

# How Revealing Rankings Affects Student Attitude and Performance in a Peer Review Learning Environment

Pantelis M. Papadopoulos<sup>1</sup>(✉), Thomas Lagkas<sup>2</sup>, and Stavros N. Demetriadis<sup>3</sup>

<sup>1</sup> Aarhus University, Aarhus, Denmark  
pmpapad@tdm.au.dk

<sup>2</sup> The University of Sheffield – CITY College, Thessaloniki, Greece  
t.lagkas@sheffield.ac.uk

<sup>3</sup> Aristotle University of Thessaloniki, Thessaloniki, Greece  
sdemetri@csd.auth.gr

**Abstract.** This paper investigates the possible benefits as well as the overall impact on the behaviour of students within a learning environment, which is based on double-blinding reviewing of freely selected peer works. Fifty-six sophomore students majoring in Informatics and Telecommunications Engineering volunteered to participate in the study. The experiment was conducted in a controlled environment, according to which students were divided into three groups of different conditions: control, usage data, usage and ranking data. No additional information was made available to students in the control condition. The students that participated in the other two conditions were provided with their usage information (logins, peer work viewed/reviewed, etc.), while members of the last group could also have access to ranking information about their positioning in their group, based on their usage data. According to our findings, students' performance between the groups were comparable, however, the Ranking group revealed differences in the resulted behavior among its members. Specifically, awareness of ranking information mostly benefited students which were relatively in favor of ranking, motivating them to further engage and perform better than the rest of the group members.

**Keywords:** Computer science education · Free-selection · Gamification · Motivation · Peer review

## 1 Introduction

This paper investigates the prospective of exploiting automatically generated ranking information in improving students' engagement and performance in the context of peer reviewing. A variety of different solutions have been proposed and evaluated by the research community for applying the peer review concept in a learning environment, aiming at the effective acquisition of domain specific knowledge, as well as of reviewing skills. It has been shown by related studies that peer reviewing can notably facilitate low level learning [1], such as introductory knowledge, besides enhancing analysis, synthesis, and evaluation that typically count as high level skills [2]. It has been reported

that peer review can significantly stimulate higher cognitive learning processes [3] and in that manner students can be engaged in a constructive and collaborative learning activity [4]. The notable benefits offered by peer reviewing in learning have driven multiple researchers in applying such a technique in various knowledge domain. For instance, related schemes have been utilized for teaching statistics [5], second language writing [6], and computer science [7]. The latter actually constitutes the knowledge domain of this specific study.

This type of studies involve participants who play different roles, more than one most of the times. These roles may be referred in different ways, since they is no globally accepted naming scheme. For instance, in [8], students are classified as “givers” or “receiver”. In [9], similar to the previous roles, students are called “assessors” or “assesses”. Regardless the naming conventions, each role is typically associated with specific learning outcomes. An indication of learning outcomes differentiation depending on the participant role is found in [9], where it was shown that the quality of the provided feedback was related with students’ performance in their final project. No such relationship was revealed between received feedback and final project quality. Similar findings were reported in [8]; students who performed reviews to other students’ writing exhibited significantly higher performance than students who just received peer reviews.

In our study, we are focusing on a peer review setting that is based on free selection, which means that there is no central authority, instructor, or information system, that assigns specific reviews to specific students. We call this peer reviewing scheme “Free-Selection” and it has been the focus of our previous work in [10], where its potential were thoroughly study. In more detail, following this protocol, it was revealed that removing any constraints dictated by specific allocations, students tend to engage more in the reviewing activity when they are given the freedom to provide feedback on answers they choose. In such a manner, the students are exposed to a larger variety of opinions and arguments, enhancing their comprehension on the taught material. As a result, students which are allowed to follow the Free-Selection model are benefited in two ways: they are provided with the opportunity to gain deeper knowledge on the domain specific field, while they are able to significantly improve their peer review skills, compared to students which were assigned review in a random way. The promising potentials of freely selecting which peer work to review have been also demonstrated by other related studies. For instance, two related online platforms, PeerWise<sup>1</sup> and curriculearn<sup>2</sup>, which are mainly focusing on facilitating students to generate their own questions, allow participants to review as many answers as they want. The recorded effect is that more reviews come from high performing students, leading to more good-quality reviews in the system [11].

Participants’ engagement is proven to play a crucial role in the success of peer reviewing systems, especially when Free-Selection is applied. Towards this direction, it is important to find efficient ways of keeping students’ interest at high levels, so that they remain as active as possible. A promising approach is adopting gamification, which is defined as the employment of game elements in non-game contexts [12]. For instance,

---

<sup>1</sup> <http://peerwise.cs.auckland.ac.nz/>.

