Collaborative Examinations for Asynchronous Learning Networks: Evaluation Results

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

[+] *Collaborative Hypermedia Research Lab*
Computer and Information Science Department
New Jersey Institute of Technology
University Heights
Newark, NJ, 07102-1982 USA

* *Ph.D. in Management Program*
Graduate School-Newark
Rutgers University
180 University Avenue
Newark, NJ 07102-1897 USA

Abstract

This paper presents the evaluation results of two student surveys on a collaborative examination process using ALN for a graduate-level course at NJIT. The exam process includes students making up questions, picking out questions, answering, grading, and appealing the grades. The process was conducted on the Virtual Classroom and Webboard during the fall 1999 and spring 2000 semesters, with some revision of the process in the second semester. The surveys following each exam elicited feedback from 138 students. Results show the majority of students felt they learned throughout the process, the exams were successful in demonstrating what they learned, and it was an enjoyable process. Students' concerns and our experiences are presented as well as suggestions for future research on this topic.

1.   Introduction

Asynchronous Learning Networks (ALN), as defined by A. F. Mayadas, are "people networks for anytime - anywhere learning" [15]. They stress two things. First, ALN provides the flexibility in accessing the learning environment at the convenience of learners, "any time, any where," through the Internet. Second, ALN emphasizes peer-to-peer interaction and collaboration in the learning process. Although this form of education has been proven effective through years of operation and studies, [1][8][9][10][17], examinations have seldom been conducted on-line, at least not at NJIT, and the issue of appropriate forms of examination for online courses is seldom treated in the published literature.   Distance learning students usually have to commute to designated campuses to take exams or have an approved proctor.

Can we take advantage of ALN to conduct exams so students can enjoy anytime-anywhere flexibility as well as maximizing their learning through collaboration? To address this question, two collaborative exams were conducted at NJIT in the fall 1999 and spring 2000 semesters, based on a similar process. This first author served as the teaching assistant for the course and assisted in all phases of the exam process in the fall semester, and conducted student surveys in both semesters. The purpose of this paper is to present evaluation results of the processes and to share our experience in conducting collaborative examinations. This section reviews characteristics of collaborative education, the goals of examinations, as well as previous research on on-line and collaborative examinations.

1.1.   Collaborative Education

Collaborative learning is characterized as learner-centered rather than teacher-centered. The role of teachers changes from transferring knowledge to students, to serving as a facilitator in the students' construction of their own knowledge [2][3][10][11]. The shifted paradigm results in more student involvement and engagement in the learning process, and thus more incentive for learning and better achievement [6] [12].

The collaborative examinations that are evaluated in this paper can be regarded as one form of a collaborative learning process. They aimed to maximize students' involvement in the process, which hopefully should also be a valuable learning experience. 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 exams were conducted through cooperation with the whole class.

1.2. Aims of Examination

Hay [7] 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 course. Ebel and Frisbie [5] point out that the main purposes of a test are to measure student achievement and to 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 itself. In addition, the process of constructing tests should cause instructors to think carefully about the goals of instruction in a course.

1.3. Previous Research on Collaborative Examination

Previous study of on-line examination processes is quite limited, and mainly conducted for computer programming courses. Though different from the collaborative examination processes that we will discuss below, some effectiveness of automated objective on-line examinations has been reported in these studies. Mason and Woit [14] highlighted the potential advantages of on-line examinations by encouraging students to attain practical skills, and reducing the level of cheating and copying. The analysis by Kumar [13] 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.

One exam process that is more like ours was conducted at the business school of California State University, Chico. [4] Students were asked to do team-based final exams using GroupSystems™ software to complete open-ended questions. The process included students' individual response to a question, comments on others' answers, and finally team-based verbal discussion with the instructor. This process differs from ours with our process also featuring student question design, student grading, Ph.D. student intermediate grading, and the collaboration of the class as a whole.

In the following section, we introduce the course in which the exams were conducted, the exam procedures, and ALN tools that were used in the process.

