers without reviews by the end of the 4-day period of the Review & Revise phase. However, this happened only twice and we asked two students with good review record (number of answer visits and submitted reviews above average) to provide the missing reviews.

4.3.2 Review Process Analysis

The Scenario.Peer scores were much higher than the Scenario.SME scores, although the FS group seemed to be closer to the raters' opinion. We decided to examine this observation, by calculating the grading agreement between students and raters. As we mentioned earlier, the Distance value for each student was the mean of the absolute differences between each of the review scores submitted by the student and the respective scores submitted by the raters. For example:

> If a student reviewed three answers suggesting the scores (2, 5, 4), and the respective scores suggested by the raters for the same answers were (3, 4, 4), then the Distance value for this student would be:

$$(\, | \, 2 - 3 \, | + | \, 5 - 4 \, | + | \, 4 - 4 \, | \,) \, / \, 3 = (\, 1 + 1 + 0 \,) \, / \, 3 = 0.67$$

Usually, a correlation coefficient is used to analyze agreement. However, in this case, the number of reviews submitted by each student is too small and varies. Therefore, calculating Pearson's correlation coefficient for each student would be incorrect or even impossible (for $SD = 0$, i.e., where the submitted grades are equal, e.g., (4, 4, 4)). This is why we propose the Distance metric as an estimation of students-raters agreement.

Results showed that indeed, the difference between Distance scores was significant (Table 4: T3). As a next step, we asked the raters to assess the quality of students' reviews in terms of helpfulness, importance, and precision and to assign a score for each review using the same 1-5 scale. Results showed that FS students submitted better reviews and t-test results confirmed that there is a trend in favor of the FS group (Table 4: T4). This finding is related to the previous one, as Pearson's correlation test results showed that the Reviewing score is negatively correlated to the Distance score (Table 4: T5).

4.3.3 Revised Answers Analysis

Finally, the raters assessed the revised answers to the scenario questions and t-test results showed that there was no significant difference between the Scenario.Revised scores of the two groups (Table 4: T6). However, when we analyzed students' improvement between initial answers (Scenario.SME score) and revised answers (Scenario.Revised score), paired-samples t-test results showed that only the difference in the Free-Selection group was significant, while there was only a trend in the difference in the Assigned-Pair group (Table 4: T7).

4.4 Post-test Phase

To investigate the group differences in the post-test, we performed one-way analysis of covariate (ANCOVA), using the Pre-test score as a covariate. Results showed that there were significant differences (Table 4: T8). Specifically, pairwise comparisons showed that there was a significant difference between NR and the two reviewing groups (AP: $p = 0.013$; FS: $p = 0.000$) and between AP and FS groups ($p = 0.048$), with the NR students scoring lower than the others, while the FS groups scored significantly higher than the others.

4.5 Interviews

Interviews lasted about 15 minutes per student and were audio recorded. We used the interview transcripts for content analysis.

All students felt comfortable with the environment and the material underlining the connection of the cases used with real-world problems. Students of both reviewing groups appreciated the comments they received from their peers and they mentioned examples where a review comment made them re-evaluate and revise their answers. Also, both groups said that the Review & Revise phase was helpful in understanding deeper the material, taking into account different perspectives, and providing improved and more comprehensive answers. Raters' opinion depicted in the difference between the Scenario.SME and the Scenario.Revised scores are in line with students' beliefs about better final answers. However, the two groups seemed to have some differences in the level of appreciation for the review microscript. AP students were divided whether the microscript was helpful or not, while all the FS expressed a very positive opinion stating that the microscript made the process clearer and simpler for them.

Furthermore, we asked an additional line of questions to students in the FS group, to analyze the way they worked during the activity. One important question was concerning students' criteria for choosing an answer over the others for reviewing. From students' responses, we identified two opposite ways of thinking presented graphically in the following statements (S1, S2):

> S1: "I was trying to find an answer that I thought it was good and complete, so that I would be able to say nice things and give a good score."

> S2: "I was trying to find an answer with a lot of flaws and mistakes, so that it would be easier for me to make some useful comments, other than just saying 'good job'."

We also asked FS students who submitted more than one review per scenario to explain their motives for such a strategy, especially since they were not awarded extra credit. First, some students said that writing down reviews and explaining their opinions to others was a good exercise to clarify their own understandings. Second, students also mentioned that after reading several answers, and since the answers were relatively short, it was easy for them to spend a little time submitting more reviews. In that way, they thought that they would increase the possibility of everyone receiving at least one review.