<sup>2</sup> <http://curriculearn.dk/src/index.php>.

the study findings reported in [13] reveal that when more than 1000 students were engaged in a PairWise-based activity which awarded virtual badges for achievements, their activity was notably increased since a larger quantity of work as well as feedback were submitted.

However, in our case such a gamification scheme would not be applicable. The main reason is related with the relatively limited number of participants (typically 50 to 60 students) as well as the short duration of the whole activity within a course (2 to 3 weeks), which do not allow the production of a sufficiently large volume of peer work and feedback. Hence, a decision was made to investigate some other innovative way of engaging students in such a learning activity; namely the provision of usage and ranking information. Specifically, we decided to use as usage indicators of students engagements the following metrics: the number of student logins in the employed online learning platform, the number of peer submissions read by students, and the number of reviews submitted by students. We refer to these metrics as usage information and we exploit them to enhance student engagement. Taking a step further, we decided to generate some metadata from the usage information, which potentially could be considered interesting by students, so that their engagement is even more enhanced. We actually refer to ranking information, which makes students aware of their relative position in their group regarding the individual usage metrics.

The rest of the paper is structured as follows. Section 2 describes in detail the adopted methodology for the whole experiment setup and results extraction. The collected results are presented and commented in Sect. 3. The following section concludes the paper by providing a thorough discussion on the findings as well as future directions.

## 2 Methodology

### 2.1 Participating Students

The specific activity took place in the context of a course titled “Network Planning and Design”. This is one of the core courses which is offered in the fourth semester of the 5-year study program “Informatics and Telecommunications Engineering” in a typical Greek Engineering School. The main learning outcomes of this course are related with the ability to analyze clients’ requirements, identify the specifications of the system, and design computer networks that comply with the above. Hence, the students taking this course need to have an adequately strong technical profile. The study was conducted with students who volunteered to participate in this learning activity and get a bonus grade for that part of the course that was associated with lab sessions. Finally, 56 students volunteered to take part and complete the activity. Participating students were randomly assigned to three groups with different treatment according to the conditions of the conducted study. Specifically, the groups were formed as follows:

Control: 18 students (11 male, 7 female)

Usage: 20 students (11 male, 9 female)

Ranking: 18 students (10 male, 8 female)

This different treatment that students received based on the above classification was exclusively related with the information that was made available to them. The tasks that had to be completed in the activity context were actually identical for all students. Specifically, no extra information related with the learning activity, either usage or ranking, was provided to the students belonging to the Control group. On the other hand, students in the Usage group received information about their usage metrics, whereas Ranking group students were provided with extra information about their ranking within their group revealing indications related to their individual progress.

## 2.2 Instruction Domain

The domain of instruction was “Network Planning and Design”, which is a typical ill-structured domain characterized by complexity and irregularity. Any Network Planning and Design project involves a number of designing decisions that are mainly based on the subjective analysis of clients requirements. In fact, the engineer is expected to solve an ill-defined problem described by the client. In order to accomplish this task, the designer needs to take numerous factors into account, which are also not fixed and are mainly related with technical details and cost constraints. In particular, a loosely describe process needs to be followed in order to proposed an appropriate network topology which combines suitable network components in a manner that requirements are adequately met. The nature of this type of project outcomes is effectively described in [14], where it is stated that the results of the followed process are actually derived by analyzing the requirements set by the client, while balancing the adopted technology against cost constraints. For these reasons, an efficient way of teaching NP&D would involve studying of realistic situations. In such a manner, students could significantly benefit by learning from existing experience and they would be provided with the opportunity to build their own technical profile. As it is supported by authors in [15], project-based instruction methods could notably benefit Computer Engineering students and help them learn to handle real world complex problems.

## 2.3 Material

**The Online Learning Platform.** This study utilizes a learning environment that was designed as a generic online platform able to support activities based on peer reviews. The eCASE platform is a custom made tool that facilitated both individual and collaborative learning in ill-structured domains of knowledge. Employing a custom-made learning environment has the obvious advantage of allowing modifications and adjustments in order to match the required conditions of the specific research activity. The suitability of eCASE for such studies has been already proven in previously published work investigating different aspects of case-based learning [16], of CSCL collaboration models [17], and of peer reviewing schemes [10].

