Traditional In-class Examination vs. Collaborative Online Examination in Asynchronous Learning Networks: Field Evaluation Results

Jia Shen (jxs1866@njit.edu)
Kung-E Cheng (kc37@njit.edu)
Michael Bieber (bieber@njit.edu)
Starr Roxanne Hiltz (hiltz@njit.edu)

Collaborative Hypermedia Research Lab
Information Systems Department
New Jersey Institute of Technology

ABSTRACT
Online courses make possible new forms of working and learning together that would be difficult or impossible to use in the classroom-based course. This paper presents field evaluation results comparing the traditional in-class examination and the collaborative online examination using asynchronous learning networks (ALN) in a graduate-level course in a U.S. university. The collaborative online exam includes students making up questions, answering, grading, and appealing the grades. A 1x2 field experiment was designed to evaluate the collaborative exam in comparison with a traditional in-class exam. Survey results (response rate = 81.6%) show an overall favorable attitude towards the collaborative exam, including enjoyability of the exam process, perceived learning, satisfaction, and recommendation for future courses. Significant correlations and differences are found among factors and between the two exam modes. Students’ concerns as well as plans for future research are also discussed in the paper.

Keywords
Collaborative examination, assessment, peer assessment, ALN, asynchronous learning networks, collaborative learning, participatory learning

INTRODUCTION
While Virtual Classrooms (Hiltz, 1994) have become a reality with the advance of networking and other computer technologies, the traditional assessment methods still dominate when it comes to exams. Instructors take full control of the exam process by designing questions, grading and setting up exams, while distance-learning students often have to travel to exam centers or take proctored exams.

Can we take advantage of Asynchronous Learning Networks (ALN) to conduct exams so students can enjoy “anytime, anywhere” flexibility as well as maximize their learning through collaboration? To investigate this question, the Collaborative Online Examination was designed based on collaborative learning theories (Leidner and Jarvenpaa, 1995) and online interaction model in ALN (Benbunan-Fich, Hiltz and Harasim, 2004). A field experiment was conducted at a graduate-level course in a U.S. university to evaluate the effectiveness of the collaborative exam. The purpose of this paper is to present our field experiment results.

Collaborative Learning
Traditional instructor-controlled exam reflects the objectivist learning model (Jonassen, 1993), where learning is regarded as the uncritical absorption of objective knowledge transferred from instructor to student. New assessment approaches are proposed based on constructivism theory where knowledge is constructed actively by learners (Leidner et al., 1995). For example, learner-centered assessment (Huba and Freed, 1999) and classroom assessment (Angelo and Cross, 1993) propose to shift the attention from instructors and teaching to students and learning in assessment.
Collaborative learning is a learner-centered and team-based approach based on constructivism and social learning theories (Vygotsky, 1962), which assumes that learning emerges as learners interact with each other. Studies have shown the superiority of collaborative learning approaches in both face-to-face settings and ALN using GSS (Group Support Systems) (Hiltz, 1988). The collaborative examination that is evaluated in this paper can be regarded as one form of the collaborative learning process. It aims to maximize students’ involvement in the process, which should result in a valuable learning experience. The collaboration discussed in this paper is not collaboration within groups where teams perform individual steps in the process, but among the class as a whole. Each student contributed and interacted with other students throughout the process, and the exam was conducted through cooperation with the whole class.

**Previous Research on Collaborative Assessment and the Collaborative Examination**

With the recognition of the collaborative nature of ALN, a small number of studies have been conducted to incorporate student active participation and collaboration into the assessment process online. With the use of GSS, students’ participation and collaboration were integrated into various phases of collaborative assessment, such as collaborative development of a grading scheme (Kwok and Ma, Feb 1999), question composition (Wilson, 2004), collaborative question answering (Shindler, 2003), and peer and self-grading (Falchikov and Goldfinch, 2000; Topping, 1998).

