Evaluation of an On-line Collaborative Examination Process

Jia Shen  
New Jersey Institute of Technology, jxs1866@njit.edu

Kung-E Cheng  
Rutgers University, kecheng@pegasus.rutgers.edu

Yooncheong Cho  
Rutgers University, yccho@pegasus.rutgers.edu

Starr Roxanne Hiltz  
New Jersey Institute of Technology, roxanne.hiltz@gmail.com

Michael Bieber  
New Jersey Institute of Technology, bieber@njit.edu

Follow this and additional works at: http://aisel.aisnet.org/amcis2000

Recommended Citation
http://aisel.aisnet.org/amcis2000/206
Evaluation of an On-line Collaborative Examination Process

Jia Shen+, jxs1866@njit.edu
Kung-E Cheng*, kecheng@pegasus.rutgers.edu
Yooncheong Cho*, yccho@pegasus.rutgers.edu
Starr Roxanne Hiltz+, roxanne.hiltz@worldnet.att.net
Michael Bieber+, bieber@njit.edu

+ Collaborative Hypermedia Research Lab
Computer and Information Science Department
New Jersey Institute of Technology

* Ph.D. in Management Program
Graduate School-Newark
Rutgers, The State University of New Jersey

Abstract

This paper presents the evaluation results of an on-line collaborative examination process. The process includes students making up questions, picking out questions, answering, grading etc. all in an on-line environment (Virtual Classroom™). Data analysis shows students have a favorable feeling towards this innovative form of exam. More than 55% of subjects agreed that they learned a great deal from having to make up the question, looking at all the other questions, looking at others’ answers, and grading another student’s answer. Also 55% of them would recommend the process be used for future exams. Observations from the instructor are also presented. Lastly, possible improvements in the examination and evaluation process are discussed.

Introduction

The Virtual Classroom™, an innovative program originated at the New Jersey Institute of Technology (NJIT), has brought the university into the homes and workplaces of student through the use of asynchronous computer-mediated communication (CMC). (Hiltz 1994). Although this form of education has been proven effective through years of operation, examinations were seldom conducted on-line. Distance learning students usually have to commute to designated campuses to take exams or have an approved proctor. Before we introduce the evaluation of an on-line examination procedure at NJIT, the paper begins with investigation of the characteristics of collaborative education, the aims of examination, as well as previous research on on-line examinations.

In the literature, collaborative learning is defined as a learning process that emphasizes group or cooperative efforts among faculty and students (Hiltz 1997). It stresses active participation and interaction for both students and instructors. In collaborative learning, instruction is learner-centered rather than teacher-centered, and knowledge is viewed as a social construct, facilitated by peer interaction, evaluation and cooperation. (Bouton 1983; Bruffee 1984; Johnson 1981; Johnson & Johnson 1975). Therefore, the role of teachers changes from transferring knowledge to students to serving as a facilitator in the students’ construction of their own knowledge (Hiltz 1998). An example of a collaborative learning strategy applied in the Virtual Classroom™ is the “seminar” type of interchange in which students become teachers. Individual or small groups of students are responsible for selecting a topic; reading material not assigned to the rest of the class; preparing a written summary for the class on the most important issues; and leading a discussion on this topic (Hiltz 1997).

Studies have shown collaborative learning strategies result in more student involvement with the course (Hiltz 1994), and more engagement in the learning process (Harasim 1990). Collaborative learning methods are more effective than traditional methods in promoting students’ incentive for learning and achievement (Johnson 1981), and enhance their satisfaction with the learning experience.

The collaborative examination process that is evaluated in this paper can be regarded as one form of a collaborative learning process. It maximized students’ involvement in the process, which hopefully should also be a valuable learning experience to them. The collaboration discussed in this paper is not collaboration within groups, but among the class as a whole. Each student had to contribute and interact with other students throughout the process, and the exam was conducted through cooperation with the whole class.