The material held by the system is organized into two main categories: scenarios and cases. Each scenario constitutes the description of a realistic NP&D problem that students are required to solve. Moreover, each scenario focuses on a number of characteristics and factors that affect designing decisions. Students are able to enhance their

understanding and acquire the knowledge required to suggest some solution for a scenario by carefully studying a number of cases that are related with it. Cases are descriptions of similar realistic NP&D problems, along with proposed solutions, which are accompanied by detailed justification. All cases that are associated with a specific case focus on the same domain factors, such as performance requirements, expected traffic type and network user profile, network expansion requirements, and financial limitations. The students are able to produce their own proposed network design as a solution for each scenario after consulting the corresponding scenarios. It is also noted that the presented scenarios and cases try to partially build on the knowledge acquired by studying the previous scenarios and cases. In that manner, the second scenario is more advanced and complex than the first one, while the third scenario is the most advanced and complex of all three.

Apart from the material presented to the students, the online platform is designed to fully support the reviewing process and provide students with usage information depending on the group they belong to. In more detail, the systems made available a review form that was used by students to provide feedback to their peers. As shown in Fig. 1, the review form provided some guidance to the students, focusing on three points, so that feedback fulfills some basic requirements. The system was also responsible for keeping all types of deliverables, for monitoring students' progress, and for controlling access. Although building our own platform definitely gave us increased flexibility, we do not consider the system as part of the analysis for this paper.

### Review Form

The review guidelines presented here aim to help you identify strengths and weaknesses also in your answers, so that you will be able to provide later a better and more comprehensive revised answer.

- 1. Which are the main points of your peer's answer?**  
List in short what your peer suggests in his/her answer.
- 2. Does your peer provide efficient argumentation?**  
Does your peer use solid arguments or does he/she base adequately his/her answer on the provided material?
- 3. Does your peer write clearly and eloquently?**  
Does your peer use correct vocabulary, grammar, phrasing? Does he/she express what he/she wants to say?

Insert your review comments in the form below. Please, follow the suggested guidelines and do not hesitate on your judgment! The goal of this task is to help you provide better revised answer at the end.

1. ---

2. ---

3. ---

Suggest a grade:

submitreset

Fig. 1. Review form.

**Tests.** The study included two test for the students. One given before starting the activity (i.e., the pre-test) and one right after the completion of the activity (i.e., the post-test). The purpose of the former test was to create a reference point for analysis by noting where students were standing regarding prior knowledge on the domain of instruction. Six open-ended questions were used for that purpose (such as “How does knowledge on the end-user profile of a network affect its architecture?”). The pre-test aimed at revealing knowledge acquired through the learning activity. It included 3 open-ended questions (such as “What kind of changes may occur in the architecture of a network, when the expansion requirements are increased?”).

**Questionnaire.** Students were asked to complete an online question after the end of the study. Our purpose was to collect their opinions on a variety of activity-related aspects. For instance, students commented on the following: their reviewers’ identity, the amount of time they devoted in the learning environment during the different activity phases, and the impact of usage/ranking information. A number of both open and closed-type questions were used for building the attitude questionnaire.

## 2.4 Design

This work adopts a widely employed study design approach which dictates comparison of participants’ performance in all groups before treatment and after treatment. The independent variables of our experiments are the usage and ranking information. The dependent variables were related with students’ performance in the learning environment, the pre-test, the post-test, and their recorded opinions in the attitude questionnaire. The whole activity was divided in 6 non-overlapping phases: Pre-test, Study, Review, Revise, Post-test, and Questionnaire.

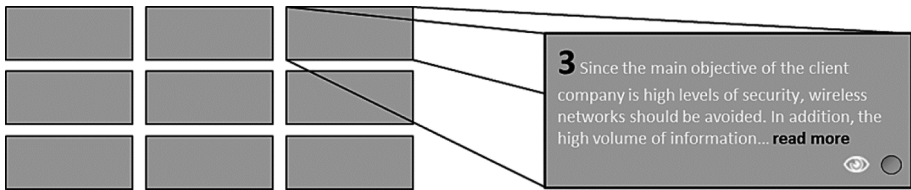
## 2.5 Procedure

The duration and the order of the 6 phases are graphically depicted in Fig. 2. The study started with the Pre-test that took place in class. Right after it we had the Study phase, which lasted for 1 week. During that phase the students were able to study the online material; they went through the past cases and proposed network designs for the corresponding scenarios. Specifically, each one of the 3 scenarios was made available to students along with the associated 2 cases for each scenario. After finishing reading the cases, they were requested to write a description of a network as solution to the scenario and then proceed to the next one.