2. Course Background, Examination Procedures, and ALN Tools

2.1. Course Background

The on-line collaborative examinations we are describing here were conducted in a graduate-level course for both masters and Ph.D. students in Information Systems at NJIT, called "CIS 677: Information System Principles," in the fall 1999 and spring 2000 semesters. 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 videotapes of the lectures. Students in both sections are required to participate in the on-line discussion, which was on Virtual Classroom™ (VC) in fall 99 and Webboard™ in spring 2000. More information on the two systems can be found below. Several conferences were established on each of them to facilitate on-line discussion of CIS 677 topics.

2.2. Examination Procedures

Previously, one exam was scheduled in the middle of the semester for CIS 677. It was a three-hour in-class proctored exam of 3-4 essay questions, with six pages of notes allowed. In the fall 1999 semester, encouraged by the program director, the instructor (Processor Michael Bieber) designed a collaborative examination process. The motivation was first to reduce the instructor's workload, which was overloaded that semester, and secondly to investigate a better examination method. This new approach was carried out in both sections that semester, with 73 students in total, where 46 were in the Face-to-Face (FTF) section and 27 in the Distance-Learning (DL) section. When the evaluation results in the first semester showed favorable feedback towards the process, a similar collaborative exam was conducted in the following spring 2000 semester. This time three in-class sections with a total of 80 students and one distance learning section with 34 students took part in the exam. The first exam process spanned five weeks and the second was streamlined to about 3.5 weeks.

The basic procedures of the examination were the same in both semesters. First, each student composed essay questions; next, each student selected one question and answered it; third, the student who created the question graded the answer and provided a justification of the grading. Subsequently, Ph.D. students enrolled in the

course did an intermediate review of the grading, and lastly, the instructor provided a final grade. If the scores of the grader and intermediate 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 himself. (The instructor ended up regrading between 20-30% of the questions.)

Based on students' feedback elicited from the first survey and instructor's experience, the following changes were made in the second semester:

• The system environment was changed from Virtual Classroom to Webboard;
• The process was made anonymous;
• Two essay questions were solicited instead of one;
• People were assigned a time-slot to pick up questions;
• The entire process was streamlined to 3.5 weeks.

Since the on-line discussion of the course in the second semester was on Webboard, the exam was also carried out there. To increase fairness, the process was changed to an anonymous one. Students were assigned a question ID and an answer ID and used them in their postings. Unfortunately, during the second semester, there was a system crash during the question reservation process. To try to recover from lost time and lost items, the instructor asked students to select and answer only one question each and graders had to be reassigned to make sure every student graded one question.

2.3. ALN Tools

In ALN, collaboration is achieved through computer networks. [1] Virtual Classroom™ (VC) and Webboard™ are the two asynchronous conferencing tools used to facilitate learning and the collaborative exam processes being discussed in the paper. VC has been used at NJIT for years and is where the first exam was conducted. Webboard™ use was begun in Spring 2000 and served as the environment where the second exam occurred. Both systems have conferences, which are separate threaded discussion areas for a topic where students can post, read and reply to comments of others. In Fall 1999 three conferences were constructed on VC for the exam. One dealt with administrative information on the exam process; one served as the main conference where all the questions, answers, grades were posted; and the last one was reserved for Ph.D. students and the instructor for grade review. In spring 2000 one main conference was created in Webboard for the main exam activities; administration information and students feedback were posted in the regular general administration and feedback conferences. In both semesters, several important announcements, such as how to compose a question and grading procedures, were posted on the course web site.

3. Evaluation Method: The Surveys

To evaluate the collaborative on-line examination processes, surveys were conducted within a couple of weeks after the examination processes finished. The questionnaires were distributed in the classroom to FTF students and also were put on the course web site for DL students. The table below shows the number of subjects registered in each session and the response ratio for the analysis in this paper.