However, we believe that other factors such as the length and coherence of an answer, the tiredness of the student, and the time schedule also played a role in the selection of the answers-to-be-reviewed. As the same strategy could be the result of various factors, we refrained from categorizing students and comparing different profiles in the FS group. For example, it is not easy to identify whether a student that read only 2-3 answers before choosing one for review does this because she actually found what she was looking for or because she was just not deeply engaged to the activity. Furthermore, students' strategies could change in the three scenarios, making the categorization into different profiles even more difficult.

5. Discussion

5.1 Hypotheses Testing

In the light of the results above, hypothesis H_{01}, concerning students' review skills, is rejected and an alternative is proposed. Hypothesis H_{02}, referring to the revised answers is tentative, since we have non-concluding results. Finally, hypothesis H_{03}, regarding students' performance in the post-test, is also rejected (Table 5).

The students scored very low in the pre-test, having minimal differences between the groups and confirming what we expected to see from novices. Similarly, the three groups had the same performance during the first week of the activity, where they had the same study conditions. As we mentioned earlier, the No-Review group logs-out of the environment in the second week and continues directly with the post-test. Since the three groups had the same performance in the study phase, we can hypothesize that the scores of the NR group in the post-test also represent the scores that the other groups would have had, if they had followed the same phase sequence. Following this rationale, we can attribute the occurred differences between the NR group and the two other groups on the Review and Revise phase.

Table 5. Hypotheses Testing

Null hypothesis	Result	Alternative
H_{01} (review): "Students in both Assigned-Pair and FS groups perform the same as reviewers in a double-blind peer review activity".	Rejected on the basis of: Table 4: T3, T4, T5.	H_{A1} (review): "Students who study in FS condition perform better as reviewers in a double-blind peer review activity".
H_{02} (revision): "Students in both AP and FS groups perform the same in revising their answers, after they received comments in a double-blind peer review activity".	Tentative. Table 4: T7 shows significant improvement for the FS group, but Table 4: T6 shows that the revised answers were close.	--
H_{03} (conceptual): "Students in both AP	Rejected on the basis of:	H_{A3} (conceptual): "Students who

and FS groups perform the same in a test on acquisition of ill-structured domain conceptual knowledge".	Table 4: T8.	study in FS condition perform better in a test on acquisition of ill-structured domain conceptual knowledge".

5.2 Review process: do students become equally good reviewers?

When we say that a student is a good reviewer, we mean that she is able to assess an answer in the same way that a domain expert would have done and that she is also capable of providing valuable comments for improvement. We have two findings that help us provide an answer to this question. First, results analysis showed that students' grades to their peers were higher than the grades the raters assigned. Although this was evident in both groups, the difference was significant only in the AP group (Table 4: T3). This means that the students in the FS group were closer to the raters' opinion about the quality of the answers. Second, when we asked the raters to assess the quality of students' reviews in terms of helpfulness, importance, and precision, we saw that there was a strong trend in favor of the FS group. As we saw earlier, the two findings are strongly correlated. This is normal, since the grade that a student suggests in a review should represent the provided comments. Hence, we can argue that the students in the FS group became eventually better reviewers than the students in the AP group.

Having in mind that the review microscript was the same for the two groups, we can say that the enhanced performance of the FS students can be attributed to the different review conditions. Students that worked in pairs were able to read only the answers that one other student provided. Hence, they had a limited chance of getting to know different perspectives and opinions about the issues raised in the scenario questions. Consequently, they had to do the reviews based mainly on the understandings they had developed during the study phase. Of course, by reading someone else's answers, a student may shift from her original answer and adopt some new ideas. However, an answer should be substantially better than the reviewer's, in both content and argumentation, to function as an eye-opener for the reviewer and make her change radically her initial opinion.

On the contrary, this was not the case for the FS group. We gave the Free-Selection students the ability to read all the answers in the group and we instructed them to review at least one in each scenario. A student could opt for a minimum effort strategy, choosing only one random answer to read and review. In this case, the student would have had the same conditions we applied in the AP group. However, usage data analysis showed us that students in the FS group read more than half the answers in each scenario and they decided to review almost double the answers that the AP group did. We need to underline here the importance of this finding. Students followed a learning strategy with increased effort without the obligation to do so. Indeed, some students read up to 12 out of the 16 available answers in a scenario, while the smallest value observed for the FS group was 3 answers per scenario (still much more than the one answer per scenario that the AP group had). According to students' statements in the interviews, the motives for this way of work were their curiosity for others' opinions and a genuine appreciation of the positive effect the acquisition of different perspectives has on their own understandings. Inevitably, by reading many different answers one should be able to understand where these answers are converging. This means, that it was easier for the students to compare answers and

grasp a clearer picture. Eventually, this way of work helped the FS students develop a better review criterion.