**Fig. 2.** Activity phases.

The following phase was about reviewing peers' work and lasted 4 days. During the Review phase students were providing feedback to their classmates through eCASE. Since the Free-Selection protocol was adopted, students were free to choose and review any peer work that belonged to the same group. The peer-review process was double-blinded, without any limitation on the number of reviews to provide. However, they were obliged to complete at least one review per scenario. To help them choose, the online platform arranged the start of peers' answers (including in approximate the first 150 words) into a grid, as shown in Fig. 3. The answers were presented in random order, with an icon indicating if the answer is already read (eye) and/or reviewed (green bullet) by the specific student.



**Fig. 3.** An example illustration of the answer grid, where peer answer no. 3 has been read and reviewed by the student.

The user was able to read the whole answer and then review it by clicking on “read more”. A review form (Fig. 1) was the shown to be filled by the student, providing some guidance by emphasizing on three main aspects of the work:

1. The basic points of the work under review.
2. The quality of the arguments provided by the author.
3. The phrasing and expression quality, as well as the overall answer clarity.

The user performing a review was also requested to suggest a grade, based on the comments he provided. The available 5-step options following the Likert scale are shown below:

1. Rejected/Wrong answer
2. Major revisions needed
3. Minor revisions needed
4. Acceptable answer
5. Very good answer

The three student groups, Control, Usage, and Ranking, received different treatment only in terms of the extra information that was presented to them. Apart from that, the actual tasks they had to complete within the activity were identical for all. In that manner, we were able to figure out the impact of the additional information. The main idea was enhancing student engagement, so in that sense we opted for metrics that were directly related with enforcing activity. More specifically, the three usage metrics that we selected aim at motivating students to: (a) visit the online platform more often, (b) collect more points of view by reading what others answered, and (c) provide more reviews to

peers' work. Accordingly, the three corresponding metrics that were made available to the students of the Usage and Ranking groups were the following:

- The number of logins in the online learning platform.
- The total read count of different peers' work.
- The total reviews they provided to different peers' work.

It is worth mentioning that these usage metrics did not provide any additional information, since it is apparent that a user that would keep track of her activity could come up with those values in a manual way. Of course, the students were not requested to perform such a task, and it clearly is much more convenient to have the metrics automatically available by the system. Regarding the extra information that was provided exclusively to the members of the Ranking group, it was revealing relative position to the group's average and as such could not be calculated by the students themselves without the system. Ranking information was presented to the users via chromatic coding to show the quartile according to the ranking (1st: green; 2nd: yellow; 3rd: orange; 4th: red).

Next, the Revise phase started right after the completion of the reviews and allowed students to consider revising their work taking into account the reviews that were now accessible. This phase lasted 3 days to allow students with enough time to read the received comments and potentially improve the network designs they proposed. They were also able to provide some feedback about the reviews they received. In case a student happened not to receive any feedback, the online platform provided her with a self-review form, before proceeding to any revisions of the initial answer. This form guided students on properly comparing their answers against others' approaches, by requesting the following:

1. List their own network components.
2. Justify choosing each component.
3. Grade themselves.
4. Identify and argue for needed revisions.

The Revise phase was considered completed by the system as soon as a student submitted the final version of her answer for each one of the three scenarios, regardless to whether she had eventually performed revisions or not.

It is already noted earlier that in the beginning of the Revision phase, one more metric was available to the Usage and Ranking groups, the average scores received from peers, along with the number of received reviews. This was the one metric that students could not affect.

The information that was shown to each group, along with some example-values, are presented in Table 1. The 2-week activity was completed with the in-class Post-test, which was then followed by the attitudes questionnaire.



**Table 1.** Example of usage/ranking information shown to students in each group.

Metric	Control	Usage	Ranking		
		Value	Value	Mean	Ranking
Number of logins in the environment	<no data>	12	15	9.4	4th
Peer work viewed in total		9	6	13.9	17th
Peer work reviewed in total		5	5	4.5	6th
Score received from peers (no of peers)		3.8 (5)	3.1 (4)	3.4	11th

### 3 Results

To avoid biases, students' answer sheets of the pre and post-test were mixed and blindly assessed by two raters that followed predefined grading instructions. The statistical analysis performed on the collected results chose a level of significance at .05. The parametric test assumptions were investigated before proceeding with any actual statistical tests. It was found that none of the assumptions were violated. On that ground, we first examined the reliability between the two raters' marks. The computed two-way random average measures (absolute agreement) intraclass correlation coefficient (ICC) revealed high level of agreement ( $>0.8$  for each variable) indicating high inter-rater reliability.