This study is based on previous research on the collaborative examination, which features student active participation in all phases of exams (Shen, Cheng, Cho, Hiltz and Bieber, 2000; Shen, Hiltz, Cheng, Cho and Bieber, 2001; Wu, Bieber, Hiltz and Han, 2004). Previous studies were conducted in a graduate-level course for both masters and Ph.D. students in Information Systems at a U.S. university. The basic procedures of the collaborative online examination take 3 to 4 weeks:

- Each student creates exam problems.
- The instructor edits the problems if necessary.
- Students choose or the instructor assigns problems to solve.
- Each student solves problems.
- Students grade the solutions to the problems they authored, writing detailed justifications.
- Ph.D. students enrolled in the class do a “second round” evaluation of each solution.
- The instructor gives a final grade for each solution.
- **optional**: Students can dispute their solution’s grade, by evaluating it themselves and writing detailed justifications.
- The instructor resolves the dispute, either keeping or adjusting the solution’s grade.

Overall positive attitudes towards the new exam process were reported in student surveys (Shen et al., 2000; Shen et al., 2001; Wu et al., 2004). This study extends previous research on the collaborative exam using a field experiment and comparison between the traditional exam and the collaborative exam.

**RESEARCH QUESTIONS AND RESEARCH DESIGN**

**Research Model, Questions, and Hypotheses**

As discussed in the previous section, the collaborative online exam can be considered as a collaborative online learning process. Based on the Online Interaction Learning Model by Benbunan-Fich, Hiltz, and Harasim (2004), our collaborative examination research model is presented in figure 1. The model has three levels of factors:

- Input factor, or independent variable: exam mode (collaborative online exam vs. traditional in-class exam)
- Moderator factors, or intervening variables: exam process enjoyability and exam grade.
- Output factors, or dependent variables: perceived learning, satisfaction, perceived fairness in grading, and recommendation for future use.
We are interested in investigating the following questions:

Q1. Did students enjoy the collaborative online exam process?
Q2. Did students think they learned through the exam process?
Q3. Did students think the grading system was fair?
Q4. Would students rather take the traditional exam than the collaborative online exam (are they satisfied)?
Q5. Would students recommend using the collaborative online exam in the future?
Q6. Do exam grades affect students’ perception of the exam process?
Q7. Does enjoyability of the exam process affect students’ perception of learning, grading fairness and satisfaction?
Q8. Do students’ perception of learning, fairness in grading, and satisfaction affect their recommendation of the process being used in the future?
Q9. Are there differences in students’ perception of exam enjoyability, learning, satisfaction, and grading fairness between those taking the traditional in-class exam and those taking the collaborative online exam?

Based on the research model and our previous studies (Shen et al., 2000; Shen et al., 2001), our hypotheses are (with corresponding numbers on the research model):

H1a: Students who participate in the collaborative online exam will report higher enjoyability of the process than students who participate in the traditional in-class exam.
H1b: Students who participate in the collaborative online exam will report higher perceived learning than students who participate in the traditional in-class exam.

H1c: Students who participate in the collaborative online exam will report higher satisfaction than students who participate in the traditional in-class exam.

H1d: Students who participate in the collaborative online exam will report lower perceived fairness in grading than students who participate in the traditional in-class exam.

H2: Students’ exam grade will positively correlate with students’ perceived enjoyability of the exam process.

H3a: Students who perceive higher enjoyability will report higher perception of learning through the exam process.

H3b: Students who perceive higher enjoyability will report higher satisfaction with the exam process.

H3c: Students who perceive higher enjoyability will report higher perception of fairness in grading.

H3d: Students who receive higher grades will report higher perception of learning through the exam process.

H3e: Students who receive higher grades will report higher satisfaction with the exam process.

H3f: Students who receive higher grades will report higher perception of fairness in grading.

H4: Students who report higher perceived learning will recommend the collaborative exam process more.

H5: Students who are more satisfied will recommend the collaborative exam process more.

H6: Students who report higher perceived fairness in grading will recommend the collaborative exam process more.

RESEARCH DESIGN: 1X2 FIELD EXPERIMENT

To evaluate the collaborative online examination process and compare it with a traditional exam, a 1x2 field experiment was designed in a graduate-level course titled “Information Systems Principles” in spring 2000 at a U.S. university. A total of 114 students were enrolled in three sections of the course, each taught by one instructor (one of the instructors was the one who pioneered the collaborative exam). The instructors randomly selected half of the students to participate in the traditional in-class exam (control group), and the other half in the collaborative online exam. Table 1 shows the number of subjects in each condition and the response rate in the survey.