Table 1. Number of Subjects in the Two Surveys

	Fall 1999	Spring 2000
FTF	41	60
DL	21	15
Total	63*	75
Return Rate	86.3%	65.8%

There is one returned questionnaire missing answers to the question on section.

As can be seen, the return rates in both surveys are good. (The instructor assigned three extra credit points towards the course grade to whomever participated in the survey.) Some questions were adapted from those used in the "Course Questionnaire – Virtual University Project" [8]. We solicited students' opinions concerning the following research questions:

Q1: Do students feel they learn from the exam process?
Q2: Are they satisfied with the process?
Q3: What do they think about the grading system?
Q4: Does the collaborative nature of the process affect students in any way?

Q5: Is the collaborative exam effective in testing the mastery of students' knowledge?
Q6: Did they like the anonymous process in the spring session?

4. Findings

The raw data, gathered from the two semesters, were analyzed using Excel™ and SPSS™. We did not mix the two sets of data in our analysis because there were several changes made in the second process. In spite of the

changes, results show most of the feedback for both processes is similar.

4.1. Learning Effects

Table 2 below addresses the question: Do students feel they learn from the exam process? The majority of students in both semesters agreed that all phases of the exam process were part of the learning process. One interesting thing to notice is that the grading process is perceived as the most valuable learning activity in fall 1999 (mean=3.82) while in spring 2000 it ranks the lowest (mean=3.41). This may be attributable to the system crash and the reassignment of graders to questions

as mentioned above. This resulted in about half students ending up grading answers to questions that were not composed by them, which frustrated them and reduced their perceived learning.

Ph.D. students were asked about whether they learned through their role as intermediate graders. Three valid answers were obtained in fall 1999 with a mean score of 3.33. Eleven answers were received in the following semester and the mean increased to 3.73.

Students were also asked to compare this exam process with the traditional one on items related to learning (Table 3).

Table 2. Students Perception of Learning Effects of the Examination Process

Item	Session	Number of Responses	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	Mean	Std. Dev.
I learned from making up a question	Fall 99	63	4.76%	17.46%	22.22%	42.86%	12.70%	3.41	1.07
	Spring 00	75	4%	18.7%	21.3%	33.3%	22.7%	3.52	1.16
I learned from looking at all the other questions	Fall 99	63	9.52%	14.29%	20.63%	46.03%	9.52%	3.32	1.13
	Spring 00	75	5.3%	10.7%	16%	48%	20%	3.67	1.08
I learned from looking at others' answers	Fall 99	63	4.76%	14.29%	20.63%	46.03%	14.29%	3.51	1.06
	Spring 00	75	6.7%	14.7%	20%	38.7%	20%	3.51	1.17
I learned from grading others' answers	Fall 99	62	3.23%	12.90%	9.68%	46.77%	27.42%	3.82	1.08
	Spring 00	73	9.6%	11%	23.3%	41.4%	15.1%	3.41	1.16

Table 3. Percentage of students reporting other learning effects of the process

Item	Session	Frequency Favorable (M + MM)*
I became more interested in the subject.	Fall 99	57.1%
	Spring 00	58.1%
I learned a great deal of factual material.	Fall 99	52.4%
	Spring 00	62.2%
I gained a good understanding of basic concepts.	Fall 99	57.1%
	Spring 00	58.1%
I developed the ability to communicate clearly about this subject.	Fall 99	68.3%
	Spring 00	64.9%
My skill in critical thinking was increased.	Fall 99	68.2%
	Spring 00	73%
My ability to integrate facts and develop generalizations improved.	Fall 99	73%
	Spring 00	75.3%
I was stimulated to do additional reading.	Fall 99	62.9%
	Spring 00	67.6%
I became more confident in expressing my ideas.	Fall 99	54%
	Spring 00	64.4%
I was motivated to do my best work.	Fall 99	65.1%
	Spring 00	68.5%

Table 4 below shows students found the collaborative online exam to be a less stressful experience than traditional exams, in both sessions.