#### 3.1 Pre-test and Post-test Results

Students' scores in both written testes are presented in Table 2. As expected, pre-test scores were notably low, since students were novice in the specific domain of knowledge. Moreover, it was shown via analysis of variance (ANOVA) that in all three groups students achieved comparable scores, indicating that the distribution of students in the three groups was indeed random. In order to compare student performance after the completion of the activity in all groups, one-way analysis of covariance (ANCOVA) was performed, employing as covariate the pre-test score. The respective results showed again no statistical differences ( $p > .05$ ) between the three groups.

**Table 2.** Scores in pre-test and post-test.

(scale: 0–10)	Control			Usage			Ranking		
	M	SD	n	M	SD	n	M	SD	n
Pre-test	2.11	(0.98)	18	1.98	(1.01)	20	2.02	(1.09)	18
Post-test	8.17	(1.01)	18	8.10	(0.96)	20	8.11	(0.93)	18

### 3.2 Performance in Learning Environment

Apart from the pre-test and the post-test, the two raters also marked students’ work in the online learning platform. Table 3 reveals statistics about the 3-scenarios average scores that students’ initial and revised answers received from the two raters. The Likert scale used was the same with the one used by reviewers (1: Rejected/Wrong answer; 2: Major revisions needed; 3: Minor revisions needed; 4: Acceptable answer; 5: Very good answer).

**Table 3.** Scores in learning environment.

(scale: 1–5)	Control			Usage			Ranking		
	M	SD	n	M	SD	n	M	SD	n
Initial	2.94	(0.58)	18	3.12	(0.75)	20	2.80	(0.77)	18
Revised	3.44	(0.67)	18	3.64	(0.58)	20	3.49	(0.77)	18

Paired-samples t-test results between the initial and the revised score for each group showed significant statistical difference ( $p < .05$ ) in all three groups. This indicates that students were benefited from the peer reviewing process. However, similarly to the writing tests score, the comparison between the three groups does not reveal any statistically significant different ( $p > .05$ ). This finding came from the one-way ANOVA test that was applied on the two variables (initial score and revised score). The inference is that presenting usage/ranking information had no significant impact on students’ performance in the learning environment. Similar conclusion can be drawn from the usage metrics presented in Table 4. As with the learning environment performance, the comparison between the three groups revealed no significant difference.

**Table 4.** Values of usage metrics.

	Control			Usage			Ranking		
	M	SD	n	M	SD	n	M	SD	n
Logins	30.29	(12.90)	18	28.73	(14.02)	20	28.06	(14.44)	18
Views	12.54	(9.26)	18	11.47	(10.57)	20	14.31	(7.69)	18
Reviews	4.31	(1.02)	18	4.18	(1.21)	20	4.21	(0.93)	18
Peer Score	3.61	(0.89)	18	3.56	(0.91)	20	3.51	(1.10)	18

### 3.3 Questionnaire Outcomes

Students’ responses to the most significant questions included in the attitude questionnaire are summarized in Table 5. Looking one by one at the results for each one of the five questions, important conclusions can be made. First of all, based on Q1, students in general did not care to know who reviewed their work. Actually, only 4 out of 56

expressed the interest to learn their reviewers' identity; three just for curiosity and one because of strong objections to the received review.

The second and third questions are about the time students spent in the activity. In fact, we have been keeping track of time students were logged in the system, in order to have an indication of their involvement. However, in many cases, the logged data may not correspond sufficiently to the actual activity duration. For instance, a logged user does not necessarily mean an active one. For that reason, we decided to contrast the logged data to students' responses. Specifically, students from all groups reported comparable time spent during the first week ( $p > .05$ ), according to responses in Q2. On the other hand, students belonging to the Ranking group spent significantly more time ( $F(2,53) = 3.66, p = .03$ ) in the online platform during the second week, according to responses in Q3. It is highly notable that students spent much more time during the first week than during the second week. However, that was expected, since the first week involves studying for the first time all the material and writing all original answers.