<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of students</th>
<th>No. of students answering the survey</th>
<th>Return Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Online Exam</td>
<td>57</td>
<td>52</td>
<td>91.2%</td>
</tr>
<tr>
<td>Traditional In-class Exam</td>
<td>57</td>
<td>41</td>
<td>71.9%</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>93</td>
<td>81.6%</td>
</tr>
</tbody>
</table>

Table 1. 1x2 Field Experiment and Number of Subjects

As can be seen, the return rate is high. The instructor assigned three extra credit points towards the course grade for participation in the survey. In addition, the convenience of filling out the survey online resulted in high response rate, especially in the collaborative exam condition where students did the exam online.

In the collaborative exam condition, students began by composing essay questions. Next each student was assigned two questions and answered them. Answers were restricted to 1250 words. Third, the student who created the question graded the answer and provided a justification of the grading. Each grade was broken into six sub-grades for the following criteria: framing of the answer, considering all sides of the issue, use of references, synthesis of these references, justification of points made, and overall answer quality. The professors gave detailed grading guidelines for each criterion. Students had access to these from the start of the exam process, so they knew what level of detail would be expected. Fourth, Ph.D. students enrolled in the course did a second round of grading. Lastly, the instructors provided final grades. If the scores of the grader and second round reviewer were within a few points of each other, the instructor assigned the higher score. If the two disagreed, then the instructor graded the question him- or herself. (The instructors ended up re-grading between 20-30%
of the questions, taking advantage of the justifications from the two student graders.) The process was made anonymous on Webboard\textsuperscript{TM} (see details in section 2.3). The entire process was streamlined to 3 weeks.

The traditional exam was closed book, but students were allowed to bring six pages of notes for the traditional exam. To match the number of questions that students in collaborative exam had, two open-ended questions were designed by the instructors. Students had three hours to answer the two questions, and the maximum length of the answers was limited to five pages, which was roughly equivalent to the 1250 word limit of the collaborative exam. Grading criteria and guidelines equivalent to those of the collaborative exam were provided in advance of the exam to students. Thus the instructors used essentially the same criteria for grading both types of exam.

Shortly following the exam, students in both conditions were asked to fill out a survey. The survey was distributed both online and on paper in the classroom for students. Two questionnaires were designed for the two exam modes respectively, with matching questions whenever possible to allow comparison. The questionnaires adopted questions that have been validated in previous studies on the collaborative exam (Shen et al., 2000; Shen et al., 2001), with some additions.

**ALN Tools**

In ALN, collaboration is achieved through computer networks (Alavi and Dufner, 2004; Benbunan-Fich et al., 2004). WebBoard\textsuperscript{TM} was the asynchronous conferencing tool used to facilitate the collaborative exam process and learning being discussed in this paper. WebBoard\textsuperscript{TM} has “conferences”, which are separate threaded discussion areas for a topic where students can post, read and reply to comments of others. One main conference was created in WebBoard for the main exam activities; administration information and student feedback were posted in the regular general administration and feedback conferences for that semester. In addition to Webboard, several important announcements, such as how to compose a question and grading procedures, were posted on the course web site.

**FINDINGS**

The raw data were analyzed using SPSS\textsuperscript{TM}. In this section, we first present the evaluation results of the collaborative online exam process, addressing research questions 1-5. Next, we use factor analysis and correlation analysis to find the underlying factors and the relationship among them, addressing questions 6 -8 and hypotheses 2-6. Finally we compare results between the collaborative online exam and the traditional in-class exam, addressing question 9 and hypotheses 1a-1d.

5-point Likert scales were used throughout the survey and the results are scored on a basis of 1-5, where:

- **SA:** Strongly Agree = 5; **A:** Agree = 4; **N:** Neutral = 3; **D:** Disagree = 2; **SD:** Strongly Disagree = 1

The total number of subjects in tables 2-6 is 52, unless noted otherwise.

