Integrating Peer Evaluation in the Project Based Adaptive Educational System MyProject

Maria Boubouka, Kyparissia Papanikolaou, Maria Grigoriadou
Department of Informatics & Telecommunications, University of Athens, Greece
{mboub, spap, gregor}@di.uoa.gr

Abstract
In this paper we investigate how peer evaluation mechanisms may promote assessment in a project-based context. In particular, we present a research conducted to investigate the effectiveness of peer evaluation tasks in the web-based educational system named MyProject. Moreover, students' satisfaction in accomplishing peer review and assessment tasks was evaluated. In MyProject, two forms of peer evaluation have been incorporated stimulating students (a) to read and review their peers' ideas or answers to assessment questions, through the accomplishment of the project, (b) after they submit their solutions to the project, to act as reviewers and evaluate their peers' project deliverables and suggest corrections, whilst they should correct their own solutions based on the comments they received from peer-reviewers and respond to the reviewers' comments. Based on the data collected, we identified different types of errors in students' project deliverables. The data analysis provide evidence about the positive influence of the peer assessment method in students' improvement and the potential of the peer review tasks in promoting students' collaboration.

Introduction
Project Based Learning (PBL) is a teaching and learning approach according to which students undertake complex tasks, based on challenging questions or problems, and work autonomously for long periods of time in order to create authentic deliverables (Thomas, 2000). The implementation of PBL requires reorganization of the entire instruction process and necessitates the use of new approaches concerning the students' assessment, as the traditional tests are not adequate in Project Based Learning Environments (Baron et al., 1998; Buck Institute of Education, 2003). In this context, the incorporation of innovative assessment methods like peer review and assessment in a project-based setting is interesting to be investigated. Gouli et al. defined peer assessment as the "activities of learners in which they judge and evaluate the work and/or the performance of their peers" (Gouli et al., 2006). In other words, students are involved in a process that in traditional instruction settings is conducted exclusively by the teacher. Thus, peer assessment contributes to the creation of more student-centered learning environments. In the active involvement of students in the assessment process a new perspective of the assessment is reflected which is no longer considered only as a tool for rating the students' performance but is also recognized as a powerful learning tool (Lai & Lan, 2006). Assessment, rather than being conducted only in the end of the course to measure the results (summative assessment) should be carried on throughout the instruction process (formative assessment) (Noiset, & Gaverni, 1977) and provide appropriate feedback for improvements both to the teacher and the learners (Bhalerao & Ward, 2001). According to researchers, feedback can improve learning outcomes as far as it is provided in time and is personalized. But what about large classrooms, where the student to teacher ratio is high? In such cases, the use of peer assessment may facilitate the provision of analytical feedback (Bhalerao & Ward, 2001). However, the positive results of the method are not limited to the students that receive feedback. It concerns the students that conduct the assessment as well, since while reading the work of their peers, they have the opportunity to reflect on their own work, realizing their errors and deficiencies (Dochy & McDowell, 1997; Sitthiworachart & Joy, 2003). Thus the method fosters awareness and reflection on the behaviour of oneself and of peers, which is a competency highly recognised in the work world (Sluijsmans & Moerkerke, 1999).

On the other hand, in order for the students to be able to accomplish a peer evaluation task, they should have or develop certain skills such as the ability to criticize, to summarize, to provide feedback and identify errors and deficiencies (Van Lehn et al., 1995). Furthermore, it is uncertain that the students will be objective during the process (Ward et al., 2004). Therefore, the method is not always welcomed by the students since they tend to believe that only the teacher is able to perform an adequate assessment and provide feedback (Zhao, 1998) and they also dislike the fact that their assessors are competitors as well (Lin et al., 2001). Such issues should be taking into account when designing or evaluating peer assessment tasks.
Peer assessment has been applied in various disciplines such as algorithms (Chinn, 2005), programming (Chalk & Adeboye, 2005; Sithiworachart & Joy 2007), psychology (Lin et al., 2002) and didactics of informatics (Gouli et al., 2006) with promising results. Furthermore, web-based systems have been created to facilitate the peer assessment process namely NetPeas (Lin et al., 2001), Web-SPA (Lin et al., 2002), Peer Grader (Gehringer, 2001), OASYS (Ward et al., 2004), PECASSE (Gouli et al., 2006). These systems could be used for the peer assessment of the final deliverables of a Project Based course but do not provide support to the students during the elaboration of the projects. On the other hand, systems that support Project Based Learning such as iExpeditions (Wang et al., 1999), Zebu (Tiessen & Ward, 1999), CommSy (Pape et al., 2002), HyperNews (Häkkinen, 2002), NetPBL (Lee & Tsai, 2004) focus mainly in the collaboration support and do not provide a specific peer assessment mechanism.