**Table 5.** Summary of questionnaire responses.

	Control	Usage	Ranking
	(n = 18)	(n = 20)	(n = 18)
	M (SD)	M (SD)	M (SD)
<b>Q1. Would you like to know the identity of your peers that reviewed your answers? (1: No; 5: Yes)</b>	1.70 (0.80)	1.95 (1.16)	1.84 (1.30)
<b>Q2. How many hours did you spend approximately in the activity during the first week (Study phase)? (in hours)</b>	10.08 (3.09)	10.12 (3.80)	9.59 (3.94)
<b>Q3. How many hours did you spend approximately in the activity during the second week (Review and Revise phases)? (in hours)</b>	2.48 (1.40)	2.33 (1.01)	3.31 (1.13)*
<b>Q4. Would you characterize the usage information presented during the second week as useful or useless? (1: Useless; 5: Useful)</b>	n.a.	4.33 (1.06)	4.26 (0.81)
<b>Q5. Would you characterize the ranking information presented during the second week as useful or useless? (1: Useless; 5: Useful)</b>	n.a.	n.a.	3.53 (1.07)

This activity is definitely more time consuming than reviewing and revising answers.

Responses to question Q4 showed that students of the Usage and Ranking groups had a positive opinion for the usage information presented to them during the second week. In fact, only 3 students claimed that they did not pay any attention and did not find the information important. A follow-up open-ended question about the way usage metrics helped students revealed that the provided information was mostly considered “nice” and “interesting” rather than “useful”. In general, almost all students agreed that being aware of usage data did not really have an impact on the way they studied or on the material comprehension.

However, Ranking group students had different opinions about the presented ranking information. According to the responses provided in Q5, most of the students ( $n = 10$ ) were positive (4: Rather Useful; 5: Useful), whereas 4 of the group members claimed that ranking information was not really useful. The students that had a positive opinion, argued that knowing their ranking motivated them to improve their performance. For instance, a student positioned in the top quartile stated: “I liked knowing my ranking in the group and it was reassuring knowing that my peers graded my answers that high.”

### 3.4 Elaborating on the Ranking Group

In this subsection, we further analyze students’ opinions about ranking information, since their responses to Q5 revealed two different standings. A closer look on how students behaved in the Ranking group showed three main patterns. Performing a case-by-case analysis on the number of logins and the reviews performed by each student, we saw that there were students that tried to improve their rankings and maintain a high position, students whose ranking were relatively stable, and students whose ranking were dropping steadily as the activity progressed.

Figure 4 shows three distinctive cases illustrating the above. Day 8 is the first day of the Review phase in the study. Students’ rankings were recorded by the system before that date, but became available for the first time to them on that date.

During the 4 days of the Review phase, Student A jumps in rankings from position 10 to position 2, with her position improving each day. This means that this student had increasingly more logins in the system than the other students. On the other hand, Student B’s position remains relatively unchanged near the middle of the group population. Finally, Student C had enough logins during the Study phase of the activity to be placed on the 7th place in the beginning of the Review phase. However, there is a constant decline in visiting the system after that.

Similar analysis on the other metrics of the study showed that students reacted differently to the ranking information. This finding, along with the explicitly stated differentiation of students in Q5 made us look further into the performance of the Ranking group.

For this reason, we split the Ranking group in two: (a) the InFavor subgroup that included 10 students who were positive towards ranking information, and (b) the Indifferent subgroup that included the other 8 students who had a neutral or negative opinion against ranking information. We focused on performance metrics, which are summarized in Table 6.

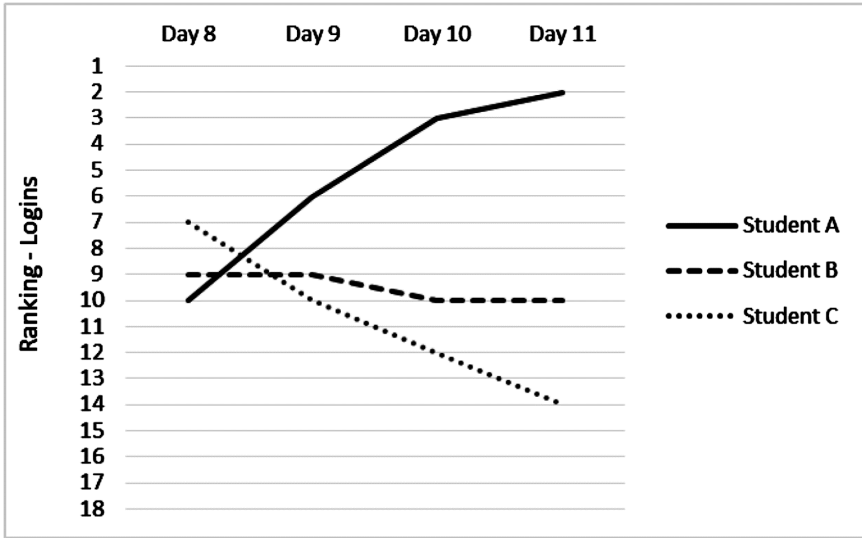


Fig. 4. Login rankings of three students of the ranking group during the review phase.