The grading procedure is quite unique in this process. Below (Table 5) are students' responses to questions regarding the grading process. Students' satisfaction with the grading procedure was lower in the spring semester than in the fall. The percentage of students in the "strongly disagree" category rose to above 20% for items regarding grading fairness, students' capability to grade, and whether the instructor should have done all the grading. Possible reasons include the reassignment of graders to questions and the grading criteria themselves. Further discussion on grading can be found in the discussion section.

Table 4. Percentage of students reporting their satisfaction with the examination process

Item	Session	Number of Response	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	Mean	Std. Dev.
Comfortable Timeframe	Fall 99	44	2.27%	2.27%	15.91%	52.27%	27.27%	4.00	0.86
	Spring 00	75	10.7%	18.7%	13.3%	44%	13.3%	3.31	1.23
Flexibility	Fall 99	44	.00%	2.27%	25.00%	38.64%	34.09%	4.05	0.83
	Spring 00	75	6.7%	2.7%	21.3%	57.3%	12%	3.65	.97
More Pressure *	Fall 99	43	30.23%	34.88%	18.60%	13.95%	2.33%	3.77*	1.11
	Spring 00	74	14.9%	33.8%	24.3%	18.9%	8.1%	3.28*	1.18

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

Table 5. Attitudes toward the grading system

Item	Session	Number of Response	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)	Mean	Std. Dev.
Grading fairness	Fall 99	44	2.27%	20.45%	25.00%	45.45%	6.82%	3.34	0.96
	Spring 00	75	20%	16%	21.3%	36%	6.7%	2.93	1.27
Students' grading *	Fall 99	44	4.55%	38.64%	15.91%	40.91%	.00%	3.07*	1.00
	Spring 00	75	6.7%	25.3%	22.7%	25.3%	20%	2.73*	1.23
Ph.D. students grading	Fall 99	44	6.82%	13.64%	31.82%	43.18%	4.55%	3.25	0.99
	Spring 00	75	6.7%	13.3%	28%	46.7%	5.3%	3.31	1.00
Instructor grading *	Fall 99	44	25.00%	15.91%	29.55%	20.45%	9.09%	3.27*	1.30
	Spring 00	75	2.7%	10.7%	29.3%	33.3%	24%	2.35*	1.05

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

Does the collaborative nature of the process affect students in any way? Table 6 below presents the results on the items: "I learned to value other points of view" and "I developed new friendships in this class."

Table 6. Student reports regarding the collaborative nature of the process

Item	Session	Number of Response	Much Less (1)	Less (2)	No Difference (3)	More (4)	Much More (5)	Mean	Std. Dev.
I learned to value other points of view	Fall 99	63	.00%	.00%	25.40%	60.32%	14.29%	3.89	0.63
	Spring 00	74	2.7%	4.1%	25.7%	47.3%	20.3%	3.78	.91
I developed new friendships in this class	Fall 99	63	9.52%	3.17%	46.03%	25.40%	15.87%	3.35	1.09
	Spring 00	74	4.1%	4.1%	39.2%	33.8%	18.9%	3.59	.98

In addition, we elicited students' opinions about doing question design by themselves. In both sessions the majority of students perceived their peers as capable in question design. The mean for the question: "I don't think students were able to design good questions that measured the learning objectives of the course," when converted to score positive responses as high, is 3.22 in both semesters.

In terms of testing mastery, the fall session yielded 59.1% of students agreeing with "The exam was successful in enabling me to demonstrate what I learned

in class," while 15.9% disagreed and 25% were neutral. The ratios on the spring session are 65.3%, 20% and 14.7% respectively.

The process in the second semester was anonymous. Did students enjoy it? The results show 89% of students agreed that they liked the anonymity provided in the process. 14.4% admitted that they tried to guess authors of the postings, and 8% admitted letting others know their question id or answer id.