Hay (1996) explains two main educational reasons for examinations: (1) to test the level of factual knowledge, (2) to test the ability to integrate material learned throughout a teaching session. Ebel and Frisbie (1986) point out that the main purposes of a test are to measure student achievement and motivate and direct student learning; and also the process of taking an exam and discussing its scoring should be a richly rewarding learning experience. In addition, the process of constructing tests should cause instructors to think carefully about the goals of instruction in a course.
Previous study of on-line examination processes is quite limited, and mainly conducted for computer programming courses. Though different from the collaborative examination process that we will discuss below, some effectiveness of on-line examination has been reported in these studies. Mason and Woit (1998) highlighted the potential advantages of on-line examinations by encouraging students to attain practical skills, and reducing level of cheating and copying. The analysis by Kumar (1999) indicated that there was a good correlation between written test and on-line test scores. Further, he demonstrated that on-line tests indeed reward better problem-solving skills, since students who finish programming projects on time score among the highest in an on-line test.

Course Background and Examination Procedures

Course Background

The on-line collaborative examination we are describing here was conducted in a graduate-level course for both masters and Ph.D. students in Information Systems at NJIT during Fall 1999, called “CIS 677: Information System Principles.” The aim of the course is to study how information systems and technology can be used effectively by people and organizations. The course covers one topic for each lecture, with related textbook chapters and articles as instruction materials. There are two sections of the course: a Face-to-Face (FTF) section where students meet once a week; and a Distance-Learning (DL) section where students watch video tapes of the lectures. Students in both sections are required to participate in the on-line discussion in Virtual Classroom™ (VC), which is a CMC environment that has been used for education at NJIT for years. Several conferences were established on VC to facilitate on-line discussion of CIS 677 topics.

Collaborative Examination Environment

Usually, one exam is scheduled in the middle of the semester for CIS 677. It is a three-hour in class exam of 3-4 essay questions, with notes of six pages allowed. In the fall 1999 session, the instructor proposed a collaborative examination process. The motivation was first to reduce his own work, which was overloaded that semester, and secondly to test a new method of conducting the exam. With the approval of the program director, this new approach was carried out in both sections, with 73 students in total, where 46 were in the FTF section and 27 in the DL section.

Basic procedures of the examination were: first, each student composes an essay question; next, each student selects one question and answers it; third, the student who created the question grades the answer and provides a justification of the grading. Later, Ph.D. students enrolled in the course do an intermediate review of the grading, and lastly, the instructor provides a final grade.

The whole process was carried out in the Virtual Classroom. The instructor created three conferences for the exam. Conferences are separate discussion areas where students can post and read the comments others post. Each conference has a number and a name. The functions of the three conferences were as follows: (1) The “Fall 99 midterm review conference” where administrative information regarding the midterm process was posted, (2) The “Midterm Question Conference,” which was the main conference where the midterm process was conducted. Students posted their questions, answers, grading justifications, and Ph.D. students’ intermediate grading in this conference and (3) The “Midterm Evaluation Discussion Conference” which was only open to Ph.D. students and the instructor for discussing the grade reviewing process. In addition, several important announcements, such as how to compose a question and grading procedures, were posted on the instructor’s web site. The main procedure of the midterm took place over a 10-day period. Ph.D. students and the instructor did the intermediate and final grading during the following 5 weeks.

Survey and Questionnaire

To evaluate the collaborative on-line examination process, a survey was conducted within a couple of weeks after the main examination process finished. The questionnaire was distributed in the classroom to FTF students and also was put on the web site of the instructor for DL students. 63 questionnaires were returned, with 19 through fax, and 44 collected in classroom. Considering the total number of students in both sections, which is 73, the return rate is good (86.3%).

The questions are modified from ones used in the “Course Questionnaire – Virtual University Project” by Dr. Starr Roxanne Hiltz. We wanted to solicit students’ opinions concerning the following research questions:

Q1: Is the collaborative exam effective in testing the mastery of students’ knowledge?
Q2: Does the exam motivate students’ learning of course material?
Q3: Do students learn from the examination process?
Q4: Are students satisfied with the exam process?

Findings

The raw data, gathered from 63 students, were analyzed using Excel™ and SPSS™. First, the analysis of
data shows the majority of students think the examination process is also a learning process for them:

- 55.6% of subjects agreed that they learned a great deal from having to make up the question.
- 55.5% of subjects agreed that they learned from looking at all the other questions.
- 60.3% of subjects agreed that they learned from looking at others’ answers.
- 74.1% of subjects agreed that they learned from grading another students’ answer.

It is noticeable that participation in the grading process is the activity from which students perceived the most learning. Table 1 shows the detailed percentages of responses in each category from strongly agree to strongly disagree.