Table 6. Performance of the two ranking subgroups.

	Logins*		Views*		Reviews		Peer Score*		Post-test*	
	M	SD	M	SD	M	SD	M	SD	M	SD
Indifferent (n = 8)	21.88	(5.17)	10.50	(6.58)	4.00	(1.00)	2.89	(0.93)	7.53	(0.78)
InFavor (n = 10)	33.78	(4.43)	18.22	(7.46)	4.45	(1.13)	3.83	(0.67)	8.59	(1.02)

We performed a t-test analysis to compare the performance of the two sub-groups. The results indicated significant difference between the subgroups in the number of logins ( $t[16] = 5.26, p = .00$ ), the number of peer work read ( $t[16] = 2.29, p = .03$ ), the score they received from peers on the initial answers ( $t[16] = 2.49, p = .02$ ), and the scores in the written post-test ( $t[16] = 2.42, p = .02$ ). There was no significant difference, however, in the number of submitted reviews ( $p > .05$ ). The inference that can be made at this point is that students belonging in the InFavor subgroup were usually representing the top 10 positions in the Ranking group, in terms of performance.

### 4 Discussion

The idea of using ranking information as a means for motivating students requires extra attention, because it can significantly affect the main direction of the whole learning activity. Typically, different schemes adopt badges, reputation scores, achievements,

and rankings to increase student engagement and performance (for instance, students may be encouraged to correctly answer a number of questions in order to receive some badge). It must be highlighted though that this type of treatment does not alternate the settings of the learning activity. This means that receiving badges or any recognition among peers does not change the way the student is acting in the learning environment. This specific aspect actually differentiates this type of rewarding to typical loyalty programs adopted by many companies (e.g. extra discounts for using the company card). Although students do not receive tangible benefits like customers do, it is important that increasing reputation can act as an intrinsic motive for further engagement. However, the specific effect on each student greatly depends on her profile and the way she reacts to this type of stimuli.

The different attitudes towards ranking information became evident from the corresponding question (Q5) in the attitude questionnaire. Our findings showed that students who responded positively, were more active in the learning environment and generally exhibited better performance during the study and at the post-test. The fact that they achieved better scores even before viewing ranking information, specifically higher marks for their initial answers, indicates that these were students who reached higher levels of understanding the domain of knowledge. Although the related statistical analysis results strongly support this argument by revealing significant differences, we report it with caution, since the studied population (students in the two subgroups of the Ranking group) is small for safe conclusions.

Usage information, on the other hand, was widely accepted from students in the Usage and Ranking groups. Despite the positive attitude, just being aware of the absolute metric value does not motivate behavior changes or improved performance. In fact, usage information was considered “nice” by the students, but not enough to make an impact. Apparently, a student was not able to assess a usage metric value about her performance without a reference point or any indication of the group behavior. However, it must be noted that this type of information could probably prove more useful in a longer activity, where a student can track her progress by using previous values for comparison.

While considering all the data that became available to us after the completion of the study, we were able to identify more details about the effect of ranking information. A case-by-case analysis revealed different attitudes even within the Ranking subgroups. For instance, there were students who took into account their rankings throughout the whole duration of the activity, while others considered them only at the beginning. An interesting case is that of two students of the Indifferent subgroup, who expressed negatively about the ranking information, arguing that they disliked it because it increased competition between the students, so they eventually ignored it. What is even more interesting is that system logs showed that both students had post-test scores above the average, while they held one of the top 10 positions in multiple usage metrics. On the other side, a student from the InFavor subgroup focused exclusively on keeping her rankings as high as possible during the whole activity. In fact, she managed to do so for most of the metrics (Logins: 66; Views: 43; Reviews: 3; Peer Score: 3.53; Post-test: 7.80). It is worth mentioning that the specific student only submitted the minimum number of reviews and scored significantly below the average in her initial answers and

the post-test, despite the fact that she had visited eCASE twice the times of her subgroup average and viewed (or at least visited) 43 out of the total 51 answers of her peers. These findings lead us to the conclusion that the student missed the aim of the learning activity (acquisition of domain knowledge and reviewing skills development) and was accidentally misguided to a continuous chase of reputation points. Moreover, the short activity periods that the system recorded in its logs show that this specific student did not really engage into the learning activity, but rather she participated superficially. This type of attitude towards the learning scheme actually matches the definition of “gaming the system” [18], according to which a participant tries to succeed by actively exploiting the properties of a system, rather than reaching learning goals.