The overall evaluations of students are shown below (table 7 & 8) on three key questions which compared the collaborative online exam with a traditional exam: "I mastered the course material;" "I enjoyed the examination process;" and "I would recommend the process in the future." Results from both sessions 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 7. Overall evaluation of students (comparing the collaborative exam process with traditional one)

Item	Session	Number of Response	Much Less (1)	Less (2)	No Difference (3)	More (4)	Much More (5)	Mean	Std. Dev.
Mastery of material	Fall 99	63	.00%	11.1%	28.6%	52.4%	7.9%	3.57	0.80
	Spring 00	73	2.7%	2.7%	26.0%	60.3%	8.2%	3.68	.78
Enjoyment of the process	Fall 99	63	9.5%	12.7%	15.9%	38.1%	23.8%	3.54	1.25
	Spring 00	74	6.8%	9.5%	32.4%	41.9%	9.5%	3.38	1.02

Table 8. Overall recommendation of the process to be used for other courses

Item	Session	Number of Response	Strongly Oppose (1)	Oppose (2)	Neutral (3)	Recom-mend (4)	Strongly Recom-mend (5)	Mean	Std. Dev.
Recommendation of the process	Fall 99	62	6.5%	16.1%	22.6%	38.7%	16.1%	3.42	1.14
	Spring 00	73	5.5%	8.2%	24.7%	45.2%	16.4%	3.59	1.04

Thus, the study indicates that students report subjective satisfaction with the collaborative exams in general.

4.2. Preliminary Factor Analysis and Correlation Analysis

To find out the underlying dimensions, or factors, in students' feedback, which might explain their perception of the process, we used factor analysis for data gathered from the second survey. Two meaningful factors were found in the scale, which can be interpreted as a "learning index" and "grading index". Learning index shows the perceived learning effect of the exam and is composed of the following five items: "I learned a great deal from having to make up a question," "I learned from looking at other questions," "I learned from looking at others' answers," "I learned from grading another student's answer," and "The exam was successful in enabling me to

demonstrate what I learned in class". Reliability analysis shows the above scale is highly reliable with its Cronbach's Alpha equal to 0.84. The grading index describes the perceived fairness of the grading system used in the process and includes the following four questions: "I felt the grading process was fair," "I don't think students were capable of grading the responses to the questions they designed," "The Ph.D. students were capable of improving the grading by other students on the questions," and "It would have been an improvement if the instructor had done all the grading". The Cronbach's Alpha for the grading scale is also fairly high (0.79). Further, correlation analysis shows both learning and grading factors are significantly related to students' overall experience, which includes enjoyment in the process, mastery of course material and recommendation for future use. Table 9 displays the results.

Table 9. Underlying factors: Learning Index and Grading Index

Factors	[1] Items	Mean	Reliability – Alpha	Overall Experience(Pearson' r)		
				Mastery of material	Enjoyment	Recommendation
Learning Index	Learned from making up a question	3.54	0.84	.69**	.67**	.67**
	Learned from looking at other questions					
	Learned from looking at others' answers					
	Learned from grading					
	Demonstrate what I learned in class					
Grading Index	Grading process was fair	2.83	0.78	.33**	.44**	.40**
	Students' capability to grading					
	Ph.D. grading					
	Instructor does all grading					

**Significant at p<.05 level (T-test, two-tailed)

4.3. Comparison between DL and FTF

In order to find out whether the mode in which the sections were offered (distance learning vs. face-to-face students) had any effects on their evaluation, T-tests were conducted on dependant variables. Table 10 shows those items that have apparent differences between groups in terms of the mean values. In both the Fall 1999 and Spring 2000 semesters, the item "I developed new friendships in this class" is significantly different between DL and FTF students. The mean for the FTF section in Fall 1999 is 0.73 higher than that for DL, while in the following semester the difference increased to 0.75. This means the collaborative exam process plus face-to-face interaction in class significantly facilitate the development of friendship among students, while this effect on the distance learning section is not so apparent.

