# Can peers rate reliably as experts in small CSCL groups?

Ioannis Magnisalis[a], Stavros Demetriadis[a], Pantelis M. Papadopoulos[b]

[a]Aristotle University of Thessaloniki, PO Box 114, 54124, Thessaloniki, Greece

[b]Aarhus University, 8200 Aarhus N, Denmark

Email: imagnisa@csd.auth.gr; sdemetri@csd.auth.gr; pmpapad@tdm.au.dk

**Abstract.** Research on the impact of peer rating (PR) has provided encouraging results, as a method to foster collaborative learning and improve its outcomes. The scope of this paper is to discuss peer rating towards two specific directions that usually are neglected in the CSCL field, namely: a) coaching of objective anonymous peer rating through a rubric, and b) provision of peer rating summary information during collaboration. The case study utilized an asynchronous CSCL tool with the two aforementioned capabilities. Initial results showed that peer rating, when anonymous, and guided, can be as reliable as off-line expert/teacher rating, with indications that this process can foster collaboration.

**Keywords:** peer rating, expert rating, computer-supported collaborative learning, asynchronous forum discussion

## 1 Introduction

Collaborative learning (CL) is important for students both for social and cognitive reasons [1]. Computer-supported collaborative learning (CSCL) is not simply implying the use of technology for communication purposes to enable CL, but aims to improve both CL peer skills and individual/group learning domain products. Efforts to implement reflection tools [2] in CSCL that foster peer interactivity (PI) and improve the collaborative learning process and outcome have been systematically reported in the literature, providing encouraging evidence on the impact of these methods to enhance student learning. In the current work, we study qualitative aspects of PI in a setting where students rate each other's posts in a Moodle forum. Peer rating (PR) is defined as the process through which students monitor and rate the performance of their fellow group members. PR reflection, assisted by visual feedback on PR, is defined as the cognitive and affective activities individuals engage in to explore their experiences and reach new understandings and appreciations of those experiences [3]. Models that capture both activity and domain aspects of PI are described in [4], [5]. In the current study, we use scheme of [4] to classify peer interactions, according to their qualitative characteristics.

Peer rating is a complex skill that does not effectively or efficiently emerge or develop in a spontaneous way [6]. To effectively and efficiently use PR, (a) simple PR tools should be used, and (b) PR process needs to be supported and guided [6]. Solely providing students with a tool to rate posts in a discussion forum, is probably not enough to alter group collaboration balance or change students' rating standards. Visual feedback can support reflection and plays an important role in individual learning processes [7], as well as in collaborative learning processes. Small rating deviations among peers may suggest that group is led to common understanding and awareness. Enhanced group awareness can lead to more effective and efficient collaboration ([8], [9], [10]). The data from [11] and [12] revealed how computer mediation can improve the reliability and validity of peer review activities, while simultaneously improving their functionality. Intra-class correlation coefficient (ICC) [13] was used to assess SWoRD's score reliability comparing teachers' and students' ratings.

Rating has been accompanied by visualization techniques in CSCL already. In [14], authors present augmented group awareness tools supporting collaborative learning. The group awareness tool provided to the small groups in one of the experimental conditions was embedded into the online discussion environment. Taking about visualizations, in [15] students, before contributing in the discussion, they gathered information about current balance over participation while in [16] authors tried to boost student motivation by building a positive sense of competition using a representation of average class performance.

Here, we present the study of a technology system that employs visualized peer rating data on posts from a Moodle forum. The main research question of this study can be stated as such: "*To what extent can peer rating be reliable, when compared with expert rating*?"

## 2 Method

### 2.1 Participants, learning environment and procedure

The study was conducted in a Second Chance school in Thessaloniki, Greece. The participants were 176 students (ages 18-50, M=42, SD=3.7), with most of them having low familiarization level with online communication tools; only 13 had used forum/chat tools before -but none of them for educational purposes- with an average computers and in-formation literacy level of 3.8 out of 10 (based on 35 questions, similar to [17], and designed according to the B-Tile [18]). The students were randomly distributed into 44 groups of 4 peers.

In this work, each group member provided anonymous feedback to peers within the same group. Each student had been attributed to a pseudo name in the system and rated posts of peers anonymously. All ratings were calculated, summarized and visualized as feedback to collaborating peers (see Fig. 1, PR at both group and individual levels are depicted). Thus, Moodle was enhanced with: (a) an anonymous peer rating tool (PRT) based on a rubric-based qualitative model, (b) a shared visualization tool (SVT) used as a feedback tool for peer ratings with an intuitive interactive interface

supporting both individual and group awareness. PRT allows the group members to rate peer cognitive contribution chunked into posts, and shares this information anonymously with all group members. Rating is the parameter our fPIv (flexible PI visualization) system monitors. Rating in fPIv is based on the same models applied in [4]. Specific examples were given in a complete guiding manual to peers on how to rate.