We wanted to know students’ satisfaction with the process. Table 2 shows the results of three related questions: "The system provides a comfortable timeframe;" "I enjoyed the flexibility in organizing my resources;" "I felt under much pressure taking the exam this way." The results show students found it an enjoyable experience in general.

Table 1. Percentage of students reporting the learning effect of the examination process

<table>
<thead>
<tr>
<th>Item</th>
<th>Number of Response</th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neutral (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I learned from making up a question</td>
<td>63</td>
<td>4.76%</td>
<td>17.46%</td>
<td>22.22%</td>
<td>42.86%</td>
<td>12.70%</td>
<td>3.41</td>
<td>1.07</td>
</tr>
<tr>
<td>I learned from looking at all the other questions</td>
<td>63</td>
<td>9.52%</td>
<td>14.29%</td>
<td>20.63%</td>
<td>46.03%</td>
<td>9.52%</td>
<td>3.32</td>
<td>1.13</td>
</tr>
<tr>
<td>I learned from looking at others' answers</td>
<td>63</td>
<td>4.76%</td>
<td>14.29%</td>
<td>20.63%</td>
<td>46.03%</td>
<td>14.29%</td>
<td>3.51</td>
<td>1.06</td>
</tr>
<tr>
<td>I learned from grading others' answer</td>
<td>62</td>
<td>3.23%</td>
<td>12.90%</td>
<td>9.68%</td>
<td>46.77%</td>
<td>27.42%</td>
<td>3.82</td>
<td>1.08</td>
</tr>
</tbody>
</table>

Table 2. Percentage of students reporting their satisfaction with the examination process

<table>
<thead>
<tr>
<th>Item</th>
<th>Number of Response</th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neutral (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comfortable Timeframe</td>
<td>44</td>
<td>2.27%</td>
<td>2.27%</td>
<td>15.91%</td>
<td>52.27%</td>
<td>27.27%</td>
<td>4.00</td>
<td>0.86</td>
</tr>
<tr>
<td>Flexibility</td>
<td>44</td>
<td>.00%</td>
<td>2.27%</td>
<td>25.00%</td>
<td>38.64%</td>
<td>34.09%</td>
<td>4.05</td>
<td>0.83</td>
</tr>
<tr>
<td>More Pressure*</td>
<td>43</td>
<td>30.23%</td>
<td>34.88%</td>
<td>18.60%</td>
<td>13.95%</td>
<td>2.33%</td>
<td>3.77*</td>
<td>1.11</td>
</tr>
</tbody>
</table>

*This negative item has been converted to positive in calculating mean.

Table 3. Percentage of students reporting their attitude toward the grading system

<table>
<thead>
<tr>
<th>Item</th>
<th>Number of Response</th>
<th>Strongly Disagree (1)</th>
<th>Disagree (2)</th>
<th>Neutral (3)</th>
<th>Agree (4)</th>
<th>Strongly Agree (5)</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I felt the grading process was fair</td>
<td>44</td>
<td>2.27%</td>
<td>20.45%</td>
<td>25.00%</td>
<td>45.45%</td>
<td>6.82%</td>
<td>3.34</td>
<td>0.96</td>
</tr>
<tr>
<td>Students were not capable of doing the grading*</td>
<td>44</td>
<td>4.55%</td>
<td>38.64%</td>
<td>15.91%</td>
<td>40.91%</td>
<td>.00%</td>
<td>3.07*</td>
<td>1.00</td>
</tr>
<tr>
<td>Ph.D. students were capable of improving the grading</td>
<td>44</td>
<td>6.82%</td>
<td>13.64%</td>
<td>31.82%</td>
<td>43.18%</td>
<td>4.55%</td>
<td>3.25</td>
<td>0.99</td>
</tr>
<tr>
<td>It would be an improvement if the instructor had done all grading*</td>
<td>44</td>
<td>25.00%</td>
<td>15.91%</td>
<td>29.55%</td>
<td>20.45%</td>
<td>9.09%</td>
<td>3.27*</td>
<td>1.30</td>
</tr>
</tbody>
</table>

*This negative item has been converted to positive in calculating mean.
The grading procedure is quite unique in this process. Table 3 is students’ responses to the following questions regarding the grading process:

What is interesting here is that although students reported learning more from participating in grading than in other phases of the process, many of them nevertheless would have preferred the instructor to do all of the grading. Reasons for this discrepancy need to be further explored in the future.

In terms of testing mastery, 59.1% of students agreed with “The exam was successful in enabling me to demonstrate what I learned in class” while 15.9% disagree and 25% were neutral.