It is unclear what impact the examination has had on developing friendships. Given that the process was anonymous in the second year, reading a person's question and answer should not bring development of new relationships. Considering that 8% of participants revealed their identity to others, however, indicates that people might have gotten together to collaborate on developing questions or on grading, and this strengthened friendships. (Another issue is whether this could be considered cheating.) These are areas we should investigate more deeply.

Table 10. Differences between Distance Learning and Face-To-Face Sections

	Fall 1999					Spring 2000				
	Section	N	Mean	Std. Dev.	Sig.	Section	N	Mean	Std. Dev.	Sig.
I developed new friendships	**DL**	21	2.9	1.00	**	**DL**	15	3.00	.85	**
	FTF	41	3.63	1.02		**FTF**	59	3.75	.96	
I enjoyed the flexibility	**DL**	18	4.44	.62	**	**DL**	15	3.73	.96	
	FTF	25	3.76	.88		**FTF**	60	3.63	.97	
I learned to value other's point of view	**DL**	21	3.86	.73		**DL**	15	3.07	1.03	**
	FTF	41	3.90	.58		**FTF**	59	3.97	.79	
I enjoyed the exam process	**DL**	21	4.05	1.16	**	**DL**	15	3.27	1.10	
	FTF	41	3.32	1.23		**FTF**	59	3.41	1.00	
I would recommend the process	**DL**	21	3.76	1.18		**DL**	15	3.47	.99	
	FTF	40	3.28	1.09		**FTF**	58	3.62	1.06	

**Significant at p<.05 level (T-test, two-tailed)

4.4. Impact of Students' Grades on Their Perceptions

We are also interested in knowing whether the grade students received affected their perception of the process. Through general linear model analysis, we find that students' enjoyment with the process and their recommendation of the process for future use are significantly (at the .05 level) related to the grade they got, with R squared for both variables equal to .99, while the grade did not affect their perception of mastery of material significantly. Further, using linear regression, the correlation coefficient R of instructor's grade and students' enjoyment as well as their recommendation for future use are R=.325 and R=. 252 respectively, both of which are significant at the .05 level.

5. Discussion

In general, students reported a favorable attitude towards their collaborative 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 students 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."

Students were asked about how they composed questions. From students' answers to the open-ended questions, most of them made up questions based on the synthesis of course topics and materials, which is a requirement from the instructor, with an emphasis on either their interest in topics, work experiences or latest hot issues. Some students regarded it as an important learning step in the process. As one student said: "It was an important part to select and compose a meaningful question. I had to do extra reading from articles, journals and online publications." It would be logical to assume that those who gave extra effort in this step also learned more from the process.

Although the evaluations provide overall favorable results, students also showed some concerns. The grading system is the most disputed component and stimulated discussions from students in both semesters. In fall 1999, 20.5% of students disagreed that the grading system was fair, and 2.3% strongly disagreed. Though the professor improved the grading criteria in the hope of making it more explicit in spring 2000, even more concerns arose. 16% of Spring 2000 students disagreed that the grading system is fair, and the ratio for "strongly disagree" increased to 20%. The mean of this item in the first and the second semester is 3.34 and 2.93 respectively. One of the reasons that might explain the decreased satisfaction with the grading system in spring 2000 was the instructor's response to the system crash, reducing the number of questions to respond to from two to one. This resulted in some students having had both of their questions answered while some had only one and others had none. The instructor ended up having to reassign graders to questions to ensure each student would grade one and only one question. This resulted in about half of the answers being graded by students who were not composers of questions.

Other reasons, which may be more important, are the grading criteria themselves. The criteria used in Fall 99 break points down to eight parts, such as synthesis quality, class readings, and background/framing, with certain points assigned to each part. Based on the results of the first exam, the instructor revised the criteria and provided categories to grade each part. These categories are described as "great," "good," "ok," "poor" and "zero," with a numeric grade assigned for each category in every part, such as: background/framing - 10 points; grading: great - 10; good - 8; ok - 5; poor – 3. Graders also had to explain each part's grade with a 2-3 sentence justification. By making the grading more explicit, students were more likely to score several parts as "good" instead of "great." This resulted in lowering scores overall from the fall semester. Someone who had answered most parts well ("good") easily could receive a grade in the 70s or low 80s, which distressed many students.