**Fig. 1.** Data representation of group collaboration in Moodle-forum.

The study lasted 4 weeks (May-June, 2015). The subject was the history of Thessaloniki. The study had 6 phases: pre-test, familiarization, study, discussion, post-test, and interviews. Pre-test was a written test, containing 20 close-type questions (e.g. "When was the White Tower built?") measuring prior subject knowledge. After pre-test, the students were introduced to the learning environment and they were allowed a 4-week prolonged study period to familiarize themselves with the system and the peer rating process on a test discussion subject. During the study phase that followed and which lasted 4 more weeks, the students had to study 10 most important sightseeing of the city within a presentation deliverable. Collaborative work and deliverable could only be produced within the Moodle forum. Peer rating was obligatory and a grade penalty was introduced for the students that did not rate their peers' posts. Then, the students filled a post-test containing 20 questions similar to the ones in the pre-test, and a questionnaire focusing on the peer rating process and the tools used. The study was concluded by interviews, in which the students had the opportunity to further elaborate their thoughts on the whole activity.

### 2.2 Data collection and analysis

Logs & expert ratings: Students' activity within each tool was monitored. Data logs per peer included: the number of posts sent, the number of posts read, time of posts read, access time of resources like forum or visualization tool, time and duration of visualizations viewed. Moreover, three teacher experts rated independently and offline the posts of all collaborating peers a few days after case study and system was

closed. These ratings were logged in the same PRT tool we introduced along Moodle-based asynchronous forum used for group discussions.

Interviews**:** All students were interviewed individually for 15 minutes with focus on deeper understanding students' comments and suggestions. The interviews were transcribed and the classification of conclusions concerning both interviews and open-type question answers was the by-product of interviews analysis.

Analysis**:** Peer ratings were compared to expert ratings. The two-way random average measures (absolute agreement) intra-class correlation coefficient (ICC) was used as a measure of inter-rater reliability. In general, data analysis followed the principles of a mixed evaluation method [19].

# 3 Results

## 3.1 Logs & expert ratings.

L1: The students read the received ratings, then they rate and finally post. This is the most common strategy followed by peers when interacting within the fPIv system provided. Students post or rate, after reading PR feedback, 97% of the time, while they read PR feedback after posting or rating 73% of the time.
L2: The students are likely to use the PR visual representation tool during the whole activity to monitor their collaboration. The average time spent (per student, per day) on PR feedback information is approximately 5 minutes (almost 4% of their activity time when logged in the system).
L3: Students did not need to spend much time in reading guide instructions (average: 5.7 minutes), and during collaboration they seemed to have "internalized" the rating guidance given (average: 2.3 views).
L4: Providing PR seems to trigger PI (i.e. posts and replies). A new post appears after a student has received peer rating(s). This is related to L1.
L5: ICC was very high for ratings among peers (.91), expert teachers (.99), and peers and teachers combined (.87).

## 3.2 Interviews

The list below shows the most important findings recorded during the interviews.
I1: Providing PR was an easy task (93%) because of the guiding rubric used.
I2: The strategy of peers was driven by PR process (83%). The students first studied the ratings of their peers and then formed their strategy posting, replying and rating. This finding is in accordance with L1 above.
I3: Raw table data should be accompanied by simple visualizations like bar charts preferably (81% of students). This helps students evaluate the raw data and draw conclusions, promoting self and group awareness during collaboration.
I5: Students wanted to have some statistical data depicting the PR feedback of the whole class (74%). Thus, PR presentation is covering three levels (see fig. 1): a) Individual, b) Group, c) All groups working in parallel in an activity.
I6: Students (93%) opted for anonymous PR (otherwise rating would be biased).

## 4 Discussion-Conclusion

In relation to the major research question posed "*To what extent can peer rating be reliable, when compared with expert rating?*" we can state that:

- A tool like PRT, when guided and anonymous, can motivate and support effectively students perform reliable on-line rating (when compared to that of external off-line expert raters)
- A tool like SVT can foster collaboration balance in small groups (up to 4 peers).

PRT anonymously shares all perceived and received ratings of peers, in order to make them more aware of the collaboration process and the way peers rate his/her posts (L1, L3, I6). SVT stimulates individual reflection on discrepancies between self and received peer ratings, and stimulates peera to reflect collaboratively upon their group performance. This reflection process allows group members to reach a shared view about what can be referred to as valuable post contributions (see L2, L4, I5).