5.3 Revision process: are the revised answers equally good?

It was obvious that the revised answers of both groups (AP, FS) were better than the initial ones. However, the difference in the Revision scores of the two groups remained non-significant. Based on this fact, we cannot reject hypothesis H_{02}. After going deeper into data analysis though, we found out that the review process was more beneficial for the FS group. We base this on the significant difference recorded between the initial and the revised answer scores for the FS group, while at the same time there was only a weak trend for the respective difference of the AP group. That is why, hypothesis H_{02} is tentative.

Going a few steps ahead in our analysis, it is clear that the two reviewing groups (AP, FS) had significant difference in the post-test, showing that probably the students had acquired different levels of knowledge. So, the question that rises is why the same difference did not occur in the revised answers of the scenarios in the learning environment. It seems that for the AP group, even reading one different answer and receiving one review, was enough to have a considerable improvement on the initial answer. Although this improvement was not significant, it kept the difference between AP and FS in a non-significant level. To answer the question, we need to examine the differences in revising an answer and having a test on domain conceptual knowledge. Working in pairs gives the chance to students to improve what they have already wrote, either by getting some useful comments or adopting good ideas from the answers they review. On the contrary, in the post-test, students have to show that they have acquired abstract domain knowledge. To perform well in the post-test, students need to have a deeper understanding of the domain and be able to generalize and see the connection between a specific instance and a general domain principle. Students in the FS group were more exposed to multiple perspectives and probably this helped them gain a better view of the field. Hence, they were able to improve significantly their initial answers and answer in a better way than the AP students the post-test questions.

5.4 Post-test: do the students demonstrate the same learning outcomes at the end?

Both AP and FS groups have far better scores in the post-test than the NR group. This is of course expected, because – as we explained earlier – NR students' Post-test scores can be perceived as snapshot of all the participants at the end of the Study phase. We have also mentioned in the previous section that the FS group scored significantly higher than the AP group (Table 4: T8).

We argue that the main reason behind FS students' better performance is their deeper involvement in the activity, meaning that they actually read more answers and performed more reviews. In other words, they intensified the treatment by tailoring it according to their needs. By following this strategy, the FS students were eventually able to develop a deeper understanding of the domain. Taking into account all the recorded differences between the AP and the FS groups in reviewing, revising, and post-test, we argue that the FS setting was indeed more beneficial for the students.

5.5 Interview: do the students develop the same opinions towards the review process?

Regarding the two different approaches in the FS group in selecting an answer for review, we can see that even in the same treatment it is possible to have students moving to completely opposite directions. The students that opted for the good answers to review tried to be pleasant to their peers, while the ones that opted for the problematic answers tried to be more useful. Both of them used their own criteria to identify the good and bad answers, meaning that their opinions were not always in accordance with what the raters thought. The different approaches in selecting answers may also be the reason for the wide spread of reviews. The result of this spread was that only two of the answers were not reviewed in the initial 4-day review period we allowed to the students.

Maybe the most interesting aspect of the interviews was our effort to decipher FS students' strategies, especially since they willingly got more engaged in the activity. Our initial expectation was to have only a small number of students reading many answers and submitting additional reviews apart from the mandatory one. However, the results drew a different picture. The students acknowledged that the task of writing down reviews, and essentially explaining to others their opinions through comments and suggestions, was a good exercise for them to clear in their minds their own understandings. This was the reason mentioned in most cases and it is very important because it mirrors students' metacognitive skills. The second reason for deeper engagement was, according to students, the fact that the whole process of reading an answer and commenting on it did not take a lot of time. This was also important and gives us insights for the relation between the workload and the engagement. Students suggested that they would not have read so many answers, if they were too long or too difficult to read. In this way, they noted some of the limitations of the FS approach (e.g., settings where students' answers span along several pages). Finally, it seems that some of the FS students felt that they were members of a community and they tried to contribute more, hoping that others will follow suit. This is clear when they say that they submitted more reviews to increase the possibility every of their peers getting one.