**Enjoyability of the collaborative online exam process**

Table 2 answers our first question: did students enjoy the collaborative online exam? Results show that students found the collaborative online exam provides flexibility in having multiple resources and a comfortable timeframe, is less stressful, and allows students to demonstrate what they learned in class.
Table 2. Enjoyability of the Collaborative Exam Process (Cronbach’s alpha=.71, Grand Mean=3.46)

+ : The negative item has been converted to positive in calculating mean.

Learning Effects

Table 3 addresses our second question: did students think they learned through the exam process? The majority of students agreed that all phases of the exam process were part of the learning process. Students learn from making up questions, grading, and from reading all of these entries online from their peers. In addition, students report that the collaborative exam increased critical thinking skills and synthesis skills. Furthermore, students report higher motivation to learn, including doing additional reading and being motivated to do their best work. As we expected from collaborative learning theories, students also report learning to value others’ points of view through this collaborative exam process.

<table>
<thead>
<tr>
<th>Item</th>
<th>SD (1)</th>
<th>D (2)</th>
<th>N (3)</th>
<th>A (4)</th>
<th>SA (5)</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I enjoyed the flexibility of having multiple reference available</td>
<td>5.8%</td>
<td>7.7%</td>
<td>11.5%</td>
<td>40.4%</td>
<td>34.6%</td>
<td>3.90</td>
<td>1.14</td>
</tr>
<tr>
<td>I felt under much pressure taking the exam this way+</td>
<td>11.5%</td>
<td>36.5%</td>
<td>25%</td>
<td>25%</td>
<td>1.9%</td>
<td>3.31+</td>
<td>1.04</td>
</tr>
<tr>
<td>The exam enabled me to demonstrate what I learned in class</td>
<td>0%</td>
<td>13.5%</td>
<td>30.8%</td>
<td>40.4%</td>
<td>15.3%</td>
<td>3.58</td>
<td>.91</td>
</tr>
<tr>
<td>The exam provided a comfortable timeframe</td>
<td>1.9%</td>
<td>5.8%</td>
<td>15.4%</td>
<td>53.8%</td>
<td>23.1%</td>
<td>3.90</td>
<td>0.89</td>
</tr>
<tr>
<td>I enjoyed the exam process</td>
<td>7.7%</td>
<td>13.5%</td>
<td>17.3%</td>
<td>46.2%</td>
<td>15.4%</td>
<td>3.48</td>
<td>1.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>SD (1)</th>
<th>D (2)</th>
<th>N (3)</th>
<th>A (4)</th>
<th>SA (5)</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I learned from making up questions</td>
<td>3.8%</td>
<td>11.5%</td>
<td>19.2%</td>
<td>50%</td>
<td>15.4%</td>
<td>3.62</td>
<td>1.01</td>
</tr>
<tr>
<td>I learned from looking at all the other questions</td>
<td>5.8%</td>
<td>15.4%</td>
<td>36.5%</td>
<td>28.8%</td>
<td>13.5%</td>
<td>3.29</td>
<td>1.07</td>
</tr>
<tr>
<td>I learned from grading students’ answers</td>
<td>5.8%</td>
<td>7.7%</td>
<td>26.9%</td>
<td>46.2%</td>
<td>13.5%</td>
<td>3.54</td>
<td>1.02</td>
</tr>
<tr>
<td>My skill in critical thinking was increased</td>
<td>1.9%</td>
<td>0%</td>
<td>42.3%</td>
<td>34.6%</td>
<td>21.2%</td>
<td>3.73</td>
<td>.866</td>
</tr>
<tr>
<td>My ability to integrate facts and develop synthesis improved</td>
<td>0%</td>
<td>1.9%</td>
<td>36.5%</td>
<td>40.4%</td>
<td>21.2%</td>
<td>3.81</td>
<td>.793</td>
</tr>
<tr>
<td>I was stimulated to do additional reading</td>
<td>0%</td>
<td>11.5%</td>
<td>30.8%</td>
<td>34.6%</td>
<td>23.1%</td>
<td>3.69</td>
<td>.961</td>
</tr>
<tr>
<td>I learned to value other people’s points of view</td>
<td>0%</td>
<td>0%</td>
<td>44.2%</td>
<td>42.3%</td>
<td>13.5%</td>
<td>3.69</td>
<td>.70</td>
</tr>
<tr>
<td>I was motivated to do my best work</td>
<td>1.9%</td>
<td>3.8%</td>
<td>42.3%</td>
<td>26.9%</td>
<td>25%</td>
<td>3.69</td>
<td>.961</td>
</tr>
<tr>
<td>I mastered the course materials</td>
<td>5.8%</td>
<td>7.7%</td>
<td>42.3%</td>
<td>38.5%</td>
<td>5.8%</td>
<td>3.31</td>
<td>.92</td>
</tr>
</tbody>
</table>
Table 3. Perceived Learning in Collaborative Exam (Cronbach’s alpha=.86, Grand Mean=3.65)