Students were also asked to compare this exam process with the traditional one on the following items that are related to learning. Table 4 shows the results:

Considering the collaborative nature of the process, we wanted to know students’ responses to the following two items comparing the on-line exam process with a traditional exam: "I learned to value other points of view" and "I developed new friendships in this class." Results show this process resulted 74.6% of students learning to value other points of view more than a traditional exam, while its impact on friendship development is not so prominent (41% said more than traditional exam). See table 5 for details.

The overall experiences of students are shown below (Figures 1, 2, and 3) with their responses to the following three questions in comparison with a traditional exam: "I mastered the course material;" "I enjoyed the examination process;" and "I would recommend the process in the future.” Results indicate the majority felt they mastered the material and enjoyed the process much more than a traditional exam, and would recommend it for future courses.

Table 4. Percentage of students reporting other learning effect of the process

<table>
<thead>
<tr>
<th>Item</th>
<th>Frequencies of favorable (M + MM)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>I became more interested in the subject.</td>
<td>57.1%</td>
</tr>
<tr>
<td>I learned a great deal of factual material.</td>
<td>52.4%</td>
</tr>
<tr>
<td>I gained a good understanding of basic concepts.</td>
<td>57.1%</td>
</tr>
<tr>
<td>I developed the ability to communicate clearly about this subject.</td>
<td>68.3%</td>
</tr>
<tr>
<td>My skill in critical thinking was increased.</td>
<td>68.2%</td>
</tr>
<tr>
<td>My ability to integrate facts and develop generalizations improved.</td>
<td>73.0%</td>
</tr>
<tr>
<td>I was stimulated to do additional reading.</td>
<td>62.9%</td>
</tr>
<tr>
<td>I became more confident in expressing my ideas.</td>
<td>54.0%</td>
</tr>
<tr>
<td>I was motivated to do my best work.</td>
<td>65.1%</td>
</tr>
</tbody>
</table>

* M: More than traditional exams.
** MM: Much More than traditional exams

Table 5. Percentage of students reporting regarding the collaborative nature of the process

<table>
<thead>
<tr>
<th>Item</th>
<th>Number of Response</th>
<th>Much Less (1)</th>
<th>Less (2)</th>
<th>No Difference (3)</th>
<th>More (4)</th>
<th>Much More (5)</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I learned to value other points of view</td>
<td>63</td>
<td>.00%</td>
<td>.00%</td>
<td>25.40%</td>
<td>60.32%</td>
<td>14.29%</td>
<td>3.89</td>
<td>0.63</td>
</tr>
<tr>
<td>I developed new friendships in this class</td>
<td>63</td>
<td>9.52%</td>
<td>3.17%</td>
<td>46.03%</td>
<td>25.40%</td>
<td>15.87%</td>
<td>3.35</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Figure 1: Distribution of responses to the question: “I mastered the course material”
In addition, DL students show even more favorable feelings toward the process than FTF students. (Table 6) The mean values of DL students on the items “I enjoyed the examination process,” “I would recommend the process,” and “I felt the grading system is fair” are higher than FTF students.

Thus, the study indicates that students report greater subjective satisfaction with the on-line exams in terms of learning process, demonstrating mastery, and satisfaction with the system.

**Instructor’s Observations**

Before being shown the student survey results, the course instructor made the following observations.

He felt that the biggest problem was the duration of the examination process. This was the first time he had conducted such a process, and he found himself developing parts of it as he went along. For example, he did not engage the Ph.D. students as evaluators until after the questions had been answered and disagreements between the students answering and grading arose. The instructor felt he needed to give the Ph.D. students extra time to do the intermediate evaluation, because this was unplanned in their schedules and several worked full time.

The instructor felt that both the initial grading and the level of Ph.D. student intermediate evaluations were inconsistent. The grading guidelines gave just the total points for each portion of the grade, e.g., “grade the synthesis of class materials out of 15 points.” But without stating how to grade each portion, the grades varied widely. Graders wrote in their justification that a portion was “good,” while awarding it anywhere between 8 and 13 points out of 15. Next time he will provide detailed guidelines for scoring each portion to ensure uniform standards.