Overall the current study presents field research evidence that the students who followed the free selection protocol have significant benefits (compared to students who follow the typical "assigned pair" protocol) regarding: domain learning, review ability and revision ability. They also report interesting perspectives on their engagement in the activity, such as that writing multiple reviews helped them "clear their own understanding" and "feel member of a community".

The above results are inline with previous studies that report beneficial learning outcomes (a) when students get multiple reviews (as opposed to getting a single review) (Cho & Schunn, 2007), and (b) when students act as "givers" (that is when they provide reviews) compared to acting as "receivers" (getting peer feedback) (Lundstrom & Baker, 2009; Li, Liu & Steckelberg, 2010; Reily, Finnerty & Terveen, 2009). Furthermore, the students' positive approach to the FS method is also in line with other studies reporting tendency of students to contribute more when not restricted by the assignment protocol (Luxton-Reilly, 2009). Summarizing, the current study strongly encourages teachers and peer review system designers to provide opportunities for students to engage in multiple review activity mainly from the perspective of review provider ("giver").

6. Implications for Design of a Free-Selection Technique

There are several characteristics of the Free-Selection approach that can be changed in various ways to accommodate different learning needs. For example, an instructor could opt for a non-blind FS technique, or make the reviews submitted available to others. Based on the findings of this study, we analyze in the following some of the most important aspects of the FS protocol we implemented.

Resistant to group size changes

Contrary to any setting with pre-assigned group formation, the FS approach can easily deal with odd or even number of students and unexpected changes in students' population. This is a clear advantage for instructors that very often need to reorganize groups and reassign reviews after sudden dropouts or last minute entries. In the FS approach, it does not really matter whether the answer grid will comprise 10, 11, or 14 answers. What is important is to allow students get multiple perspectives and maybe disable the review functionality to answers from students that are out of the activity, so that the reviews will be channeled to students that really need them.

Small groups to increase engagement and diminish randomness

Although it could be beneficial to students to read different opinions, we should have a threshold so that they will not be overwhelmed by a large number of answers. The higher the number of available answers, the higher the chance of students picking answers for review randomly. In case of a large class, the instructor can divide the students into smaller groups and apply the FS approach independently to each of them.

Preferred for short deliverables

It was very important in the study that the students' answers were relative short, rarely filling a whole page. This was commented by the students in the interviews, saying that it was easy to read a lot of answers because of their length. One should expect that as the answers become longer, the possibility of students reading more of them gets lower. This in term diminishes one of the main purposes of the FS approach, which is to make the students learn more by analyzing different opinions.

Implementing a at-least-one-review-per-submission policy

This is a difficult issue to tackle, as it affects the very nature of the protocol. Ideally, we would like to have each answer reviewed at least one time, so that every student will receive comments tailored exactly on what she submitted. On the other hand, we do want to let the students decide which answer to read and review. These two needs are contradicting and it is clear that there is no solution that can satisfy them both. There are many different ways an instructor could choose from. For example, guiding students to the non-reviewed answers, by applying a first-come first-served approach, where the reviewed answers are noted or even excluded for additional reviews until all the grid is reviewed. In this study, we decided to give more attention on providing complete freedom of selection to the students by following a simple approach. We allowed a 4-day period where the students could choose freely answers to review, and

we assigned the non-reviewed answers to students in a 3-day period that followed. The random positioning of answers in the grid has maybe helped in the wide spread of reviews, so at the end we had only two answers to assign. Of course, the outcome from the 4-day period could be much worse, with a lot more non-reviewed answers to assign. This should also be expected. However, looking back in all the data, we believe that the students were benefited more by reading the different answers than by getting reviews to their answers. After all, even the students that only got one review per answer read many more answers while reviewing. A review can be helpful or not, since it may or may not include good suggestions for improvement. On the other hand, when the students are exposed through reading to many different opinions, they tend to compare and search for the dominant opinion or for the one that fits theirs better. Through this process, the students get an indirect feedback. In other words, we believe that the FS students would benefit from the process, even without a policy that ascertains that each answer gets at least one review.