Perceived Fairness in Grading

Table 4 addresses our third question: Did students think the grading system was fair? Students’ responses fluctuated around neutral on these questions, similar to results shown in previous studies (Shen et al., 2000; Shen et al., 2001). Students question the grading capability of their peers (mean of 2.63), and express favor towards having the instructors do all grading (mean of 2.25). It is noticeable that the grading criteria given by the instructors are considered as explicit (mean of 3.81), and second round (Ph.D. student) graders’ ability to improve grading fairness was agreed to by the majority (mean of 3.12). This may suggest that having qualified second round graders is a good approach in this exam process.

<table>
<thead>
<tr>
<th>Item</th>
<th>SD (1)</th>
<th>D (2)</th>
<th>N (3)</th>
<th>A (4)</th>
<th>SA (5)</th>
<th>Mean</th>
<th>S. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I felt the grading process was fair</td>
<td>11.5%</td>
<td>25%</td>
<td>19.2%</td>
<td>40.4%</td>
<td>3.8%</td>
<td>3.00</td>
<td>1.14</td>
</tr>
<tr>
<td>Students were not capable of grading</td>
<td>1.9%</td>
<td>19.2%</td>
<td>34.6%</td>
<td>28.8%</td>
<td>15.4%</td>
<td>2.63</td>
<td>1.03</td>
</tr>
<tr>
<td>The grading criteria given by the professor were explicit enough</td>
<td>1.9%</td>
<td>7.7%</td>
<td>25%</td>
<td>38.5%</td>
<td>26.9%</td>
<td>3.81</td>
<td>.99</td>
</tr>
<tr>
<td>Intermediate graders improved grading fairness</td>
<td>7.7%</td>
<td>19.2%</td>
<td>32.7%</td>
<td>34.6%</td>
<td>5.8%</td>
<td>3.12</td>
<td>1.04</td>
</tr>
<tr>
<td>It would have been an improvement if the instructor had done all the grading</td>
<td>0%</td>
<td>15.4%</td>
<td>19.2%</td>
<td>40.4%</td>
<td>25%</td>
<td>2.25</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Table 4. Perceived Fairness in Grading in Collaborative Exam (Cronbach’s alpha = .71, Grand Mean=3.40)

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

Overall satisfaction and Recommendation

Table 5 and 6 address our next two questions: would students prefer to take the traditional exam than the collaborative online exam? Would students recommend using the collaborative online exam in the future? Results indicate that students report subjective satisfaction with the collaborative exam in general, and are willing to recommend using this process in the future.

<table>
<thead>
<tr>
<th>Item</th>
<th>SD (1)</th>
<th>D (2)</th>
<th>N (3)</th>
<th>A (4)</th>
<th>SA (5)</th>
<th>Mean</th>
<th>S. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would rather take the traditional exam</td>
<td>23.1%</td>
<td>21.2%</td>
<td>30.8%</td>
<td>15.4%</td>
<td>9.6%</td>
<td>3.33</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Table 5. Overall Satisfaction with the Collaborative Exam Process

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly Oppose (1)</th>
<th>Oppose (2)</th>
<th>Neutral (3)</th>
<th>Recommend (4)</th>
<th>Strongly Recommend (5)</th>
<th>Mean</th>
<th>S. D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would recommend using this process in the future</td>
<td>3.8%</td>
<td>7.7%</td>
<td>30.8%</td>
<td>32.7%</td>
<td>25%</td>
<td>3.67</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Table 6. Overall Recommendation of Using the Collaborative Exam Process for Other Courses

Factor analysis and correlation analysis

Factor analysis was conducted to find the underlying factors. Three factors were found through principle component factor analysis, each corresponding to one aspect of the students’ responses we discussed earlier. Reliability analysis shows all three factors are highly reliable (Cronbach’s Alpha >.60).