The instructor actually reviewed only the exams for which the initial grade differed from the intermediate evaluation by more than three points out of 100. Out of approximately 70 exams, he only had to review 15. In these conflicting cases, at first he felt he should remain within the range of the two conflicting grades. After

<table>
<thead>
<tr>
<th>Group Statistics</th>
<th>Section</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment**</td>
<td>DL</td>
<td>21</td>
<td>4.05</td>
<td>1.16</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>FTF</td>
<td>41</td>
<td>3.32</td>
<td>1.23</td>
<td>.19</td>
</tr>
<tr>
<td>Recommend</td>
<td>DL</td>
<td>21</td>
<td>3.76</td>
<td>1.18</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>FTF</td>
<td>40</td>
<td>3.28</td>
<td>1.09</td>
<td>.17</td>
</tr>
<tr>
<td>Fairness in grading</td>
<td>DL</td>
<td>18</td>
<td>3.56</td>
<td>1.04</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>FTF</td>
<td>25</td>
<td>3.16</td>
<td>.90</td>
<td>.18</td>
</tr>
</tbody>
</table>

**Enjoyment is significant at p<.05 level (t-test)**
disagreeing with both grades on a few papers, however, he changed his mind and re-scored a handful of the exams below both the original grade and intermediate evaluation. And in two cases he actually re-scored above both grades.

In reviewing the different intermediate evaluations, he also noted that not all of the Ph.D. students graded well. Two out of the nine Ph.D. students were much too lenient in his opinion, and sometimes he disagreed with much of their written justifications.

Again, the instructor feels that providing detailed grading guidelines for each portion of the grade will bring more consistency and counteract all three of the above problems.

Despite these problems, the grading was well-received by the students. Students could appeal their final grade if they included a self-grade together with a written justification identical in format to those the original grader and intermediate evaluation had to do. Only three students challenged their grades, a far lower number than usual for the instructor's traditional exams, where students also had to submit a written justification as part of the grade change request.

The on-line environment proved more difficult to manage than the instructor originally anticipated. He found the need to make a list of who posed, answered and gave an intermediate evaluation on each question. Perhaps this is obvious in retrospect. The list assisted in addressing emailed problems by students (to see which questions they were referring to). It also helped him realize who had not participated by the various deadlines so he knows whom to email with reminders. (He ended up deducting points from students who did not participate on time.)

Confounding this was a lack of specific subject headers on on-line entries. It was difficult to tell which entries contained answers, intermediate evaluations, etc., which added to the difficulty of handling emailed problems by students. Next time he will give each exam question an ID, and ask students to include the ID and a comment type (question, grade, intermediate evaluation, challenge, etc.), so entries will be easy to identify and search for.

Discussion

In general, students reported a favorable attitude towards this on-line examination experience. Some of the things that they liked best about the process included the active involvement in the exam process, the flexibility and reduction in tension. To quote from their answers to open-ended questions in the questionnaire, they said: “Every student must take part in this process actively and can learn from others. Also student can learn how to evaluate other student’s papers.” “The best was you could really think about the question ahead of time and then posting your version of the answer with thorough organizing and proofreading.” “I don’t have to go to campus. I did the exam actually from India.” “I didn’t have the exam tension. I was confident that I could answer the question.”

Although the evaluation provides favorable results, students also provided suggestions to improve the process. The greatest concern comes from the grading system. 40.9% students agree with “students were not capable of doing the grading.” At the same time, Ph.D. students’ capability to do intermediate grading was agreed to by only 48% of students. Using an assistant grader might solve the problem, provided concrete and detailed grading criteria are designed by the instructor and followed by intermediate graders. But this would remove the learning students gained from grading one another. Other problems included the congestion of the network when all students were trying to get a question; the inequality of questions, etc. Improving network capacity and asking for more than one question from each student may reduce the above problems.

There is also room left for improvement in our evaluation process. Given time constraints, we could only conduct one survey. Using various methods of data collection would provide multiple perspectives on an issue, supply more information on emerging concepts, allow for cross checking, and yield stronger substantiation of constructs (Orlikowski 1993). If more than one exam was provided, as well as if the data were gathered from several courses, longitudinally, the results might be more generalizable.

Acknowledgement

Special thanks to Professor Murray Turoff, Osama Eljabiri, and Morgan Benton for their innovative ideas in constructing the exam process and continuous support throughout the conduction of the evaluation process.

Partial support for this research was provided by an Alfred P. Sloan Foundation grant and by the New Jersey Center for Multimedia Research.

References