The role of technology

When peer review is practiced in the classroom, it is relatively easy for the teacher to implement simple assignment protocols like "assigned pair" without any support from technology tools. However, in e-learning settings or when more complex protocols are applied, there is a significant overhead caused by the administrative and management costs of the method, that specifically designed technology systems can alleviate (Ballantyne et al., 2002; Luxton, 2009). In this study, we developed an appropriate technology-enhanced learning environment to implement and analyze the two review protocols. The degree of complexity introduced by the Free-Selection protocol makes the use of technology necessary. However, what is important is not the specific environment we used, but the way technology supports the implementation of the phases that constitute the FS protocol. namely: (a) support students in producing the initial deliverables, (b) make these deliverables available for review to all students, (c) send reviews back to authors, (d) support students in revising their work and present the final deliverables.

      The first phase concerns the presentation of the learning material to the students, along with appropriate tools (e.g., instructions, forms, examples). At this point, technology can be used instead of printed material. Additional, hypertext can support domains where a linear method for presenting concepts is not enough (e.g., ill-structured domains). In the "Network Planning and Design" domain, each scenario was connected to a number of supporting resources (i.e., advice-cases evolving around domain themes that also appeared in the scenario). A domain theme can appear across many advice-cases and scenarios and it is important for students to be able to browser freely in the material, while studying.

      The role of technology becomes essential, though, in the next phases of the protocol. Making students' work available to other is not an easy task to do without the managerial support of a technological system. For example, an instructor applying a paper-based FS protocol in a class of 10 students has to make 9 copies of each student answer and give everyone a set containing all the answers of their peers. This is of course impractical, since many of the copies are not going to be read or reviewed. In the next phase, the instructor has to collect all the copies back, along with review comments, and give them to the authors so that they will be able to provide their revised answers. The technology can easily lift the weight of managing

the access to multiple deliverables and sending the appropriate reviews back to the authors.

However, the role of technology spans beyond document management. A big advantage for the instructor is that through technology the whole activity can be monitored and valuable information about students' strategies can be recorded. For example, in this study it would be impossible to know how many answers a student read before submitting a review, without an appropriate module in the learning environment we used. Similar data (e.g., number of logins, page visits, etc.) may be useful for the instructor in evaluating students' effort in the activity.

Technology can also be used to adjust the degree of coercion in each step of the learning activity. Our students had to successfully complete the first week of study by submitting answers to the three scenarios, to get access to the answer grid and be able to read what their peers submitted. Similarly, they were able to see the reviews they received only after completing the review phase (i.e., submitting at least one review per scenario).

In this study, we focused on a simple version of a FS protocol, applying a double-blinded review process and controlling the data available to the students. This means that, although our learning environment recorded in detail students' activity, this data was not available to them. However, an interested instructor could apply a more complex version where review comments are publicly available, Scenario.Peer scores are visible in each scenario answer and updated after each review, and authors and reviewers can communicate directly with each other. This type of implementation cannot be done without the use of appropriate technological support.

Future research

Based on the above we can recommend a series of future studies to analyze further the strengths and limitations of the Free-Selection method. First, students noted in several occasions that the length of the deliverables were relative short (the length of students' answers was usually less than a page). This clearly affected the strategies they applied, since reading and reviewing answers was an easy task, in terms of time. One could argue that the effectiveness of the FS protocol would drop in research designs where peer work is more complex and spans over several pages. Future studies can test this assumption. We would expect a drop in the number of answers read and reviewed, although we believe that even in these cases the implementation of the FS approach would be more beneficial than the assigned-pair method.

Second, the research question that came up in this study and is going to be our focal point in a following study is the impact of indirect feedback in the FS method. We already presented our belief that it was more beneficial for our students to perform many reviews (and thus to get multiple perspectives), than to receive multiple comments from their peers. Lundstrom and Baker (2009) have already reported that students playing only the role of reviewers gained significantly more in a writing class than students playing only the role of authors. In FS protocol, students play both roles and it would be interesting to analyze the impact of indirect feedback on students that submitted reviews, but did not receive comments for their work as authors.

7. Conclusions

In this study, we compared the potential of an alternative to the typical Assigned-Pair technique seen very often in peer review and collaborative learning settings. Our

scope was to deal with the shortcomings of the AP approach, without increasing the instructor's overhead, and in the process to support the students more efficiently. The Free-Selection technique seems to be a good step towards this direction, since it keeps instructor's involvement and workload to a bare minimum. Additionally, it gives control to the students, resulting to better strategies. The students are responsible for the volume of the material they are going to study, fulfilling of course some basic requirements (at least one review per scenario). Eventually, it seems that the Free-Selection technique is more beneficial to students, since they demonstrate enhanced performance concerning both domain specific (conceptual) and domain general (reviewing) knowledge.