- Enjoyability of the Exam Process (Cronbach’s Alpha = .71)
• Perceived Learning (Cronbach’s Alpha=.86)
• Perceived Fairness in Grading (Cronbach’s Alpha=.71)

Next we conducted correlation analysis to find the relationship between these factors. First we are interested in the question (Q6): Do exam grades affect students’ perception of the exam process? It is necessary to point out that students received their grades before they answered the survey, and we hypothesize that students’ exam grade is positively correlated with students’ perceived enjoyability of the exam process (H2). In addition, we hypothesize that students who receive higher grades would report higher perception of learning through the exam process (H3d), higher satisfaction with the exam process (H3e), and higher perception of fairness in grading (H3f). In this analysis, the average of the student grader’s grade, the intermediate grader’s grade, and the instructor’s final grade was used as the grade for the collaborative exam students. Given that the grade was not normally distributed, we used nonparametric correlation analysis and find exam grade is significantly correlated with the enjoyability factor (Spearman’s rho = .387, p<.05). In addition, the exam grade is also significantly correlated with perceived learning (Spearman’s rho = .321, p<.05), and the correlation with satisfaction and perceived fairness in grading are not significant. The result supports hypotheses 2 and 3d, but not 3e or 3f.

Second in our correlation analysis, we are interested in the question (Q7): Does enjoyability of the exam process affect students’ perception of learning, grading fairness and satisfaction? We hypothesize that students who perceive higher enjoyability report higher perception of learning through the exam process (H3a), higher satisfaction with the exam process (H3b), and higher perception of fairness in grading (H3c). Given that all factors in this analysis are normally distributed, we used parametric correlation analysis and find that the enjoyability factor is significantly correlated with perceived learning (Pearson’s r = .486, p<.01), satisfaction (Pearson’s r = .597, p<.01), and perceived fairness in grading (Pearson’s r = .376, p<.01). The results support hypotheses 3a, 3b, and 3c.

Third, we investigate in the question (Q8): Do students’ perception of learning, fairness in grading, and satisfaction affect their recommendation of the process being used in the future? We hypothesize that recommendation of the process is positively associated with perceived learning (H4), satisfaction (H5), and perceived fairness in grading (H6). Given all factors in this analysis are normally distributed, we used parametric analysis again, and find that recommendation is significantly correlated with perceived learning (Pearson’s r = .431, p<.01), satisfaction (Pearson’s r = .71, p<.01), and perceived fairness in grading (Pearson’s r = .424, p<.01). The results support hypotheses 4, 5, and 6.

Correlation analysis results are shown in figure 2.

Comparing the Collaborative Exam and Traditional Exam

The analysis so far draws from data of the 52 subjects in the collaborative exam condition. Although overall favorable attitudes towards the collaborative exam are revealed, we are also interested in finding how students compare the collaborative exam with a traditional in-class exam. This section investigates the question (Q9): Are there differences in students’ perception of exam enjoyability, learning, satisfaction, and grading fairness between those taking the traditional in-class exam and the collaborative online exam? We hypothesize that compared with the traditional exam condition, students in the collaborative online exam report higher enjoyability of the process (H1a), higher perceived learning (H1b), higher satisfaction (H1c), and lower perceived fairness in grading (H1d).