This actually raises a cultural issue concerning "grade inflation." Many students in the United States have come to expect a grade of "good" to translate to a numeric grade in the high 80s or low 90s, whereas in many other countries, a grade in the 70s or low 80s is actually very "good." Making the grading criteria explicit forced the student graders to critically examine examination answers and determine which components were excellent and which were not—quite possibly for the first time in their academic careers. Students thus clearly bumped up against the "grade inflation" issue. Resolving it is a task that goes far beyond the collaborative examination process. (As a partial solution, the instructor is considering building in a lot of extra credit in the grading criteria. Because these explicit grading criteria are announced far in advance, students should be prepared to

add extra credit components to their answers, and thus raise their grades.)

One possible improvement might be, instead of requiring students to give a numeric grade, instruct them to only critique the answer and provide justification. In this way they are still given the chance to learn from reading and critiquing another's answers. This would preserve much of the learning that students get from the learning procedure as well as increase perceived fairness in grading. This would require the instructor, however, to grade a different essay question from every student.

One of the greatest fears students expressed going into the process was grading : were they qualified to grade their peers? Is it appropriate for them to assume a role traditionally delegated to the instructor? We strongly believe that graduate students, many of whom hold full-time jobs, are quite capable of evaluating others. But applying evaluation from the workplace and other everyday experiences to grading essay questions is a non-trivial skill transfer. We do realize that evaluation is a skill that must be learned, and we hope the explicit grading criteria provided the means to begin.

We used Ph.D. students as intermediate-level graders in part to monitor how well the masters students evaluated, and in part to begin developing their own evaluation skills. Not every Ph.D. student did high-quality evaluations, and the instructor had to work closely with a few to hone these skills. (Many professors wish that they had had opportunities to learn how to grade before entering their first teaching job!)

6. Future Research

For courses that are based on discussion and collaborative learning, a high-pressure individual examination imposes an examination process that is not in keeping with the desired learning processes. However, devising an online collaborative examination process that will be "as good or better" in all respects is not an easy task. We plan to continue field research on how best to structure and manage such a process.

We believe the idea of a collaborative examination is a good one for ALN based courses. Our experience suggests that improved success may lie in the following key factors:
- Selection of an appropriate and reliable computer-mediated environment
- Improvement in the grading procedures
- Improving the Instructor's role in the process

Both VC and Webboard were incapable of handling large numbers of students logging in at the same time, at least on our servers, and it happened in both question selection processes that the system became extremely slow or even crashed. This caused stress and chaos for students. A software-hardware system that can better support up to 100 users logging in simultaneously can alleviate the problem. Or, an improved time-slot procedure can be used to constrain the number of students logging in at the same time.

The grading system needs the most improvement, as it received the most complaints.

Last but not least, the instructor's role in the process is extremely important. This collaborative exam stresses the students' active participation throughout the process, from making up questions to grading, with which most of them had no prior experience. How to ensure that the process is a beneficial learning process and achieves the purpose of testing mastery depends on careful planning by the instructor.

There is also room left for improvement in our evaluation process. First, the survey results can only report the subjective learning effects, which are perceived learning, instead of objective or actual learning. [1] Since there is no traditional exam carried out in addition to the on-line exam, there is no objective measurement available to test the real learning effect of the exam process and its effectiveness in testing mastery of students. A field experiment might possibly be used in the future which uses both types of exams and compares their results. Secondly, because of the changes and their consequences in the second semester, there are confounding variables in our analysis that might explain the difference between the surveys. Examples are the change from VC to Webboard, the Webboard crash, students' confounding of the overall evaluation of the course instead of the midterm procedure itself, overall experience in class, and the fact that spring semester students were told previous students were most satisfied with the results. Third, using various methods in addition to questionnaires in data collection would provide multiple perspectives on an issue and yield stronger substantiation of constructs. Fourth, If the data were gathered from several courses, longitudinally, the results might be more generalizable.