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. NY: Longman

Anderson, N., & Shneiderman, B. (1977). Use of peer ratings in evaluating computer program quality. In Proceedings of the 15th Annual SIGCPR Conference, 218-226, Arlington, Virginia, USA. New York, NY: ACM.

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.

Carlson, P.A., & Berry, F.C. (2005). Calibrated Peer Review: A Tool for Assessing the Process as Well as the Product in Learning Outcomes. In Proceedings of 2005 American Society for Engineering Education Annual Conference & Exposition.

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.

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

Crespo, R. M., Pardo, A., & Kloos, C. D. (2004). An Adaptive Strategy for Peer Review. Paper presented at ASEE/IEEE Frontiers in Education Conference. Savannah, GA.

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.

Demetriadis, S., Egerter, T., Hanisch, F., & Fischer, F. (2011). Peer review-based scripted collaboration to support domain-specific and domain-general knowledge acquisition in computer science. Computer Science Education, 21(1), 29-56.

DiPardo, A., & Freedman, S. W. (1988). Peer response groups in the writing classroom: theoretic foundations and new directions. Review of Educational Research, 58, 119-149.

Dossin, M. M. (2003). Among Friends: Effective Peer Critiquing. The Clearing House, 76 (4), 206-207.

Falchikov, N. (2001). Learning together: Peer tutoring in higher education. London: Routledge Falmer.

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. Review of Educational Research, 70(3), 287-322.

Fitzgerald, J. (1987). Research on revision in writing. Review of Educational Research, 57, 481-506.

Gehringer, E., Ehresman, L., Conger, S. & Wagle, P. (2007). Reusable Learning Objects Through Peer Review: The Expertiza approach. Innovate, 3 (5). Retrieved September 10, 2009 from: http://www.innovateonline.info/index.php?view=article&id=365

Gendron, M., & Jarmoszko, A.T. (2003). Teaching data communications and networking to school of business undergraduates: A pedagogical framework for juxtaposing theory and practice. Proc. Informing Science IT Education (InSITE) Conf., 1145–1153.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. Learning and Instruction, 20, 304-315.

Goldin, I. M., & Ashley, K. D., (2011). Peering Inside Peer Review with Bayesian Models. In G. Biswas et al. (Eds.): AIED 2011, LNAI 6738, pp. 90–97. Springer-Verlag: Berlin.

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.

Hayes, J. R., Flower, L., Schriver, K. A., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Eds.), Reading, writing, and language processing. Advances in applied psycholinguistics, 2, 176-240. Cambridge, UK: Cambridge University Press.

Hundhausen, C., A. Agrawal, Fairbrother, D., & Trevisan M. (2009). Integrating pedagogical code reviews into a CS 1 course: an empirical study. Proceedings of the 40th ACM technical symposium on Computer science education. 291-295. Chattanooga, TN, USA, ACM.

Linge, N., & Parsons, D. (2006). Problem-based learning as an effective tool for teaching computer network design. Education, IEEE Transactions on, .49(1), 5- 10.

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

Liu, J., & Hansen, J.G. (2002). Peer Response in Second Language Writing Classrooms. In Belcher, D., & Liu, J. (Eds.), Michigan Series on Teaching Multilingual Writers. Ann Arbor, MI: U of Michigan P.

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.

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

Loddington, S. P., Wilkinson, N., Bates, M. R. N., Crawford, A R. and Willmot, P. (2008) Discover WebPA: an online self and peer assessment system, EE2008: International Conference on Innovation, Good Practice and Research in Engineering Education, 14-16 July 2008, Loughborough, UK. Available at http://www.engsc.ac.uk/downloads/scholarart/ee2008/w005-loddington.pdf [accessed 31 July 2009].

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.

MacArthur, C. A., Graham, S., & Schwartz, S. (1991). Knowledge of revision and revising behavior among students with learning disabilities. Learning Disability Quarterly, 14, 61-73.

Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? ELT Journal, 46, 274-284.

Martinez-Mones, A., Gomez-Sanchez, E., Dimitriadis, Y.A., Jorrin-Abellan, I.M., Rubia-Avi, B., & Vega-Gorgojo, G. (2005). Multiple Case Studies to Enhance Project-Based Learning in a Computer Architecture Course. Education, IEEE Transactions on, 48(3), 482- 489.