The common items used in the two questionnaires for the two conditions are used for comparison. Similar factors are extracted from the traditional exam condition, such as enjoyability of the exam process (3 items), perceived learning (6 items), and perceived fairness in grading (2 items). Details cannot be included in this paper due to word limit, but are available upon request. T-test is used in testing the significance of the differences between the two conditions. Results (Table 7) show compared with the traditional exam, students in the collaborative online exam condition significantly enjoyed the process more (mean at 3.11 and 3.46, p<.05), and are significantly more satisfied with the process (mean at 2.78 and 3.33, p<.05). This supports hypothesis H1a and H1c. Students in the collaborative exam condition have lower perception of the fairness in grading as predicted in H1d, (mean at 3.62 and 3.40), however this result is not significant. Our prediction of higher perceived learning is not supported (mean at 3.81 and 3.65), and the result is not significant.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Exam Type</th>
<th>N</th>
<th>Mean</th>
<th>S. D.</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyability of the Exam Process</td>
<td>Traditional</td>
<td>41</td>
<td>3.11</td>
<td>.70</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>52</td>
<td>3.46</td>
<td>.81</td>
<td></td>
</tr>
<tr>
<td>Perceived Learning</td>
<td>Traditional</td>
<td>41</td>
<td>3.81</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>52</td>
<td>3.65</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td>Perceived Fairness in Grading</td>
<td>Traditional</td>
<td>41</td>
<td>3.62</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>52</td>
<td>3.40</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>Overall Satisfaction</td>
<td>Traditional</td>
<td>41</td>
<td>2.78</td>
<td>1.37</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Collaborative</td>
<td>52</td>
<td>3.33</td>
<td>1.26</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Comparison between Traditional Exam and Collaborative Exam

*: Significant at p<.05 level

Figure 2 shows the research model in figure 1 with results. The numbers beside lines are correlation coefficients. Because grade is not normally distributed, the three correlation coefficients associated with grade are Spearman's rho. Others are Pearson's R. These coefficients reflect the survey results of the 52 students in the collaborative exam. Using survey results of all 93 students, significant differences are found and highlighted with bold box outlines in figure 2.

Figure 2. Research Model with Results

*: significant at p<.05 level   **: significant at p<.01 level

Bold box outlines indicate significant difference between the two exam conditions.
DISCUSSION AND FUTURE RESEARCH

Students’ answers to the two-opened questions provide further insights into what they liked and disliked most about the collaborative exam process. For example, students report the main advantages include: “I believe that it (the collaborative exam) provided the true experience of distance learning,” “It gives me less pressure, and I think it is more flexible to really learn the knowledge,” “The best part about the online exam process was the capability and the time to thoroughly review and research the articles that was covered during the semester. Memorization skills was not a requirement for this type of exam which most traditional close-book would have required,” “The great thing is that, we are given time for critical thinking,” “The best I liked was that I wrote my point of views and gather others view on a particular subject.”

In describing what they liked least about the process, student concerns center around the grading process. For example, students report: “some of the grading is very subjective based on the answer...if the grader agrees with your ideas then the grade will tend to be higher, if not, watch out below....” “Students should not be allowed to grade papers as they are not necessarily capable of providing the correct judgment on what constitutes a right answer for the questions they set.” While the validity and reliability of peer grading has been established in previous meta-analysis studies (Falchikov and Goldfinch, 2000; Topping, 1998), students still seem to have a lack of confidence in themselves in grading other’s answers, nor do they trust their peers for grading their answers. We think the ability to critically evaluate other’s work is essential in people’s professional life, and should be fostered in school education. In addition, further revisions to our grading system may be helpful (such as providing secondary-level grading by advanced students) in enhancing the perceived fairness in grading.

We plan to further our investigation of the collaborative examination through a series of studies. One area of research currently underway is based on collaborative learning theories (Swan and Shea, 2004) and online interaction studies in ALN research (Hiltz and Goldman, 2004). We want to investigate whether the current collaborative exam process can be further enhanced through small group activities. Field experiments have been designed and will be conducted. While our studies so far have used only graduate-level courses, we believe the method can be extended to undergraduate students, probably with more scaffolding by the instructor. Further investigations of the adoption of collaborative exam in different disciplines, cultures and nations are also being explored. We believe the collaborative examination approach is beneficial across a wide variety of course levels, disciplines and cultures.

ACKNOWLEDGEMENTS

Many thanks to Hee-kyung Cho for her excellent advice on statistical analysis. We gratefully acknowledge partial funding support for this research by the United Parcel Service Foundation, the New Jersey Center for Pervasive Information Technology, the New Jersey Commission on Science and Technology, and the National Science Foundation under grants IIS-0135531 and DUE-0226075.

REFERENCES