Because group members' peer ratings are shared in SVT, all group members receive information on their peers' contributions. The strength of PRT and SVT emerge from its ability to make implicit aspects of collaboration (e.g., rated posts among peers) explicit for all group members. PRT and SVT enhance students' awareness of performance, by providing them with explicit information concerning their performance (e.g., contributing low quality work). Based on aforementioned findings we can state that a PR process that is anonymous and guided can provide a good experience to students. Students' ratings compared to teachers' ratings exhibited scarce and small deviations. Relevant tests reveal that on-line PR, if performed with these prerequisites, can be as respectable as offline expert rating (L5). That is the key element that allows for building an on-line reflection tool like SVT.

A shared visualization tool helps discussion stay on-task. From log files (see L1), we notice that a peer before writing a post and/or rating peers, he/she studies on PR visual feedback. This is aligned with study [15], where students, before contributing in the discussion, they gathered information about current balance over participation and formed a strategy to achieve a balanced participation over their discussion.

In this work, we reached similar conclusions as in [16]. There, authors tried to boost their motivation by building a positive sense of competition using a representation of average class performance. That is why -in fPIv- we have opted to use information to the student not only for the group he/she is in but also for the class he/she participates in. In our work, we notice that enhancing interpersonal behavior positively affects the group's balance (L2, L3 and L4) positively.

Overall, the effects of PRT and SVT on group members' individual behavior and their social group performance look very promising. To our knowledge, there is no concise conclusion in previous research to what extent peer rating assessment and reflection feedback affect group behavior and performance, and what kind of reflection feedback lead to effective reflection processes (e.g., [6], [7]). Therefore, this work contributes as an instigator towards studies that examine a combination of guided and anonymous peer rating supplemented with visualized feedback.

## 5 References


1. Dillenbourg, P.: What do you mean by collaborative learning?. Collaborative-learning: Cognitive and Computational Approaches., 1-19 (1999)
2. Vatrapu, R., Teplovs, C., Fujita, N., Bull, S.: Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge, pp. 93-98, ACM (2011)
3. Boud, D., Keogh, R., Walker, D.: Reflection: Turning learning into experience (1985).
4. De Wever, B., Schellens, T., Valcke, M., & Van Keer, H.: Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. Computers & Education, vol. 46, pp. 6-28. (2006)
5. Veerman, A., Veldhuis-Diermanse, E.: Collaborative learning through electronic knowledge construction in academic education. Collaborative learning, reasoning, and technology, 323-354 (2006).
6. Prins, F. J., Sluijsmans, D. M., Kirschner, P. A., Strijbos, J. W.: Formative peer assessment in a CSCL environment: A case study. Assessment & Evaluation in Higher Education, vol. 30, no 4, pp. 417-444 (2005)
7. Chen, N. S., Wei, C. W., Wu, K. T., & Uden, L.: Effects of high level prompts and peer assessment on online students' reflection levels. Computers & Education, vol. 52, no.4, pp. 283-291. (2009).
8. Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. Computers in Human Behavior, 27(3), 1046-1058.
9. Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. Journal of Applied Psychology, 84(1), 58.
10. Kollock, P.: The economies of online cooperation. Communities in cyberspace, 220 (1999)
11. Cheng, R., Vassileva, J. Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. User Modeling and User-Adapted Interaction, vol. 16, pp. 321-348. (2006).
12. Cho, K., Schunn, C. D., Wilson, R.W.: Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. Journal of Educational Psychology, vol. 98(4), pp. 891. (2006)
13. McGraw, K. O., Wong, S. P.: Forming inferences about some intraclass correlation coefficients. Psychological methods, vol. 1, pp. 30. (1996)
14. Dehler, J., Bodemer, D., Buder, J., Hesse, F.W.: Guiding knowledge communication in CSCL via group knowledge awareness. Computers in Human Behavior, vol. 27, pp. 1068-1078 (2011).
15. Bachour, K., Kaplan, F., Dillenbourg, P. Reflect: An interactive table for regulating face-to-face collaborative learning. In Times of Convergence. Technologies Across Learning Contexts, pp. 39-48. Springer Berlin Heidelberg (2008).
16. Falakmasir, M. H., Hsiao, I. H., Mazzola, L., Grant, N., Brusilovsky, P.: The Impact of Social Performance Visualization on Students. In IEEE 12th International Conference on Advanced Learning Technologies (ICALT), pp. 565-569 (2012).
17. Computer Literacy Survey, http:// www.fscj.edu/tutorials/media/ online_skill_assess.pdf (accessed January 2, 2016)
18. Beile, P.M.: Development and Validation of the Beile Test of Information Literacy for Education (B-TILED)". College of Education, Central Florida Univ., PhD Thesis, 2005.
19. Creswell J.W.: Research design: Qualitative, quantitative, and mixed methods approaches. Edition (2nd ed.). Sage Publications, Inc., 2002.