McConnell, J. (2001). Active and cooperative learning. Analysis of Algorithms: An Active Learning Approach. Jones & Bartlett Pub

McCutchen, D., Hull, G. A., & Smith, W. L. (1987). Editing strategies and error correction in basic writing. Written Communication, 4, 139-154.

Noor, R. (2003). Implementing a problem based learning for undergraduate course: A first experience. Presented at The Learning Conf. 2003: What Learning Means, Jul. 15-18, 2003, Institute of Education, Univ. of London.

Norris, M. & Pretty, S. (2000). Designing the Total Area Network. New York: Wiley.

O'Neill, D. (2001). Peer Reviews. Encyclopedia of Software Engineering. NY: John Wiley & Sons.

Reily, K., Finnerty, P.L., & Terveen, L. (2009). Two Peers are Better Than One: Aggregating Peer Reviews for Computing Assignments is Surprisingly Accurate. Proceedings of GROUP'09, May 10–13, 2009, Sanibel Island, Florida, USA.

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

Silva, E. & Moriera, D. (2003). WebCoM: a tool to use peer review to improve student interaction. Journal on Eduational Resources in Computing, 3(1), 1-14.

Sluijsmans, D.M.A., Brand-Gruwel, S., & Van Merriënboer, J.J.G. (2002). Peer assessment training in teacher education: effects on performance and perceptions. Assessment and Evaluation in Higher Education, 27, 443-454.

Sommers, N. I. (1980). Revision strategies of student writers and experienced writers. College Composition and Communication, 31, 378-387.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In T. Duffy & D. Jonassen (eds.), Constructivism and the technology of instruction (pp. 57–75). Hillsdale, NJ: Erlbaum.

Topping, K. (1998). Peer assessment between students in colleges and universities. Review of Educational Research, 68, 249-276.

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.

Turner, S. & Pérez-Quiñones, M.A. (2009). Exploring Peer Review in the Computer Science Classroom, Computing Research Repository. Retrieved November, 30, 2009 from http://arxiv.org/ftp/arxiv/papers/0907/0907.3456.pdf

Turner, S., Pérez-Quiñones, M.A., Edwards, S., & Chase, J (2010). Peer Review in CS2: Conceptual Learning. *Proceedings of SIGCSE'10*, March 10–13, 2010, Milwaukee, Wisconsin, USA

Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Peer assessment in university teaching: evaluating seven course designs. Assessment and Evaluation in Higher Education, 31, 19-36.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. Learning and Instruction, 20(4), 316-327.

Whittington, J. (2004). The process of effective critiques. Computers & Graphics, 28(3), 401-407, Elsevier.

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

Ziu, E.Z.F., Lin, S., Chiu, C.H., & Yuan, S.M. (2001). Web-based peer review: the learner as both adapter and reviewer. IEEE Transactions on Education, 44, 246-251.

Appendix A
Example of Studied Material

The following is an excerpt from a resource material in the Network Planning and Design domain. This and other material of similar complexity, adapted from the literature, was presented to students as past experience. This particular excerpt presents how the expansion requirements affected network design decisions.

Expansion requirements of the Metalux computer network

The expansion requirements affect to a great degree the network design decisions.

From the beginning of the design process, the management of Metalux clarified to the Direct Comm designers that the development of an easily expandable network is of high priority. The intention to increase workers and add a new production unit was a significant factor for the majority of the designing choices of Direct Comm.

For example, the expected inclusion of more employees influenced the decision on the number of workstations. So, the network planners considered 3 extra computers for the accounting department, 4 extra computers for the sales department, and 4 extra computers for the production department.

Moreover, Metalux preferred to acquire a 24-port switch for every of its departments aiming to the future interconnection of more devices. Notice that the initial number of devices that needed to get connected was much lower.

The type of the network devices (switches and routers) was chosen to be GbE, despite that the majority of the links operated at Fast Ethernet mode. The concept was to avoid link replacements in a future network upgrade.

Following the same concept, the designers decided that the performance of the standard Cat5 cables was not adequate for the Fast Ethernet connections. Thus, they included full Cat5e cabling, in order to support more efficiently future GbE connections.

Furthermore, the expansion requirements greatly affected decisions regarding subnetting and addressing. The network designers defined a redundant number of available host addresses (126 for each department). In fact, a set of addresses corresponding to a subnet was intentionally left unused, so that the second production unit network could make use of them in the future. Besides, the purpose of the inclusion of a router for connecting just one switch (the one belonging to the production department) based on the same concept: allowing the future connection of that router to the switch of the second production unit.