The three cases discussed above constitute extreme examples for our study. However, they make clear that depending on the underlying conditions the use of ranking information might sometimes have the opposite effect of the one expected by the instructor. Adding to these three cases, some students mentioned that low ranking would motivate them to improve their performance in order to occupy a higher relative position. Furthermore, it was also reported by students that a high position was reassuring, actually limiting incentives from improvement. One way or another, it is obvious that the point of view greatly depends on the fact that many students have a very subjective opinion about the quality of their work, which is quite different from the raters’ view. Hence, solely relying on ranking information can potentially prove a misleading factor.

Conclusively, the use of ranking information as engagement enhancer for the students could significantly benefit them, particularly in cases that students have a positive attitude towards the revelation of this type of information. In such cases, the intrinsic motivation of students is strengthened. Nevertheless, it is important to closely monitor students’ behavior and reactions during the learning activity, in order to identify cases where ranking information could cause the opposite effect resulting in disengaging.

## References

1. Turner, S., Pérez-Quiñones, M.A., Edwards, S., Chase, J.: Peer review in CS2: conceptual learning. In: Proceedings of SIGCSE 2010, Milwaukee, Wisconsin, USA, 10–13 March 2010
2. Anderson, L.W., Krathwohl, D.R. (eds.): A Taxonomy for Learning, Teaching, and Assessing: a Revision of Bloom’s Taxonomy of Educational Objectives. Longman, NY (2001)
3. Scardamalia, M., Bereiter, C.: Computer support for knowledge-building communities. *J. Learn. Sci.* **3**(3), 265–283 (1994)
4. McConnell, J.: Active and Cooperative Learning. Analysis of Algorithms: An Active Learning Approach. Jones & Bartlett Pub, Burlington (2001)
5. Ashley, K.D., Goldin, I.M.: Peering inside peer review with bayesian models. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 90–97. Springer, Heidelberg (2011)
6. Hansen, J., Liu, J.: Guiding principles for effective peer response. *ELT J.* **59**, 31–38 (2005)
7. Liou, H.C., Peng, Z.Y.: Training effects on computer-mediated peer review. *System* **37**, 514–525 (2009)
8. Lundstrom, K., Baker, W.: To give is better than to receive: the benefits of peer review to the reviewer’s own writing. *J. Second Lang. Writ.* **18**, 30–43 (2009)

9. Li, L., Liu, X., Steckelberg, A.L.: Assessor or assessee: how student learning improves by giving and receiving peer feedback. *Br. J. Educ. Technol.* **41**(3), 525–536 (2010)
10. Papadopoulos, P.M., Lagkas, T.D., Demetriadis, S.N.: How to improve the peer review method: free-selection vs assigned-pair protocol evaluated in a computer networking course. *Comput. Educ.* **59**, 182–195 (2012). Elsevier
11. Luxton-Reilly, A.: A systematic review of tools that support peer assessment. *Comput. Sci. Educ.* **19**(4), 209–232 (2009)
12. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining “gamification”. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek 2011)*, pp. 9–15. ACM, New York (2011)
13. Denny, P.: The effect of virtual achievements on student engagement. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM (2013)
14. Norris, M., Pretty, S.: *Designing the Total Area Network*. Wiley, New York (2000)
15. Martinez-Mones, A., Gomez-Sanchez, E., Dimitriadis, Y.A., Jorin-Abellan, I.M., Rubia-Avi, B., Vega-Gorgojo, G.: Multiple case studies to enhance project-based learning in a computer architecture course. *IEEE Trans. Educ.* **48**(3), 482–489 (2005)
16. Demetriadis, S.N., Papadopoulos, P.M., Stamelos, I.G., Fischer, F.: The effect of scaffolding students’ context-generating cognitive activity in technology-enhanced case-based learning. *Comput. Educ.* **51**(2), 939–954 (2008). Elsevier
17. Papadopoulos, P.M., Demetriadis, S.N., Stamelos, I.G.: Analyzing the role of students’ self-organization in scripted collaboration: a case study. In: *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning – CSCL 2009*, pp. 487–496. ISLS, Rhodes, Greece (2009)
18. Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K.: Why students engage in “gaming the system” behavior in interactive learning environments. *J. Interact. Learn. Res.* **19**(2), 185–224 (2008). Chesapeake, AACE, VA