In this paper, we present a research conducted to investigate the effects of incorporation of peer evaluation tasks in the web-based educational system named MyProject which is a project-based platform. Peer evaluation mechanisms are incorporated in MyProject as well as for reviewing peers answers to the project’s driving questions and small problems embedded in the educational material as for assessing students’ final deliverables of the project. This study was conducted through the formative evaluation of MyProject. Results of this study provide evidence about the effectiveness of the method and the positive attitude of students about peer evaluation tasks.

The Adaptive Educational System MyProject

MyProject (Figure 1) (available in http://hermes.di.uoa.gr:8080/myproject) derived from the formative assessment of the Adaptive Educational System ProSys (Papanikolaou & Grigoriadou, 2006). Both systems’ design is based on constructivist learning theories. In MyProject, learners work with a project and the system proposes them a learning cycle to follow i.e. a sequence of learning activities and aids (such as a case library and individualised navigation advices). The learning cycle (Figure 1) contains four different stages: i) Introduction, ii) Generate Ideas, iii) Multiple Perspectives & Research, and iv) Solution & Evaluation, appearing as links in the left side of the Myproject screen. In particular, in the Introduction stage the project is presented to the learners. The project is an ill-structured problem, not fully defined in order to involve learners into its definition by making their own assumptions. In the Generate Ideas stage, one or more driving questions are posed (Figure 1). The driving questions aim to make the learner think on the basic concepts required for the elaboration of the project. The learners submit their answers to the system. After the submission, they are allowed to access the answers of their peers and they are can state whether they agree, disagree or find one answer indifferent (Figure 1). In the same time, links to the corresponding educational material appear as a scaffold. The learners are allowed to change their answers to the driving questions whenever they want, but their initial answer is always available so that they can monitor the evolution of their ideas. Educational content in the form of realistic cases (Kolodner & Guzdial, 1999), or self-assessment questions (usually embedded in the cases) is provided at the Multiple Perspectives & Research stage. The self-assessment questions may be of multiple choice form or open–ended. The learners’ answers to the multiple choice questions are automatically evaluated by the system. The open ended questions can be reviewed by their peers in the same way as the answers to the driving questions in the Generate Ideas stage, whilst the teacher’s answer becomes available when the learner submits his/her answer. Lastly, the Solution and Evaluation stage is based on a peer review approach which will be described in detail further down. The sequence of the stages within the learning cycles is indicatory and the learners are able to navigate freely throughout the cycle.

The points that differentiate MyProject with ProSys are:

- The learning cycle of MyProject contains four instead of the five stages of ProSys, since the stages Introduction and Project Description of the latter were merged to the Introduction stage of MyProject.
- Additional scaffolding is provided, such as the links between the driving questions with the corresponding educational content, visual cues informing the learners about their progress of study, peer evaluation of self-assessment questions.
- The Solution and Evaluation stage was completely reformed, since in ProSys the Evaluation was conducted using multiple choice questions. Thus, the students assessment was not based at the final deliverable of the project but it was inferred by the students answers to relevant questions. The peer assessment method was implemented in order to assess directly the project’s deliverables.
Peer Evaluation in MyProject

Two forms of peer evaluation are incorporated in MyProject:

i) Peer review of the students’ answers to the driving questions in the Generate Ideas stage and to the open-ended questions in the Multiple Perspectives & Research stage. Each learner’s answer to these questions is followed by the total number of peers that stated agreement, disagreement or indifference for it. On the other hand, all the statements that the learner has submitted to the system concerning the answers of his/her peers are available in his/her learner model where all the commented answers are listed along with data concerning the author of the answer, the date and time the comment was submitted and the comment itself.

ii) Peer assessment of the students’ final deliverables in the Solution and Evaluation stage. The peer assessment is divided in four separate phases. In the first phase, Prosose, the learners can upload their work. In the second phase, Evaluate, they have access to the work of peers that is assigned to them for evaluation. The assignment is done automatically by the system in a random way. For each assigned project they produce a text document, containing their evaluation and feedback, which they upload to the system. In the third phase, Revise, they receive the evaluation of peers for their work and create a revised version. In the same phase, the learners have to assess the evaluation that they received for their work, and in case they decide to ignore suggestions provided by their evaluators they have to justify this choice. Finally, in the fourth phase Reflect the learners have access to other peers’ evaluations for the work they evaluated in the evaluate phase (Figure 2). A time bar is constantly available in the Solution & Evaluation stage, in order to inform students for the current phase of the peer assessment process and the important timestamps. The timestamps as well as the number of evaluations one student has to perform and receive are defined by the teacher through MyProject’s Authoring tool.
Formative evaluation of MyProject