Overall, the collaborative examination did achieve its goal of increasing learning during the examination process. It reduced the workload of the instructor, allowing him to focus attention on other parts of the learning process and the course. It gave students experience in evaluating academic work and their peers. Through this, students gained a better understanding of a professor's job. Many remarked that they never realized how challenging it is to be a thorough and fair grader. Collaborative examinations hold the promise of teaching students in a new way. Once refined, they could form a critical part of a technology-supported curriculum.

Acknowledgements

References

[1]   R. Benbunan-Fich and S. R. Hiltz, "Learning Effects of Asynchronous Learning Networks: A Comparison of Groups and Individuals Solving Ethical Case Scenarios." In the *Proceedings of the 31st Hawaii International Conference on System Science,* 1998.

[2]   C. Bouton and R. Y. Garth, "Learning in Groups," *New Directions in Teaching and Learning*, No. 14, San Francisco: Jossey-Bass, 1983.

[3]   K. A. Bruffee, "Backgroud and History to Collaborative Learning in American Colleges," *College English*, 7, 1984, pp. 635-652.

[4]   G. Corbitt, L. Wright, and B. Martz, "Addressing the Challenges of the Future: Implemeting a Collaborative Student Enviroment at a University Business School." In the *Proceedings of the 32nd Hawaii International Conference on System Science,* 1999.

[5]   R. L. Ebel and D. A. Frisbie, *Essentials of Educational Measurement*, Fourth edition, Prentice - Hall Inc. Englewood Cliffs, NJ, 1986.

[6]   L. Harasim, (Ed.) *On-Line Education: Perspectives on a New Medium*. New York: Praeger/Greenwood, 1990.

[7]   I. Hay, "Examinations: Achievement Tests," *Journal of Geography in Higher Education*, Vol. 20, 1, 1996, pp.137-142.

[8]   S. R. Hiltz, *The Virtual Classroom,Learning Without Limits via Computer Networks*, Ablex Publishing Corporation, Norwood, NJ, 1994.

[9]   S. R. Hiltz, "Impacts of College-level Courses via Asynchronous Learning Networks: Some Preliminary Results," *Journal Of Asynchronous Learning Networks, 1997.*

[10] S. R. Hiltz, "Collaborative Learning in Asynchronous Learning Networks: Building Learning Communities," New Jersey Institute of Technology, Copyright, 1998. www.eies.njit.edu/~hiltz.

[11] D. W. Johnson and R. T. Johnson, *Learning Together and Alone: Cooperation, Competition, and Individualization*, Engelwood Cliffs, Prentice Hall, 1975.

[12] D. W. Johnson, "Student-student Interaction: The neglected variable in education," *Educational Research*, Vol. 10, 1, 1981, pp. 5-10.

[13] A. N. Kumar, "On changing from Written to On-line Tests in Computer Science I: An assessment," *Proceedings of the 4th annual SIGCSE/SIGCUE on Innovation and technology in computer science education* , 1999, Pages 25 - 28.

[14] D. V. Mason and D. M. Woit, "Integrating Technology into Computer Science Examinations," *Proceedings of the twenty-ninth SIGCSE technical symposium on Computer science education* , 1998, Pages 140 - 144

[15] A. F. Mayadas, "What is ALN?," 2000, http://www.aln.org/alnweb/aln.htm

[16] R. Rosenthal and R. L. Rosnow, *Essentials of Behavioral Research: Methods and Data Analysis*, Second Edition, McGraw-Hill Series in Psychology, 1991.

[17] S. J. Sloffer, B. Dueber, and T. M. Duffy, "Using Asynchronous Conferencing to Promote Critical Thinking: Two Implementations in Higher Education." In the *Proceedings of the 32nd Hawaii International Conference on System Science,* 1999.