An empirical study was conducted as a part of the formative evaluation of MyProject during the spring semester 2006-2007. In the context of this study, we investigate i) in which extent the peer assessment method conducted through MyProject during the Solution & Evaluation stage supports the improvement of the students’ final deliverables and ii) the students opinions about peer evaluation features of MyProject. Forty undergraduate students of the Department of Informatics and Telecommunications of the University of Athens participated in the study. The students had enrolled in the course Informatics and Education. The aim of this course is to examine how technologies of Informatics and Telecommunications can be integrated in education in order to enhance learning. In the framework of the final assignment of this course, the students had to work for a four-week period with MyProject, elaborating an algorithmic project on the loop structure. The assigned project required the creation of an algorithm to process the results of running races sports events, in order to find the best performance and the athlete that achieved it. More specifically, the students had to study the educational material, fulfil and submit the specific tasks proposed by the system during the Generate Ideas and Multiple Perspectives & Research stages and submit their solutions for the project which went under the peer assessment process. The peer assessment, lasted for ten days, and was further divided into two periods of five days each: in the first (Evaluate phase of the Solution & Evaluation stage of MyProject) each student had to evaluate the submitted solutions of two peers providing analytical feedback, though in the second (Revise phase of the Solution & Evaluation stage of MyProject) each student had to submit a revised solution taking into account the feedback he/she had received and evaluate the received evaluation and feedback.

In parallel with the elaboration of the project, the students had to fulfil a questionnaire created by MyProject’s designing team in order to investigate users’ opinions about the peer evaluation features of MyProject. A few days after the test session, the students granted a semi-structured interview to the researchers/system designers in groups of four. The interviews were structured around specific open questions concerning the usefulness of the peer review and peer assessment features of MyProject, the students’ motivation to engage in these activities and their suggestions for the improvement or extension of these functionalities. Additionally, ambiguous answers to open-ended questions of the questionnaire were clarified.
Data collection and analysis

In order to study the effectiveness of the peer assessment method, we analysed the algorithms that the students submitted as their initial solutions in order to find the most common errors and limitations. Next, we analyzed their revised solutions in order to see whether the defective points were improved. For every defective initial solution we examined also the evaluations that the student received and the solutions he/she evaluated in order to see if he/she received relative feedback from at least one of his/her evaluators and if he/she had evaluated at least one program that was correct concerning the particular error.

On the other hand, the examination of the students’ opinions on the system and its peer review features was based on their answers to the questionnaires and the interviews.

Results

The analysis of the students’ initial solutions revealed different types of errors/limitations that we tried to categorise. Two general categories were revealed: errors linked with the learning outcomes of the project (domain dependent) and those linked with the project-based approach (project dependent). Specially, the second category may be divided in two subcategories: limitations linked with the general context of the project (context-based) or with the project definition itself (project definition). In particular, below we refer instances of the specific categories. Analysing students’ deliverables we acknowledge that the following limitations occurred frequently:

- Searching for the maximum value instead of the minimum (project dependent, context-based)
- Handling of only one race instead of many races (project dependent, project definition)
- Omission of the input data checks (domain dependent)

The first error has to do with the general context of the project, reflecting students’ ignorance of the rules of racing games or negligence and not due to ignorance of the algorithm, as the maximum and minimum algorithms are practically the same. As far as the number of races handled by the algorithm (second limitation/error) is concerned, it may have occurred due to the project-definition interpretation, as the requirement for the handling of more than one races was implicit and not explicit. However we decided to keep track of the improvement concerning this point as the handling of more than one games makes the program more general and the solution is finer. Finally, the third error is related to the omission of input data checks which implies that the students did not perform sufficient testing of their solutions with various data sets.

Comparing the students’ initial with their final solutions we observed that the peer assessment process managed to eliminate the error concerning the searching for the maximum value from the final solutions (Figure 3). Besides, ten students that in their initial solutions handled one race revised them in order to handle many races in their final solutions. One of them had received only relevant feedback from at least one of his/her evaluators (i.e. he didn’t have the chance to evaluate correct peer solutions), though the remaining nine in addition to the relevant feedback, they had also evaluated at least one correct peer solution. What is worthy of remark is that 2 out of the 4 students that had received feedback containing a suggestion to revise their solution in order to handle more than one race, did not accept the suggestion. In their response to reviewer’s comments, they commented that they would rather insist on their interpretation of the definition of the project. Moreover, none of the four students that did not receive relevant feedback, changed the number of races from one to many in their final solution, regardless of whether they evaluated solutions that handled many races or not. As far as the data check omission error is concerned, although the total number of improved solutions was very impressive (22 final solutions contained input data checks compared with the only 5 initial solutions) the majority of the students that revised successfully their solutions had received relevant feedback. However, one of the four students that evaluated correct solutions without receiving relevant feedback, managed to revise his/her solution as well. Finally, eleven of the students that did not revise their answers had neither feedback nor correct solutions at their disposal, raising a question about the number of reviewers and the quality of feedback provided.

Summarizing our findings, we can say that through the peer assessment process, students managed to fully eliminate the context-based error (due to their ignorance of context-based information) whilst it considerably helped them to correct the other types of errors as well. As far as the error related with the project definition is concerned, relevant feedback seems to strongly influence students in revising their solutions towards an optimum one (although in some cases, the students refused to revise their solutions despite the relevant feedback insisting in their own interpretation of the project description). But reviewing optimum solutions, seems also as a potential factor in influencing students towards an optimum solution, which needs further consideration. Finally, concerning the domain depended error,
the feedback received is proven to be an important factor, but the evaluation of correct solutions seems to influence the students as well.

**Figure 3:** The number of final solutions that are correct (left diagram) or defective (right diagram) according to each particular error/limitation with respect to whether the student I. evaluated at least one correct peer solution, II. received relevant feedback from at least one of his/her evaluators, III. neither evaluated correct solutions nor received relevant feedback IV. both evaluated correct solutions and received relevant feedback, or V. had submitted an initial solution that was correct according to the particular error/limitation.

As far as the investigation of the students’ opinions on the peer review features of My Project is concerned we examined the students’ answers in the questionnaires and the interviews.

According to the students’ answers to the questionnaires, the peer review feature of MyProject concerning the students’ answers in the driving questions in the Generate Ideas stage and the open ended questions in Multiple Perspectives & Research stage managed in general to support active learning and reflection (see Table 1) recognizing as the most outstanding factor the option of revising their answers. However, during the interviews where this peer review feature was further discussed, even though the vast majority of students (29 out of 39 students) agreed that they find very appealing the possibility to receive comments on their answers by their peers, 15 out of 34 admitted that they did not engage in commenting their peers’ answers. The same discussion brought up interesting points about the motives for commenting peers’ answers and additional functionality required from the system to support it. In particular, they answered that the most salient reason which would motivate them to comment on their peers answers is to express disagreement. Thus, they would like to have the opportunity to provide elaborate comments either through a comment mechanism or through a discussion forum.

**Table 1:** Students’ opinions on the features of MyProject that promote Reflection and Active Learning as expressed in relevant question in the questionnaire. The numbers indicate number of students per answer.

<table>
<thead>
<tr>
<th>MyProject fosters Reflection and Active learning by enabling:</th>
<th>Not at all (-2)</th>
<th>Very much (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>the submission of my answers in the driving questions and open-ended questions in a data base where peers can see them and review them</td>
<td>2 0 1</td>
<td>18 18</td>
</tr>
<tr>
<td>the opportunity to see and review answers of peers in the driving questions and open-ended questions</td>
<td>2 1 6</td>
<td>13 17</td>
</tr>
<tr>
<td>the revision of my answers</td>
<td>1 0 1</td>
<td>6 31</td>
</tr>
</tbody>
</table>
Moreover, the students’ opinions about the peer assessment in the *Solution & Evaluation* stage were rather positive, as 80% answered that the peer assessment process managed to support successfully the assessment and self-assessment of the project in the relative answers of the questionnaire. Besides, all of them agreed that the peer assessment fosters critical thinking, stressing the contribution of the activities of peers’ evaluation and revision of their own projects (table 2).

**Table 2:** Students’ opinions on how the particular activities in which the students engage in the Peer Assessment process during the *Solution & Evaluation* stage of MyProject contribute in the promotion of Critical Thinking as expressed in relevant question in the questionnaire. The numbers indicate number of students per answer.

| Critical thinking is being fostered during the Peer Assessment through the engagement in: | Not at all | Very much |
| --- | --- | --- | --- | --- |
| | -2 | -1 | 0 | 1 | 2 |
| o the evaluation of peers’ solutions | 2 | 0 | 1 | 18 | 18 |
| o the evaluation and commenting of the received feedback | 2 | 1 | 6 | 13 | 17 |
| o the revision of my solution | 1 | 0 | 1 | 6 | 31 |

**Conclusions - Future Plans**

The integration of peer review and assessment features in a Project Based Learning Environment was quite promising. Results from the empirical study conducted with undergraduate students in a project on algorithms provided evidence for the potential of peer assessment in promoting students improvement. Furthermore, the students that used the system commented positively the peer assessment and review features of the system, and in particular the possibility to revise their work based on their peers’ comments and evaluation.

In the near future we plan to apply the system to more complex projects on a larger sample of students in order to examine the effectiveness of the peer review features. Furthermore, we plan to revise MyProject in order to provide the additional functionalities for supporting communication to meet the requirements that arose from the discussion with the students that participated in the empirical study for the evaluation of the system.

**References**


